PTO/SB/05 (05-03)
Approved for use through 04/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office. U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	AOL0142
First Inventor	Fish
Title	Search Enhancement System Having...
Express Mail Label No.	EL 816157440 US

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO: **Commissioner for Patents, Mail Stop Patent Application, P.O. Box 1450, Alexandria VA 22313-1450**

1. [X] Fee Transmittal Form (e.g., PTO/SB/17) *(Submit an original and a duplicate for fee processing)*
2. [ ] Applicant claims small entity status. See 37 CFR 1.27.
3. [X] Specification [*Total Pages* 34]
   (preferred arrangement set forth below)
   - Descriptive title of the invention
   - Cross Reference to Related Applications
   - Statement Regarding Fed sponsored R & D
   - Reference to sequence listing, a table, or a computer program listing appendix
   - Background of the Invention
   - Brief Summary of the Invention
   - Brief Description of the Drawings *(if filed)*
   - Detailed Description
   - Claim(s)
   - Abstract of the Disclosure
4. [X] Drawing(s) (*35 U.S.C. 113*) [*Total Sheets* 15]
5. Oath or Declaration [*Total Sheets* 2]
   a. [ ] Newly executed (original or copy)
   b. [ ] Copy from a prior application (37 CFR 1.63(d)) *(for continuation/divisional with Box 18 completed)*
      i. [ ] DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) name in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).
6. [ ] Application Data Sheet. See 37 CFR 1.76
7. [ ] CD-ROM or CD-R in duplicate, large table or Computer Program *(Appendix)*
8. Nucleotide and/or Amino Acid Sequence Submission *(if applicable, all necessary)*
   a. [ ] Computer Reader Form (CRF)
   b. Specification Sequence Listing on:
      i. [ ] CD-ROM or CD-R (2 copies); or
      ii. [ ] Paper
   c. [ ] Statements verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

9. [ ] Assignment Papers (cover sheet & document(s))
10. [ ] 37 CFR 3.73(b) Statement *(when there is an assignee)* [X] Power of Attorney
11. [ ] English Translation Document *(if applicable)*
12. [ ] Information Disclosure Statement (IDS)/PTO-1499 [ ] Copies of IDS Citations
13. [ ] Preliminary Amendment
14. [ ] Return Receipt Postcard (MPEP 503) *(Should be specifically itemized)*
15. [ ] Certified Copy of Priority Document(s) *(if foreign priority is claimed)*
16. [ ] Nonpublication Request under 35 U.S.C. 122 (b)(2)(B)(i). Applicant must attach form PTO/SB/35 or its equivalent.
17. [ ] Other: .....

18. If a CONTINUING APPLICATION, *check appropriate box, and supply the requisite information below and in the first sentence of the specification following the title, or in an Application Data Sheet under 37 CFR 1.76:*

[ ] Continuation [ ] Divisional [ ] Continuation-in-part (CIP) of prior application No.: .....

Prior application information: *Examiner* _____ *Art Unit:* _____

For CONTINUATION OF DIVISIONAL APPS only; The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 5b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

19. CORRESPONDENCE ADDRESS

[X] *Customer Number or Bar Code Label* 22862 *(Insert Customer No. or Attach bar code label here)* *OR* [ ] *Correspondence address below*

Name					
Address					
City		*State*		*Zip Code*	
Country		*Telephone*		*Fax*	

Name (Print/Type)	Michael A. Glenn	*Registration No. (Attorney/Agent)*	30,176
Signature		*Date*	10/14/2003

PTO/SB/17 (01-03)
Approved for use through 04/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL for FY 2003

Effective 01/01/2003. Patent fees are subject to annual revision.

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 953.00

Complet if Known	
Application Number	Unassigned
Filing Date	Herewith
First Named Inventor	Fish
Examiner Name	Unassigned
Art Unit	Unassigned
Attorney Docket No.	AOL0142

METHOD OF PAYMENT *(check all that apply)*

☐ Check ☐ Credit card ☐ Money Order ☐ Other ☐ None

☑ Deposit Account:

Deposit Account Number: 07-1445

Deposit Account Name: Glenn Patent Group

The Commissioner is authorized to: *(check all that apply)*

☑ Charge fee(s) indicated below ☑ Credit any overpayments

☑ Charge any additional fee(s) during the pendency of this application

☐ Charge fee(s) indicated below, except for the filing fee to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	750	2001	375	Utility filing fee	770.00
1002	330	2002	165	Design filing fee	
1003	520	2003	260	Plant filing fee	
1004	750	2004	375	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	

SUBTOTAL (1) ($) 770.00

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

		Extra Claims		Fee from below	Fee Paid
Total Claims	33	-20** = 13	x	18.00	= 183.00
Independent Claims	3	- 3** = 0	x	86.00	= 0
Multiple Dependent					=

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	** Reissue independent claims over original patent
1205	18	2205	9	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($) 183.00

**or number previously paid, if greater; For Reissues, see above

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet	
1053	130	1053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for *ex parte* reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	410	2252	205	Extension for reply within second month	
1253	930	2253	465	Extension for reply within third month	
1254	1,450	2254	725	Extension for reply within fourth month	
1255	1,970	2255	985	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	320	2402	160	Filing a brief in support of an appeal	
1403	280	2403	140	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive - unavoidable	
1453	1,300	2453	650	Petition to revive - unintentional	
1501	1,300	2501	650	Utility issue fee (or reissue)	
1502	470	2502	235	Design issue fee	
1503	630	2503	315	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	750	2809	375	Filing a submission after final rejection (37 CFR 1.129(a))	
1810	750	2810	375	For each additional invention to be examined (37 CFR 1.129(b))	
1801	750	2801	375	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify) _____

*Reduced by Basic Filing Fee Paid SUBTOTAL (3) ($) 00.00

SUBMITTED BY				(Complete *if applicable*)	
Name (Print/Type)	Michael A. Glenn	*Registration No. (Attorney/Agent)*	30,176	*Telephone*	650-474-8400
Signature				*Date*	10/14/2003

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

This collection of information is required by 37 CFR 1.17 and 1.27. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, Washington, DC 20231.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

PTO/SB/05 (05-03)
Approved for use through 04/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office. U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

19270 U.S.PTO 10/685749 101403

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	AOL0142
First Inventor	Fish
Title	Search Enhancement System Having...
Express Mail Label No.	EL 816157440 US

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO: **Commissioner for Patents, Mail Stop Patent Application, P.O. Box 1450, Alexandria VA 22313-1450**

1. [X] Fee Transmittal Form (e.g., PTO/SB/17) *(Submit an original and a duplicate for fee processing)*
2. [ ] Applicant claims small entity status. See 37 CFR 1.27.
3. [X] Specification [*Total Pages* 34] *(preferred arrangement set forth below)*
   - Descriptive title of the invention
   - Cross Reference to Related Applications
   - Statement Regarding Fed sponsored R & D
   - Reference to sequence listing, a table, or a computer program listing appendix
   - Background of the Invention
   - Brief Summary of the Invention
   - Brief Description of the Drawings *(if filed)*
   - Detailed Description
   - Claim(s)
   - Abstract of the Disclosure
4. [X] Drawing(s) (*35 U.S.C. 113*) [*Total Sheets* 15]
5. Oath or Declaration [*Total Sheets* 2]
   a. [ ] Newly executed (original or copy)
   b. [ ] Copy from a prior application (37 CFR 1.63(d)) *(for continuation/divisional with Box 18 completed)*
      i. [ ] DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) name in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).
6. [ ] Application Data Sheet. See 37 CFR 1.76
7. [ ] CD-ROM or CD-R in duplicate, large table or Computer Program *(Appendix)*
8. Nucleotide and/or Amino Acid Sequence Submission *(if applicable, all necessary)*
   a. [ ] Computer Reader Form (CRF)
   b. Specification Sequence Listing on:
      i. [ ] CD-ROM or CD-R (2 copies); or
      ii. [ ] Paper
   c. [ ] Statements verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

9. [ ] Assignment Papers (cover sheet & document(s))
10. [ ] 37 CFR 3.73(b) Statement *(when there is an assignee)* [X] Power of Attorney
11. [ ] English Translation Document *(if applicable)*
12. [ ] Information Disclosure Statement (IDS)/PTO-1499 [ ] Copies of IDS Citations
13. [ ] Preliminary Amendment
14. [ ] Return Receipt Postcard (MPEP 503) *(Should be specifically itemized)*
15. [ ] Certified Copy of Priority Document(s) *(if foreign priority is claimed)*
16. [ ] Nonpublication Request under 35 U.S.C. 122 (b)(2)(B)(i). Applicant must attach form PTO/SB/35 or its equivalent.
17. [ ] Other: .....

18. If a CONTINUING APPLICATION, *check appropriate box, and supply the requisite information below and in the first sentence of the specification following the title, or in an Application Data Sheet under 37 CFR 1.76:*

[ ] Continuation [ ] Divisional [ ] Continuation-in-part (CIP) of prior application No.: .....

Prior application information: *Examiner* _____ *Art Unit:* _____

For CONTINUATION OF DIVISIONAL APPS only; The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 5b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

19. CORRESPONDENCE ADDRESS

[X] *Customer Number or Bar Code Label* 22862 *(Insert Customer No. or Attach bar code label here)* OR [ ] *Correspondence address below*

Name					
Address					
City		*State*		*Zip Code*	
Country		*Telephone*		*Fax*	

Name (Print/Type)	Michael A. Glenn	*Registration No. (Attorney/Agent)*	30,176
Signature		*Date*	10/14/2003

This collection of information is required by 37 CFR 1.53(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Mail Stop Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/SB/17 (01-03)
Approved for use through 04/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL for FY 2003

Effective 01/01/2003. Patent fees are subject to annual revision.

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 953.00

Complet if Known	
Application Number	Unassigned
Filing Date	Herewith
First Named Inventor	Fish
Examiner Name	Unassigned
Art Unit	Unassigned
Attorney Docket No.	AOL0142

METHOD OF PAYMENT *(check all that apply)*

☐ Check ☐ Credit card ☐ Money Order ☐ Other ☐ None

☑ Deposit Account:

Deposit Account Number: 07-1445

Deposit Account Name: Glenn Patent Group

The Commissioner is authorized to: *(check all that apply)*

☑ Charge fee(s) indicated below ☑ Credit any overpayments

☑ Charge any additional fee(s) during the pendency of this application

☐ Charge fee(s) indicated below, except for the filing fee to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	750	2001	375	Utility filing fee	770.00
1002	330	2002	165	Design filing fee	
1003	520	2003	260	Plant filing fee	
1004	750	2004	375	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	

SUBTOTAL (1) ($) 770.00

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

		Extra Claims		Fee from below	Fee Paid
Total Claims	33	-20** = 13	x	18.00	= 183.00
Independent Claims	3	- 3** = 0	x	86.00	= 0
Multiple Dependent					=

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	** Reissue independent claims over original patent
1205	18	2205	9	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($) 183.00

**or number previously paid, if greater; For Reissues, see above

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet	
1053	130	1053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for *ex parte* reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	410	2252	205	Extension for reply within second month	
1253	930	2253	465	Extension for reply within third month	
1254	1,450	2254	725	Extension for reply within fourth month	
1255	1,970	2255	985	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	320	2402	160	Filing a brief in support of an appeal	
1403	280	2403	140	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive - unavoidable	
1453	1,300	2453	650	Petition to revive - unintentional	
1501	1,300	2501	650	Utility issue fee (or reissue)	
1502	470	2502	235	Design issue fee	
1503	630	2503	315	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	750	2809	375	Filing a submission after final rejection (37 CFR 1.129(a))	
1810	750	2810	375	For each additional invention to be examined (37 CFR 1.129(b))	
1801	750	2801	375	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify) _____

*Reduced by Basic Filing Fee Paid SUBTOTAL (3) ($) 00.00

SUBMITTED BY				(Complete *if applicable*)	
Name (Print/Type)	Michael A. Glenn	*Registration No. (Attorney/Agent)*	30,176	*Telephone*	650-474-8400
Signature				*Date*	10/14/2003

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

This collection of information is required by 37 CFR 1.17 and 1.27. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, Washington, DC 20231.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

SEARCH ENHANCEMENT SYSTEM HAVING PERSONAL SEARCH PARAMETERS

FIELD OF THE INVENTION

The invention relates to the search and retrieval of information or content in a network environment. More particularly, the invention relates to the enhancement of search results, based upon information received from a user and/or an external source.

BACKGROUND OF THE INVENTION

Conventional search engines compare input search terms against metadata, to identify displayable results. Some search processes also allow for refined searching in input terms, against particular identified types of metadata. For example, during a search query at a search engine, a user may be able to enter either a word string, *e.g.* "serial number", or a corresponding abbreviation, *e.g.* "SN", to indicate that subsequent search terms should be applied against serial number metadata. Furthermore, some conventional search engines permit comparison of input search terms against full or partial text.

When applying conventional search technology, users typically obtain several pages of search results for any given search query, necessitating an extended period of review. For example, a common problem which is often encountered with conventional search queries is that the found set of matching sites or information sources, or documents is often too large, *i.e.* too broad, such as if too few search terms are entered within a search string, or if the search terms are too general. A user must often either manually browse through a large number of found sources to find relevant sites, or must perform a different search, typically having different terms and/or additional terms, in the hopes of more accurately finding the desired sites and/or information.

A similar problem that is also encountered with conventional search inquires is that the found set of matching sites, information sources, or documents is often too small, *i.e.* too narrow, such as if too many search terms are entered within a string, or if the search terms are too narrow in scope. A user is then typically required to perform one or more subsequent searches, typically having different terms and/or fewer terms, in the hopes of finding a larger found set of desired sites and information.

Several structures and methods have been described for the searching and retrieval of information in a network environment.

J. Breese and C. Kadie, *Methods and Apparatus for Tuning a Match Between Entities Having Attributes*, U.S. Patent No. 6,144,964, describe a matching of "entities having attributes, some of which have associated values. The values of the attributes may be adjusted based on number of entities that have values for a particular attribute so that the values decrease as the number increases. The attributes of the entities may be harmonized and provided with default values so that entities being matched have common attributes defined by the union of the attributes of the entities being matched. The attributes of the entities may be expanded and provided with default values so that the entities being matched have attributes that neither had originally. Match values may be normalized to provide a weight value which may be used to predict an attribute value of a new entity based on known attribute values of known entities. The weight values may be tuned such that relatively high weights are amplified and relatively low weights are suppressed."

B. Hazlehurst, S. Burke, and K. Nybakken, *Intelligent Query System for Automatically Indexing in a Database and Automatically Categorizing Users*, U.S. Patent No. 6,289,353 B1, describe a system which "develops multiple information spaces in which different types of real-world objects (e.g., documents, users, products) can be represented. Machine learning techniques are used to facilitate automated emergence of information spaces in which objects are represented as vectors of real numbers. The system then delivers information to users based upon similarity measures applied to the representation of the objects in these information spaces. The system simultaneously classifies documents, users, products, and other objects. Documents are managed by collators that act as classifiers of overlapping portions of the database of documents. Collators evolve to meet the demands for information delivery expressed by user feedback. Liaisons act on the behalf of users to elicit information from the population of collators. This information is then presented to users upon logging into the system via Internet or another communication channel. Mites handle incoming documents from multiple information sources (e.g., in-house editorial staff, third-party news feeds, large databases, World Wide Web spiders) and feed documents to those collators which provide a good fit for the new documents."

V. Berstis and H. Rodriguez, *Blocking Saves to Web Browser Cache Based on Content Rating*, U.S. Patent No. 6,510,458 B1, describe a process in which a "user sets preference parameters that filter web page contents from being stored in the cache. The preferences relate to the web page's contents and attributes. Before caching the

web page, the contents and attributes of the web page are filtered solely as a function of the web browser. Cache filters take a variety of forms, such as ratings filters, web page identifier filters, and key word filters, which scan accessed contents of a web page for user selected terms. The filtered web page is then blocked from entry in the browser's cache based on the filtering process. Conversely, a user sets preference parameters that filter web page contents to override the block from cache preferences. The browser responds by storing the filtered web pages that were previously designated as web pages not to be cached."

Other structures and methods for the searching and retrieval of information include: Y. Freund et al., An Efficient Boosting Algorithm for Combining Preferences, AT&T Labs, MIT Laboratory for Computer Science; J. Shavlik et al., Building Intelligent Agents for Web-Based Tasks: A Theory-Refinement Approach, University of Wisconsin-Madison; and J. Shavlik, et al., Intelligent Agents for Web-based Tasks: An Advice-Taking Approach, University of Wisconsin-Madison.

Several other structures and methods provide background information in regard to the search and retrieval of information, such as: European Patent Application No. EP 1 288 795 A1, Query systems; D. Reed, P. Heymann, S. Mushero, K. Jones, J. Oberlander, and D. Banay, *Computer-Based Communication System and Method Using Metadata Defining a Control Structure*, U.S. Patent No. 5,862,325; B. Hazlehurst, S. Burke, and K. Nybakken, *Intelligent Query System for Automatically Indexing Information in a Database and Automatically Categorizing Users*, U.S. Patent No. 5,974,412; J. Breese and C. Kadie, *Methods and Apparatus for Matching Entities and for Predicting an Attribute of an Entity Based on an Attribute Frequency Value*, U.S. Patent No. 6,018,738; D. Donoho, D. Hindawi, and L. Lippincott, *Method and Apparatus for Computed Relevance Messaging*, U.S. Patent No. 6,256,664 B1; D. Donoho, D. Hindawi, and L. Lippincott, *Inspector for Computed Relevance Messaging*, U.S. Patent No. 6,263,362 B1; A. Lang and D. Kosak, *Integrated Collaborative/Content-Based Filter Structure Employing Selectively Shared, Content-Based Profile Data to Evaluate Information Entities in a Massive Information Network*, U.S. Patent No. 6,308,175 B1, A. Lang and D. Kosak, *Collaborative/Adaptive Search Engine*, U.S. Patent No. 6,314,420 B1; J. Breese and C. Kadie, *Methods and Apparatus, Using Expansion Attributes Having Default, Values, for Matching Entities and Predicting an Attribute of an Entity*, U.S. Patent No. 6,345,264 B1; D. Reed, P. Heymann, S. Mushero, K. Jones, J. Oberlander, and D. Banay, *Computer-Based Communication System and Method Using Metadata Defining a Control-Structure*, U.S. Patent No. 6,345,288 B1; J. Breese and C. Kadie, *Method and Apparatus, Using Attribute Set Harmonization and Default Attribute Values, for Matching Entities and Predicting an Attribute of an Entity*, U.S.

Patent No. 6,353,813 B1; D. Donoho, D. Hindawi, and L. Lippincott, *Relevance Clause for Computed Relevance Messaging*, U.S. Patent No. 6,356,936 B1; E. Steeg, *Coincidence Detection Method, Products and Apparatus*, U.S. Patent No. 6,493,637 B1; System and Method for Data Collection, Evaluation Information Generation, And Presentation, U.S. Patent No. 6,539,392 B1; Baudisch, P.; The Profile Editor: Designing a Direct Manipulative Tool for Assembling Profiles; Institute for Integrated Information and Publication Systems IPSI, German National Research Center for Information Technology GMD, Germany; J. Budzik et al.; User Interactions with Everyday Applications as Context for Just-in-time Information Access; Intelligent Information Laboratory, Northwestern University; J. Budzik et al.; Watson: Anticipating and Contextualizing Information Needs; Northwestern University; E. Glover et al.; Improving Category Specific Web Search by Learning Query Modifications; NEC Research Institute, Princeton, NJ, EECS Department, University of Michigan, Ann Arbor, MI, Information Sciences and Technology, Pennsylvania State University; Pazzani et al., A Framework for Collaborative, Content-Based and Demographic Filtering; Department of Information and Computer Science, University of California, Irvine; T. Bauer et al.; Real Time User Context Modeling for Information Retrieval Agents, Computer Science Department, Indiana University; J. Shavlik et al.; An Instructable, Adaptive Interface for Discovery and Monitoring Information on the World-Wide Web; University of Wisconsin-Madison; J. Budzik et al.; Watson: An Infrastructure for Providing Task-Relevant, Just-In-Time Information; Department of Computer Science, Northwestern University; and D. Nahl, Ethnography Of Novices' First Use Of Web Search Engines: Affective Control In Cognitive Processing; Internet Reference Services Quarterly, vol. 3, no. 2, p. 51-72, 1998.

It would be advantageous to provide a system and an associated method which provides an enhancement to a search system, wherein the results from the search engine are refined or reorganized, based upon information from an identified secondary source. The development of such a search enhancement system would constitute a major technological advance.

It would also be advantageous to provide a system and an associated method which provides an enhancement to a search system, wherein information from an identified secondary source is integrated with a search query, such that results from the search engine are refined or organized, based upon the information from the identified secondary source. The development of such a search enhancement system would constitute a major technological advance.

In addition to search parameters which may be unique to a particular search, there are often parameters that are commonly relevant for a plurality of searches, such as relating to personalized information regarding the user or to similarities between the subject matter of a search. A user is often required to repeatedly input such parameters, along with other parameters that are unique to search.

Several structures and methods have been described for the searching and sorting of information, based on relevance, personal information, or profiles.

J. Driscoll, *Method and System for Searching for Relevant Documents from a Text Database Collection, Using Statistical Ranking, Relevancy Feedback and Small Pieces of Text*, U.S. Patent No. 5,642,502, describes a search system and method "for retrieving relevant documents from a text data base collection comprised of patents, medical and legal documents, journals, news stories and the like. Each small piece of text within the documents such as a sentence, phrase and semantic unit in the data base is treated as a document. Natural language queries are used to search for relevant documents from the data base. A first search query creates a selected group of documents. Each word in both the search query and in the documents are given weighted values. Combining the weighted values creates similarity values for each document which are then ranked according to their relevant importance to the search query. A user reading and passing through this ranked list checks off which documents are relevant or not. Then the system automatically causes the original search query to be updated into a second search query which can include the same words, less words or different words than the first search query. Words in the second search query can have the same or different weights compared to the first search query. The system automatically searches the text data base and creates a second group of documents, which as a minimum does not include at least one of the documents found in the first group. The second group can also be comprised of additional documents not found in the first group. The ranking of documents in the second group is different than the first ranking such that the more relevant documents are found closer to the top of the list."

T. Gerace, *Method and Apparatus for Determining Behavioral Profile of a Computer User*, U.S. Patent No. 5,848,396, describes a computer network method and apparatus which "provides targeting of appropriate audience based on psychographic or behavioral profiles of end users. The psychographic profile is formed by recording computer activity and viewing habits of the end user. Content of categories of interest and display format in each category are revealed by the psychographic profile, based on user viewing of agate information. Using the profile (with or without additional user demographics), advertisements are displayed to appropriately selected users. Based

on regression analysis of recorded responses of a first set of users viewing the advertisements, the target user profile is refined. Viewing by and regression analysis of recorded responses of subsequent sets of users continually auto-targets and customizes ads for the optimal end user audience."

F. Herz, *System for Customized Electronic Identification of Desirable Objects*, U.S. Patent No. 6,029,195, describes "customized electronic identification of desirable objects, such as news articles, in an electronic media environment, and in particular to a system that automatically constructs both a "target profile" for each target object in the electronic media based, for example, on the frequency with which each word appears in an article relative to its overall frequency of use in all articles, as well as a "target profile interest summary" for each user, which target profile interest summary describes the user's interest level in various types of target objects. The system then evaluates the target profiles against the users' target profile interest summaries to generate a user-customized rank ordered listing of target objects most likely to be of interest to each user so that the user can select from among these potentially relevant target objects, which were automatically selected by this system from the plethora of target objects that are profiled on the electronic media. Users' target profile interest summaries can be used to efficiently organize the distribution of information in a large scale system consisting of many users interconnected by means of a communication network. Additionally, a cryptographically-based pseudonym proxy server is provided to ensure the privacy of a user's target profile interest summary, by giving the user control over the ability of third parties to access this summary and to identify or contact the user."

A. Lang and D. Kosak, *Collaborative/Adaptive Search Engine*, U.S. Patent No. 6,314,420 B1, describe a search engine system "for a portal site on the internet. The search engine system employs a regular search engine to make one-shot or demand searches for information entities which provide at least threshold matches to user queries. The search engine system also employs a collaborative/content-based filter to make continuing searches for information entities which match existing wire queries and are ranked and stored over time in user-accessible, system wires corresponding to the respective queries. A user feedback system provides collaborative feedback data for integration with content profile data in the operation of the collaborative/content-based filter. A query processor determines whether a demand search or a wire search is made for an input query."

D. Kravets, L. Chiriac, J. Esakov, and S. Wan, *Search Data Processor*, U.S. Patent No. 6,363,377 B1, describe a "tool to be used with a search engine for a information management system includes methods for refining, filtering, and organizing search

queries and search results. A query tuner in the tool allows a user to automatically reformulate a query in order to find a reasonable number of matching documents from the search engine by selectively modifying individual search terms to be weaker or stronger and concurrently requesting a plurality of searches, each with a respectively different modified query. The tool also uses a dynamic filter which employs a dynamic set of record tokens to restrict the results of an arbitrary search query to selectively include or exclude records which correspond to the set of record tokens. The tool also includes a results organizer which aids the user in understanding and visualizing a large number of matching documents returned in response to a search query by clustering like items returned from the search. The query tuner, dynamic filter and results organizer may be used individually or in conjunction. The searched information management system may be consolidated or distributed and may span a global information network such as the Internet."

P. Biffar, *Self-Learning and Self-Personalizing Knowledge Search Engine That Delivers Holistic Results*, U.S. Patent No. 6,397,212 B1, describes a search engine which "provides intelligent multi-dimensional searches, in which the search engine always presents a complete, holistic result, and in which the search engine presents knowledge (i.e. linked facts) and not just information (i.e. facts). The search engine is adaptive, such that the search results improve over time as the system learns about the user and develops a user profile. Thus, the search engine is self personalizing, i.e. it collects and analyzes the user history, and/or it has the user react to solutions and learns from such user reactions. The search engine generates profiles, e.g. it learns from all searches of all users and combines the user profiles and patterns of similar users. The search engine accepts direct user feedback to improve the next search iteration One feature of the invention is locking/unlocking, where a user may select specific attributes that are to remain locked while the search engine matches these locked attributes to all unlocked attributes. The user may also specify details about characteristics, provide and/or receive qualitative ratings of an overall result, and introduce additional criteria to the search strategy or select a search algorithm. Additionally, the system can be set up such that it does not require a keyboard and/or mouse interface, e.g. it can operate with a television remote control or other such human interface."

G. Cullis, *Personalized Search Methods*, U.S. Patent No. 6,539,377 B1, describes a "method of organizing information in which the search activity of previous users is monitored and such activity is used to organize articles for future users. Personal data about future users can be used to provide different article rankings depending on the search activity and personal data of the previous users."

Other structures and methods have been described which provide background information regarding the searching and sorting of information, based on relevance, personal information, or profiles, such as: J. Pitkow et al.; Personalized Search, Communications of the ACM, vol. 45, no. 9, p. 50-5, September 2002; J. McGowan et al., Who Do You Want To Be Today? Web Personae for Personalised Information Access; Adaptive Hypermedia and Adaptive Web-Based Systems. Second International Conference, AH 2002. Proceedings (Lecture Notes in Computer Science Vol. 2347), p. 514-17, 2002; S. Kalajdziski et al.; Intelligent Recommendation in Digital Library, Proceedings of the IASTED International Conference Intelligent Systems and Control, p. 408-12, ACTA Press, Anaheim, CA, USA, 2001; L. Kerschberg et al., A Semantic Taxonomy-Based Personalizable Meta-Search Agent, Proceedings of the Second International Conference on Web Information Systems Engineering, vol. 1, p. 41-50; IEEE Comput. Soc., Los Alamitos, CA, USA, 2002; C. Dichev, A Framework for Context-Driven Web Resource Discovery, Modeling and Using Context, Third International and Interdisciplinary Conference, Context 2001, Proceedings (Lecture Notes in Artificial Intelligence, vol. 2116), p. 433-6, Springer-Verlag, Berlin, Germany, 2001; X. Meng et al., Feasibility of Adding Filtering Process in Web Browser to Improve Web Search Accuracy, Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, PDPTA '2000, vol. 4, p. 1809-15, CSREA Press, Athens, GA, USA, 2000; K. Kim et al., Development of a Personalized Link-Based Search Engine Using Fuzzy Concept Network, Journal of KISS: Computing Practices, vol. 7, no. 3, p. 211-19, Korea Inf. Sci. Soc., June 2001; C. Yang et al., A Hybrid Document Clustering for a Web Agent, Journal of KISS: Software and Applications, vol. 28, no. 5, p. 422-30, Korea Inf. Sci. Soc., May 2001; K. Kim et al.; A Personalized Web Search Engine Using Fuzzy Concept Network with Link Structure; Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), vol. 1, p. 81-6; IEEE , Piscataway, NJ, USA; 2001; A. Scime et al.; WebSifter: An Ontology-Based Personalizable Search Agent for the Web; Proceedings 2000 Kyoto International Conference on Digital Libraries: Research and Practice, p. 203-10; IEEE Comput. Soc., Los Alamitos, CA, USA; 2000; Z. Wei-Feng et al., Personalizing Search Result Using Agent, Mini-Micro Systems, vol. 22, no. 6, p. 724-7, Mini-Micro Syst., China, June 2001; P. Chen et al., An Information Retrieval System Based on a User Profile, Journal of Systems and Software, vol. 54, no. 1, p. 3-8, Elsevier, September 30, 2000; X. Meng et al., Personalize Web Search Using Information On Client's Side, Fifth International Conference for Young Computer Scientists, ICYCS'99, Advances in Computer Science and Technology, vol. 2, p. 985-92; Int. Acad. Publishers, Beijing, China, 1999; P. Chen et al.; A Personalized Information Retrieval System; Computational Intelligence for Modelling, Control and Automation, Intelligent Image

Processing, Data Analysis and Information Retrieval (Concurrent Systems Engineering Series, vol. 56), p. 247-53, IOS Press, Amsterdam, Netherlands, 1999; S. Laine-Cruzel et al., Improving Information Retrieval by Combining User Profile and Document Segmentation, Information Processing & Management, vol. 32, no. 3, p. 305-15; Elsevier, May 1996; D. Boley et al., Document Categorization and Query Generation on the World Wide Web Using WebACE; Department of Computer Science and Engineering, University of Minnesota; and A. Pretschner, Ontology Based Personalized Search, Dipl.-Inform., RWTH Aachen, Germany, 1998.

Other documents provide background information regarding advancements in search engine structures and processes, such as: European Patent Application No. EP 1 072 982 A2, *Method and System for Similar Word Extraction And Document Retrieval*; European Patent Specification No. EP 1 095 326 B1, *A Search System and Method for Retrieval of Data, and the Use Thereof in a Search Engine*; European Patent Application No. EP 1 284 461 A1, *Meta-Document Management System With User Definable Personalities*; European Patent Application No. EP 1 288 795 A1, *Query systems*; A. Lang and D. Kosak, *System and Method Employing Individual User Content-Based Data and User Collaborative Feedback Data to Evaluate the Content of an Information Entity in a Large Information Communication Network*, U.S. Patent No. 5,983,214; A. Lang and D. Kosak, *Multi-Level Mindpool System Especially Adapted to Provide Collaborative Filter Data for a Large Scale Information Filtering System*, U.S. Patent No. 6,029,161; M. Tso, D. Romrell, And D. Gillespie, *System for Distributing Electronic Information to a Targeted Group of Users*, U.S. Patent No. 6,047,327; G. Culliss, *Personalized Search Methods*, U.S. Patent No. 6,182,068 B1; A. Lang and D. Kosak, *Integrated Collaborative/ Content-Based Filter Structure Employing Selectively Shared, Content-Based Profile Data to Evaluate Information Entities in a Massive Information Network*; U.S. Patent No. 6,308,175 B1; D. Chen, *Cooperative Topical Servers With Automatic Prefiltering and Routing*, U.S. Patent No. 6,349,307 B1; D. Judd, P. Gauthier, and J. Baldeschwieler, *Method and Apparatus for Retrieving Documents Based on Information other than Document Content*, U.S. Patent No. 6,360,215 B1; K. Risvik, Search *System and Method for Retrieval of Data, and the Use Thereof in a Search Engine*, U.S. Patent No. 6,377,945 B1; E. Marwell and R. Pines, *Personalized Assistance System and Method*, U.S. Patent No. 6,404,884 B1; A. Weissman and G. Elbaz, *Meaning-Based Information Organization and Retrieval*, U.S. Patent No. 6,453,315 B1; J. Lee, L. Morgenstern, M. Pedlaseck, E. Schonberg, and D. Wood, *System and Method for Collecting and Analyzing Information About Content Requested in a Network (World Wide Web) Environment*, U.S. Patent No. 6,466,970 B1; S. Edlund, M. Emens, R. Kraft, and P. Yim, *Labeling and Describing Search Queries for Reuse*, U.S. Patent No. 6,484,162 B1; J. Zhang and M. Ott,

Method and Apparatus for Active Information Discovery and Retrieval, U.S. Patent No. 6,498,795 B1; L. Nikolovska, J. Martino, and A. Camplin, *Search User Interface with Enhanced Accessibility and Ease-Of-Use Features Based on Visual Metaphors*, U.S. Patent No. 6,505,194 B1; M. Bowman-Amuah, *Piecemeal Retrieval in an Information Services Patterns Environment*, U.S. Patent No. 6,550,057 B1; Callan, J. et al.; Document Filtering with Inference Networks; Computer Science Department, University of Massachusetts; Goker, A.; Capturing Information Need by Learning User Context; School of Computer and Mathematical Sciences; The Robert Gordon University; Chen, L. et al.; WebMate: A Personal Agent for Browsing and Searching; The Robotics Institute, Carnegie Mellon Institute; September 30, 1997; Cooley, R. et al.; Web Mining: Information and Pattern Discovery on the World Wide Web; Department of Computer Science and Engineering, University of Minnesota; Simons, J.; Using a Semantic User Model to Filter the World Wide Web Proactively; Nijmegen Institute for Cognition and Information, University of Nijmegen, The Netherlands; Tanudjaja, F. et al.; Persona: A Contextualized and Personalized Web Search; Laboratory of Computer Science at MIT, Cambridge, MA; June 1, 2001; Yan, T. et al.; SIFT - A Tool for Wide-Area Information Dissemination; Department of Computer Science, Stanford University, February 16, 1995; Bianchi-Berthouze, N.; Mining Multimedia Subjective Feedback; Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies, vol. 19, no. 1, p. 43-59; Kluwer Academic Publishers; July 2002; Widyantoro, D.H. et al.; A fuzzy Ontology-Based Abstract Search Engine and Its User Studies; 10th IEEE International Conference on Fuzzy Systems. (Cat. No.01CH37297), vol. 2, p. 1291-4; IEEE, Piscataway, NJ, USA; 2001; Tanudjaja, F. et al.; Persona: A Contextualized and Personalized Web Search; Proceedings of the 35th Annual Hawaii International Conference on System Sciences, p. 1232-40; IEEE Comput. Soc, Los Alamitos, CA, USA; 2002; Widyantoro, D.H. et al.; Using Fuzzy Ontology for Query Refinement in a Personalized Abstract Search Engine; Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), vol. 1, p. 610-15; IEEE , Piscataway, NJ, USA; 2001; Ho, M. et al.; A GA-Based Dynamic Personalized Filtering for Internet Search Service on Multi-Search Engine; Canadian Conference on Electrical and Computer Engineering 2001, Conference Proceedings (Cat. No.01TH8555) vol. 1, p. 271-6; IEEE , Piscataway, NJ, USA; 2001; Pogaenik, M. et al.; Layered Agent System Architecture for Personalized Retrieval of Information from Internet; Signal Processing X Theories and Applications. Proceedings of EUSIPCO 2000. Tenth European Signal Processing Conference, vol. 1, p. 421-4; Tampere Univ. Technology, Tampere, Finland; 2000; Ho, M. et al.; An Agent-Based Personalized Search on a Multi-Search Engine Based on Internet Search Service; Intelligent Data Engineering and Automated - IDEAL 2000, Data Mining, Financial Engineering, and Intelligent Agents, Second

International Conference, Proceedings (Lecture Notes in Computer Science Vol.1983), p. 404-9; Springer-Verlag, Berlin, Germany; 2000; Wei-Feng, Z. et al.; Personalizing Search Result Using Agent; Mini-Micro Systems, vol. 22, no. 6, p. 724-7; Mini-Micro Syst., China; Overmeer, M.A.C.J.; My Personal Search Engine; Computer Networks, vol. 31, no. 21, p. 2271-9; Elsevier, November 10, 1999; Pretschner, A. et al.; Ontology Based Personalized Search; Proceedings 11th International Conference on Tools with Artificial Intelligence, p. 391-8; IEEE Comput. Soc., Los Alamitos, CA, USA, 1999; Lee, E.S. et al.; Agent-Based Support for Personalized Information with Web Search Engines; Design of Computing Systems: Cognitive Considerations. Proceedings of the Seventh International Conference on Human-Computer Interaction (HCI International '97), vol. 2, p. 783-6; Elsevier, Amsterdam, Netherlands, 1997; and Berger, F.C. et al., Personalized Search Support for Networked Document Retrieval Using Link Inference, Database and Expert Systems Applications. 7th International Conference, DEXA '96 Proceedings, p. 802-11, Springer-Verlag, Berlin, Germany, 1996.

It would be advantageous to provide a system and an associated method which provides an enhancement to a search system, wherein a user may specify one or more search parameters, and wherein the user-specified search parameters are integrated into a search query, based on the subject matter of that query. The development of such a search enhancement system would constitute a major technological advance.

As well, it would be advantageous to provide a system and an associated method which provides an enhancement to a search system, wherein a user may specify one or more search parameters, and wherein the user-specified search parameters are integrated into a search query, based on the subject matter of that query, in which the subject matter is either explicitly determined or is implicitly determined, based upon user input. The development of such a search enhancement system would constitute a further technological advance.

Furthermore, it would be advantageous to provide a system and an associated method which provides an enhancement to a search system, in which general search parameters are solicited from a user before a particularized search, and wherein the solicited search parameters are associated with the user and are available for automatic integration into future particularized searches initiated by the user. The development of such a search enhancement system would constitute a further technological advance.

In addition, it would be advantageous to provide a system and an associated method which provides an enhancement to a search system, in which general search parameters

are solicited from a user before or after a particularized search, and wherein the solicited search parameters are associated with user selectable editorial content, such as for delivery to the user and/or to other recipients, and are available for automatic integration into future particularized searches, such as initiated by the user or by other recipients that are associated with the user USR, such as a network of friends, family, peers, students, neighbors, people or entities within a zip code region, and/or business associates. The development of such a search enhancement system would constitute a further technological advance.

SUMMARY OF THE INVENTION

The system and method comprises enhancement of results for a search engine, wherein the results from the search engine are refined or reorganized, based upon information from an identified secondary source. The results obtained using a conventional search are compared against the identified secondary source, *e.g.* a ratings service, and are filtered and/or sorted appropriately. In some embodiments, identification of the secondary source, such as a ratings service comprising information which may supplement the subject of a search query, is based upon information entered by the user. In alternate embodiments, the secondary source is associated with a user, as part of general user-specified search parameters, wherein one or more parameters are consulted automatically for searches for appropriate subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic view of a system for customizing results received from a search engine, wherein the customization comprises refinement of the search results based upon information received from an external source;

Figure 2 is a schematic view of an alternate system for customizing results received from a search engine, wherein the customization comprises an organization of the search results based upon information received from an external source;

Figure 3 is a schematic diagram of user identification, *i.e.* selection, of an external source within a system for further acting upon results received from a search engine;

Figure 4 is a schematic view of a system for acting upon results received from a search engine implemented within an integrated application;

Figure 5 is a schematic view of an alternate modular system for customizing results received from a search engine implemented in conjunction with a conventional search engine;

Figure 6 is a functional block diagram of operation within a system for customizing results received from a search engine implemented within an integrated application;

Figure 7 is a functional block diagram of operation within an alternate modular system for customizing results received from a search engine implemented in conjunction with a conventional search engine;

Figure 8 is a flowchart showing a process for identification of one or more external sources, and for refining search results based upon information received from the identified sources;

Figure 9 is a schematic diagram of an enhanced primary search input screen;

Figure 10 is a schematic diagram of an enhanced search system source selection screen;

Figure 11 is a schematic diagram of a primary search result screen further comprising enhanced source solicitation control;

Figure 12 shows user specification of secondary search parameters;

Figure 13 is a schematic diagram of an enhanced search parameter and subject validities;

Figure 14 is a detailed schematic diagram of an exemplary enhanced search parameter and subject validities;

Figure 15 is a functional block diagram of an enhanced search system comprising personal search parameters; and

Figure 16 is a flow chart of an enhanced search process comprising personal search parameters.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Figure 1 is a schematic view of a system 10a for acting upon, *i.e.* enhancing or customizing 22 (FIG. 3), results 14 received from a search engine 12, wherein the

customization 22 comprises refinement 22a of the search results 14 based upon information 18 received 20 from an external source 16. Figure 2 is a schematic view of an alternate system 10b for customizing 22 results received from a search engine 14, wherein the customization 22 comprises an organization, *i.e.* ranking 22b, of the search results 14, based upon information 18 received 20 from an external source 16.

The search enhancement system 10 improves current search methodologies, by refining 22a and/or organizing 22b the results 14 of a search engine 12, in compliance with information 18 from one or more sources 16. In a typical system embodiment 10, a user USR (FIG. 3) selects 33 (FIG. 3) an information source 16 to be consulted by a search application 12, in the process of performing a search 24.

In some system embodiments 10, the results 14 of a search 24 are further refined 22a and/or organized 22b, based upon information 18 received from an external source 16. For example, in a user-initiated search 13 for lodging in Austin, Texas, a user specified source 16 may preferably comprise rating information 18 of lodgings, *e.g.* such as available through American Automobile Association, Inc. (AAA). The rating information 18, from the external source 16, *e.g.* AAA, is then used to refine 22a and/or organize 22b the results 14 of a general search 24 for any lodging that otherwise meets the search parameters 106,108, *e.g.* 108a,108b (FIG. 9, FIG. 10, FIG. 11) within a search query 13. Results from a conventional search engine 12 may therefore be refined 22a and/or reorganized 22b, based on data 18 independently maintained by a ratings service 16.

In alternate system embodiments 10, information 18 received from an external source 16 is integrated within a search query 13, such that the search 24 is enhanced by the information 18, whereby the results 14 of the search 24 may be inherently refined 22a and/or organized 22b as a function of the information 18 received from an external source 16. For example, information 18 received from a user-selected source 16 may be consulted when performing a search 24, such that a set of one or more search results 14 complies with a search query 13 comprising both search parameters 108,106 entered by a user USR, in addition to meeting parameters imposed by the information 18 received from the external source 16.

For example, in a similar user-initiated search 13 for lodging in Austin, Texas, a user specified source 16 which comprises rating information 18 of lodgings can alternately be included in the search 24, *i.e.* to refine or organize the search results 14, whereby the results of the general search 24 meet the user-specified search parameters 108,106, and also comply with rating information 18 provided by a selected ratings service 16.

In some system embodiments 10, the external information 18 can be combined with other external information 18. In the above example, in a user-initiated search 13a for lodging in Austin, Texas, wherein a user specified source 16 comprises rating information 18 of lodgings, the rating information 18 can be combined with policy information 18, such as to further refine or organize the search results 14, to lodging which is approved by a secondary external source 16, *e.g.* an accounting department, associated with the user USR, to be within a specified cost per diem amount.

Supplementary external information 18 may also correspond to people or entities which are associated with the user USR, *e.g.* such as a network of friends, family, peers, students, neighbors, people or entities with a zip code region, and/or business associates. For example a user USR may be interested in the enhanced results 40 based on ranking information 18 from:

- Expert entities, *e.g.* Zagats, AAA, or a movie critic;
- Celebrities, *e.g.* Michael Jordan, John Cusack, or Sarah Michelle Gellar; or
- People similar to the user USR, *i.e.* "people like me", such as local people of a similar age and/or education level, immediate friends or friends of friends.

In system embodiments 10 in which information 18 from an external source 16 is combined with information 18 from one or more other external sources 16, the enhanced search results 40 preferably yield a composite refinement or ranking 22, for a user USR. For example, in a search for local services or people, *e.g.* a roofing contractor, a user USR can combine a general search for local contractors within a desired area, *e.g.* within a city, county, or zip code, and can rank the results based on information from a ratings service 18, and/or with information 18 from other external sources 16, such as review information from people, *e.g.* such as neighbors who have used roofing contractors, other users USR, recipients RCP (FIG. 3), and/or experts.

Figure 3 is a schematic diagram 30 of user identification, *i.e.* selection 33 of an external source 16 within a system 10 for further acting 22 upon results 14 received from a search engine 12. A user USR typically interacts with the system 10 through a terminal 32, such as a personal computer, laptop computer, or other networked device, such as a personal digital assistant, a network enabled portable phone, or other wired or wireless device.

Through user identifier input 33, the system 10 determines, *i.e.* identifies 34 one or more selected external sources 16, either directly, *e.g.* through explicit entry 33 of the identity of a source 16, or indirectly, *e.g.* through an implicit determination of an identity of a

source 16, such as through the determination of subject matter of a search query 13, and a determination of one or more sources 16 that have information 18 which pertains to the determined subject matter.

The system 10 retrieves information 38 from an external source 16, such as though a an information query 36. Based upon data received 38 from a selected source 16, the system 10 returns 41 enhanced search results 40, *e.g.* such as by returning 41a to the user USR, through the terminal 32, and/or by delivering results 41b to one or more recipients RCP, such as through terminals 32.

In some system embodiments 10, recipients RCP are explicitly determined by the user USR. In other system embodiments 10, recipients RCP may be inferentially determined by the user USR, such as comprising one or more recipients RCP that are associated with the user USR, *e.g.* such as a network of friends, family, peers, students, neighbors, people or entities with a zip code region, and/or business associates. For example, a query 13 from a user USR regarding museums in Paris, France may be refined 22a or ranked 22b, and then may be forwarded 40b to recipients RCP, such as to recipients RCP that choose 86 (FIG. 7) to receive 41b, or are chosen to receive 41b, the information.

In some system embodiments 10, recipients RCP may be inferentially determined by the search parameters or search results, such as comprising a one or more recipients RCP for which editorially ranked 22a or sorted 22b content 40 is determined to be valid. For example, a query 13 from a user USR regarding museums in Paris may be refined 22a or ranked 22b, and then may be forwarded 40b to recipients RCP, such as to recipients RCP that have expressed interest in art, and/or France, such as through recipient input 86 (FIG. 7).

The search enhancement system 10 may therefore be preferably used to provide editorially refined 22a or ranked 22b results, as a result of a user selectable editorial search 13,22, for delivery to the user USR and/or to one or more recipients RCP.

As discussed above, in some system embodiments 10, the results 14 of a general search query 13 are acted upon 22 by the information 18 from the selected source 16, while in alternate system embodiments 10, the information 18 from the selected source 16 is integrated within the query 13, to provide search results 14 that correspond to both the general search parameters 108,106, as well as to the supplemental information 18 from the selected source 16.

User Source Selection and Delivery of Enhanced Search Query Results. Figure 4 is a schematic view 42 of a system 10c for acting upon results 14 received from a search engine 12 implemented within an integrated search structure 44, *i.e.* an enhanced search engine 44. Figure 5 is a schematic view 60 of an alternate modular system 10d for customizing results 14 received from a search engine 12 implemented in conjunction with a modular application component 46.

As seen in Figure 4, system source identification 34 and information processing 22 in a system 10c are readily implemented within an application component 46 which is integrated 44 with a search engine 12. When a user USR initiates 52 a search 24, the application component 46 shown in Figure 4 provides source identification 34, based upon user selection 33, and either acts 22 upon the results 14 of a general query 13 that meets user search criteria 106,108 (FIG. 9), or alternately modifies the query 13, based upon information 18, *e.g.* 18b, from one or more selected sources 16, *e.g.* 16b.

As seen in Figure 5, the system source identification 34 and information processing 22 in a system 10d are implemented within a modular, *i.e.* distinct, application component 46 which is in associated with a discrete search engine 12. For example, an application component 46 for source identification 34 and information processing 22 may operate as a separate component in relation with an existing search engine 12, whereby information 18 from a selected source 16 is either integrated into a search query 13, *e.g.* such as through additional Boolean string elements, or is used to process 22 the results 14 of a query 13.

System Operation. Figure 6 is a functional block diagram 70 of operation within a system 10e for acting upon results 14 received from a search engine 12 implemented within an integrated search structure 44. Figure 7 is a functional block diagram 82 of operation within an alternate modular system 10f for customizing results 14 received from a search engine 12 implemented in conjunction with a modular application component 46.

Figure 8 is a flowchart showing an exemplary process 90 for an enhanced search system 10, comprising identification 33 of one or more external sources 16, and the enhancement 22 of search results 14 based upon information 18 received from one or more identified sources 18.

As seen in Figure 6 and Figure 7, a user USR interacts 71 with the enhanced search system 10, typically between a user terminal 32 and an application module 46. Typical interactions 71 between a user terminal 32 and the application module 46 comprise

search initiation 52, the return of standard, *i.e.* non-enhanced, search results 73, source prompts 74, source identification inputs 33, and/or the return 41a (FIG. 3) of enhanced results 40.

In some embodiments of the enhanced search system 10, as seen in Figure 7, one or more recipients RCP at terminals 32 may also interact 86 with the application module 46. Typical interactions 86 between a recipient terminal 32 and the application module 46 comprise establishment of relationships 87, *e.g.* such as between users USR and recipients RCP, and/or an input of preferences or interest in the receipt 41b of enhanced system content 40.

The application module 46 has access 72, *e.g.* 72a-72n, to one or more sources 16, *e.g.* 16a-16n, having associated data 18, *e.g.* 18a-18n. The sources 16a-16n are typically accessible across a network, *e.g.* such as but not limited to the Internet. In some system applications 10, the associated data 18 is sent 72 to the application module 46 before source identification 33 from a user USR, such that information 18 associated with a source 16 is internally available within the application module 46. In other system applications, the associated data 18 is sent 72 to the application module 46 upon source identification 33 from a user USR, wherein information associated with a source 16 is typically queried 36 (FIG. 5) and retrieved 38 by the application module 46.

As seen in Figure 6 and Figure 7, a search engine 12 is associated with application module 46. The search engine 12 has access 76, *e.g.* 76a-76k, to one or more external sites, sources or documents 78, *e.g.* 78a-78k, having associated content 80, *e.g.* 80a-80k. The search engine 12 typically retrieves information content 80 that corresponds to a search query 13.

In the exemplary process 90 shown in Figure 8, when a user USR initiates a search query process 52, the application module 46 typically solicits 74 the identification 33 of one or more sources 16 comprising data 18 which may be used to enhance the value of a search 24, *e.g.* to improve the quality and/or ordering of search results 14. In some system embodiments 10, a source solicitation 74 comprises a choice of one or more selectable sources 16, typically comprising sources 16 that are either explicitly available to the user USR, or are implicitly determined, *e.g.* such as travel related sources 16 if a search query 13 comprises one or more search parameters 108 within a search string 106 (FIG. 9 to FIG. 11), which indicate that the user USR is searching for lodging or travel accommodations.

Upon a receipt 33 of source identification from a user USR, the application module 46 typically sends a data query 36 to any identified sources 16, if the available data 18 from an identified source 16 is not yet available. Upon a data query 36, the data 18 is sent from an identified source 16, either to be included in a search query 13, or to be used in the processing 22 of search results 14. The application module 46 produces 22 the enhanced results 40, which are then sent 41a to the user terminal 32, and/or sent 41b to recipients RCP.

Conventional search engines typically compare input search terms 108 against content or metadata 80, to identify displayable results. Some search processes also allow for refined searching in input terms 108, against particular identified types of content or metadata 80. For example, when performing a patent search, a user USR is able to enter "SN" to indicate that subsequent search terms 108 should be applied against serial number metadata. Furthermore, some of the conventional search engines permit comparison of input search terms against full or partial text.

When applying conventional search technology, users USR typically obtain several pages of search results for any given search query, necessitating an extended period of review. For example, a common problem which is often encountered with conventional search queries is that the found set 138 (FIG. 11) of matching sites or information sources 78 is often too large, *e.g.* such as if too few search terms 108 (FIG. 11) are entered within a string 106 (FIG. 11), or if the search terms 108 are too general. A user USR must often either manually browse through a large number of found content 80 to find relevant sites 78, or must perform a different search 24, typically having different terms 108 and/or additional terms 108, in the hopes of more accurately finding the desired sites 78 and information 80.

A similar problem that is also encountered with conventional search inquires is that the found set 138 of matching sites or information sources 78 is often too small, *e.g.* such as if too many search terms 108 are entered within a string 106, or if the search terms 108 are too narrow in scope. A user USR then is typically required to perform another search, typically having different terms 108 and/or less terms 108, in the hopes of finding a larger found set 138 of desired sites 78 and information 80.

Figure 9 is a schematic view 100 of an enhanced primary search user entry screen 102 for a search engine 12, in which a user USR may preferably select one or more sources 16 at the same time as primary search parameters 108, *e.g.* 108a,-108b, are entered. As seen in Figure 9, an input screen 102 comprises a parameter input window 104, wherein a user USR can input one or more search parameters 108, *e.g.* 108a,108b,

such as within a Boolean string format 106. If a primary search 24 is desired, *i.e.* without a selection of secondary sources 16, a search control 110 may preferably be activated, such that the primary search 24 is based only upon the primary search parameters 108, *e.g.* 108a,108b, such as within a search string 106.

The enhanced primary search user entry screen 102 shown in Figure 9 also comprises secondary source selection 112, comprising one or more subject sources 116a-116j within one or more search subject groups 114a-114k. For example, within a travel subject group 114a, one or more travel subject sources 116a-116j are selectable by the user USR, such that corresponding sources 16 are referenced in association with a search 24 corresponding to the primary search parameters 108. A search subject group 114 may comprise any of a wide variety of selectable subjects 114, such as but not limited to travel, shopping, business, technology, or personal sources 114. The displayed selection of subjects 114 and subject sources may reflect general subject areas, *i.e.* for general user audiences, or may alternately reflect more specialized professional or personal interests, such as internet-based opinion, review, and/or ratings sources 16.

The search subject groups 114a-114k shown in Figure 9 also comprise corresponding options control 118a-118k, such as to add or subtract desired source choices 116, and/or to select options based upon a source 16, such as to select a desired rating level of lodging, *e.g.* 4 stars, based upon a selected source 16.

As seen in Figure 9, the source selectors 116 allow selection of one or more secondary sources 16. In some system embodiments 10, information 18 associated with a source selection 116 accompanies the general search parameters 106,108 during a search 24. In alternate system embodiments 10, information 18 associated with a source selection 116 is used to enhance 22 the results 20 of a general search 24 that is based upon the parameters 106,108.

In some system embodiments, preliminary source selectors 116 comprise selectable choices of external sources 16, such as ranking sources 16, such as a ratings service 16 for restaurants, *e.g.* zagats.com, available through Zagat Survey LLC, of New York, NY. When searching for a restaurant using a search engine 12, a user USR may find it helpful to filter 22a or sort 22b results based on a rating from a ratings service 16. More specifically, the user USR can search for and display only restaurants with a Zagats rating higher than two, or the user USR may search for all restaurants meeting a specified criteria, and sort all restaurant hits based on the Zagats rating. Similarly, when searching for lodging using a search engine 12, a user USR may find it helpful to filter 22a or sort

22b results based on a rating from a travel related ratings service 16, *e.g.* such as ratings provided by American Automobile Association, Inc. (AAA).

Some system embodiments 10 allow express entry by the user USR of information in a search string, such as within the primary input window 104, to enable identification of such a source. In an alternative implementation, a source, *e.g.* Zagats or AAA, and an appropriate rating for a source 16, *e.g.* a Zagats rating equal to 2, may be associated with a user USR as a part of general user-specified search parameters 148, *e.g.* 148a (FIG. 12), in which one or more user-specified parameters 148 may be consulted automatically for searches of appropriate subject matter. In alternate system embodiments 10, the preliminary source selectors 116 comprise selectable choices of user-defined sources 16, such as to include one or more-user-selected parameters 148 (FIG. 12).

In some embodiments, the selected 116 external sources 16 are included along with the primary search parameters 108. In alternate embodiments, selected external sources 16 are referenced to refine 22a and/or reorganize 22b results 14 of a search 24 based upon the primary search parameters 108 within the search string 106, *i.e.* the search engine 12 conducts a search 24, based upon parameters 106,108, wherein the results of a search 24 typically include all sites or sources 78 which meet the search parameter set.

Figure 10 is a schematic diagram 120 of an enhanced search system source selection screen 122. As described above, a user USR may initiate 52 a search 24, such as based upon one or more search parameters 108 within a search string 106. In some embodiments of an enhanced search system 10, a solicitation, *i.e.* source prompt 74 may be made, such that a user USR can select one or more sources 16 which can be used to refine 22a and/or organize 22b the results of a search 24.

As seen in Figure 10, the enhanced search system source selection screen 120 preferably displays entered search parameters 106,108, and may also display a search subject 124, *e.g.* such as but not limited to travel, cuisine, technical, biographical, cultural, or business subjects. The search subject 124 may be determined either explicitly or implicitly from the search parameters 106,108, or may otherwise be selected or determined, such as by user subject selection control 125. The system source selection screen 122 shown in Figure 10 also comprises a secondary source selection 112, from which one or more subject selections 116a-116j may be made by a user USR. The exemplary subject selection option 114a shown in Figure 10 corresponds to one or more travel related source selections 116, *e.g.* 116a-116j, based on the determined

subject 124. The enhanced search system source selection screen 120 also comprises a refine results control 126 and an enhanced search control 128, whereby a user USR can control search refinement or organization 22, based upon source selections 116.

The secondary source selection 112, as shown in Figure 10, may alternately correspond to people or entities which are associated with the user USR, *e.g.* such as a network of friends, family, peers, students, neighbors, people or entities with a zip code region, and/or business associates. For example, the secondary source selection 112 may provide system access to external information or input 18 from one or more recipients RCP that are associated with the user USR. In another example, the secondary source selection 112 may provide system access to external information or input 18 from:

- Expert entities, *e.g.* Zagats, AAA, or a movie critic;
- Celebrities, *e.g.* Michael Jordan, John Cusack, or Sarah Michelle Gellar; and/or
- People similar to the user USR, *i.e.* "people like me", such as local people of a similar age and/or education level, immediate friends or friends of friends.

Figure 11 is a schematic diagram 130 of a primary search result screen 132 further comprising enhanced source solicitation control 126,128. As described above, a search 24 which comprises only primary search parameters 108, *e.g.* 108,108b, such as within a search string 106, may often yield a large found set 138 of results 134, *e.g.* 134a-134j. In some embodiments of the enhanced search system 10, such as integrated with a conventional search engine 12, an enhancement of a search 24 may comprise processing 22, *e.g.* refinement 22a and/or organization 22b, of a found set 138 from a search 24. The enhanced source solicitation control 126,128 shown in the primary search result screen 132 allows a user USR to operate 22 on the results of a search 24, such as by navigation to a enhanced search system source selection screen 120, as seen in Figure 10.

The source solicitation screen 132 may alternately comprise a selection sources 16 which are implicitly determined, such as based on entered search parameters 108. For example, in a user USR entered search string 106 which includes a term lodging, a choice of travel specific sources 16 may be provided for the user USR, such as to refine a search based upon ratings from one or more travel-related sources 16.

The search enhancement system 10 is readily implemented to provide a great value for a user USR, in which information from one or more secondary sources 16 can be explicitly or implicitly accessed and integrated to refine or organize the results of a search. The search enhancement system 10 improves current search methodologies, since a user USR can specify one or more user-selected information sources to be

consulted by a search application module 46 when performing a search. Results 14 from a conventional search engine 12 may be refined or re-organized based on data independently maintained by a selected source 16, such as a ratings service 16.

In either case, the results obtained using a conventional search may be compared against the identified source ratings service 16 and filtered 22a and/or sorted 22b appropriately. System functions may be performed by an integrated search engine 44, or alternatively, by an application module 44 associated with a search engine 12, such that no modification is necessary to the conventional search engine 12.

Since conventional search engines 12 allow only explicit entry of search terms 106, 108, such as within a search string interface 104, a user USR is typically required to repeat searches using a plurality of combinations of search parameters 108 and search strings 106, in order to receive an acceptable quality and quantity of search results, *i.e.* hits.

The enhanced search system 10 provides structures and associated processes which allow a user USR to enhance either the search or the results of a search, based upon information from one or more selected sources. A wide variety of selectable sources 16, from which supplementary information 18 is accessed, may be used, such as external services 16, e.g. ratings services, or user-specified sources, e.g. such as user-defined ratings or search parameters.

Enhanced Search System having Personal Search Parameters. Some preferred embodiments of the enhanced search system 10, such as 10g (FIG. 15), comprise the selection 33 of one or more user-specified search parameters 148, *e.g.* 148a (FIG. 12).

Some embodiments of the enhanced search system 10 having personal search parameters 148 comprise a solicitation of general search parameters 148 from a user USR, before a particularized search 24 is initiated 52. The solicited search parameters 148 are thereafter associated with the user USR, such that parameters 148 are available for automatic integration into future particularized searches initiated 52 by the user USR.

Figure 12 is a schematic diagram 140 which shows user specification 146a-146p of one or more secondary search parameters 148a-148p, such as through a user interface 142. The specified parameters 148a-148p are typically stored 150, such as at one or more locations, which can be located at a wide variety of locations within an enhanced search system 10, such as within a user terminal 32, at an enhanced search application module

46, in combination with a search engine 12, or at one or more locations throughout the system, such as at a service provider or a personal web site.

The generalized search parameters 148 may pertain to a variety of different subject matters 168 (FIG. 13), and represent information useful in enhancing a search for a user USR and/or other recipients RCP, such as by filtering, further filtering, or sorting search results obtained when performing searches 54. For example, generalized search parameters 148 may include the user's address and health insurance carrier, such that future particularized searches for medical care providers may be automatically refined or organized based on proximity and eligibility.

In a conventional search environment, a user USR must often enter detailed personal search parameters within a search string, if personal criteria are to be considered at the time of a search. Therefore, a user USR is often required to understand the search engine, and to remember the parameters at the time of the search.

The use of user specified generalized search parameters 148 readily provides an improved search environment, since a user USR is not required to manually manipulate a conventional search engine 12, through the entry of detailed search parameters 108, to consider personal criteria at the time of a particularized search. As well, a user USR is not required to understand the detailed string parameter format 106 of a search engine 12, nor is a user USR required to remember and enter personalized parameters 148 at the time of a search.

The use of generalized search parameters 148 relieves the user USR of the burden of sifting through pages of search results that are not relevant or customized to their needs, for instance, medical care providers the are not proximate to their home or eligible under their insurance in the example above.

Figure 13 is a schematic diagram 160 of an user-specified search parameter 148 and subject validities, *i.e.* rankings 170. A search parameter 148 typically comprises a parameter value 166 entered by a user USR, such as within a parameter value entry window 164 within a user interface 142 (FIG. 12).

A search parameter 148 may also preferably comprise an entered or determined ranking 170 for one or more subjects 168, *e.g.* 168a-168s, such that an applicability or validity of the parameter 148 can be explicitly or implicitly determined, *i.e.* such as in a determination of inclusion within a an enhanced search 24, or within search result refinement 22a and or sorting 22b. The search parameter 148 shown in Figure 13

comprises applicability rankings 170, *e.g.* 170a-170t for at least one subject matter 168. An exemplary quantized ranking 170 may be ranked as Yes or No, one or more divisions between 0% to 100%, a numeric value of 1 to 5, or another rating value scale 170. A ranking 170 is preferably associated with each of the generalized search parameters 148 for a particular subject matter 168, so that results satisfying several of the criteria may be appropriately sorted.

Based upon the determined subject matter 168 of a search query 13, a user specified parameter 148 having a rankings 170 which is determined to be applicable to the search may either be used in conjunction with primary search terms, *i.e.* to further limit search results, or may be used to sort the results of a query based upon primary search criteria, wherein the sort is based upon the applicability ranking 170 of one or more generalized search parameters 148.

Figure 14 is a detailed schematic diagram 172 of an exemplary enhanced search parameter 148, having a parameter value 166 of "95103" for a home address zip code, along with subject validities, *i.e.* rankings 170 for a plurality of subjects 168, such as local services 168a, online shopping 168b, and health care 168s. As seen in Figure 14, a ranking 170t of 100 percent applicability is associated with local services 168a, such that a search for a local service may preferably include the home address parameter 166 of the user USR.

To enable automatic association of appropriate parameters with future particularized searches, generalized user-specified search terms are preferably stored or associated with a label or type. In the example above, for instance, search terms may be stored as follows:

Health Insurance Carrier:	Kaiser
User's Home Address:	Street Address City, State, Zip
User's Work Address:	Street Address City, State, Zip

Thereafter, at the time of a particularized search by the user USR, the subject matter of the search 24 is identified or through explicit entry by the user USR, the relevant types of generalized search parameters 148 are identified based on the subject matter 124 of the search 24. Again using the example above, if the system 10 determines that a user

USR seeks to search for a medical care provider, the generalized search parameters 148 of location and medical insurance provider may preferably be identified as relevant to this particularized search 24. Where available, user-specified information 18 related to those parameters are extracted from the general search parameters 148 associated with the user USR. Parameters 148 that are determined to be relevant can be used in any of a variety of ways, such as to return 22a or organize 22b better search results, or to perform a search 24 using the available subset of optimal parameters 148.

In some system embodiments, the system 10 solicits the user USR to provide any missing parameters at the time of the particularized search. For example, if the user-specified parameters do not include the insurance carrier of a user USR, a exemplary search for a medical care provider may be performed by supplementing the user's input with location alone, or the user USR may be asked for their medical insurance provider at the time of the particularized search.

Figure 15 is a functional block diagram of an enhanced search system 10g comprising personal search parameters 148. Figure 16 is a flowchart showing an exemplary process 200 for an enhanced search system 10g, comprising personal search parameters 148, and the refinement 22 of search results 14 based upon one or more personal search parameters 148 which are determined to be relevant to the search.

As seen in Figure 15, a user USR interacts 71 with an application module 46, typically between a user interface 142 at a user terminal 32 and the application module 46. Typical interactions 71 between a user terminal 32 and the application module 46 comprise input and definition 182 of parameters 148, search initiation 52, the return of standard, *i.e.* non-enhanced, search results 73, parameter prompts 184, source identification inputs 33, and/or the return 41a,41b of enhanced results 40.

Some embodiments of the enhanced search system 10, such as the enhanced search system 10g shown in Figure 15, preferably provide user selected editorial searches, *e.g.* such as editorially ranked content, *i.e.* editorial commentary, corresponding to user input, which can then be returned 40a to the user USR, or sent 40b to recipients RCP, such as to send to selected peer recipients RCP of the user USR. In some system embodiments, the recipients RCP comprise one or members of a network of people associated with a user USR, *e.g.* such as node recipients RCP in a network of people with similar interests, or a chain of friends, *e.g.* such as established through the Internet, *e.g.* friendster.com.

The enhanced search system 10g shown in Figure 15 may alternately provide enhanced results 40 which are filtered in part, *i.e.* refined 22a and/or organized 22b, based upon information 18 received from an external source 16, such as from the most popular people and/or places in the user's personal network.

The enhanced search system 10g and an associated method 200 therefore provide an enhancement to a search system, in which general search parameters are solicited from a user USR, either before or after a particularized search, such as through button selection within a user interface 142. In some system embodiments, the solicited search parameters are preferably associated with user selectable editorial content, such as for delivery to the user USR and/or to other recipients RCP, and are available for automatic integration into future particularized searches, such as initiated by the user USR or by other recipients RCP that are associated with the user USR, such as a network of friends, family, peers, students, neighbors, people or entities with a zip code region, and/or business associates.

In the exemplary process 200 shown in Figure 16, a user USR is preferably able to initially store 202 user-specified search parameters 148. When a user USR initiates a search query process 52, an identification of the subject matter may also be identified 204, such as through explicit entry, *e.g.* through subject selection 125 (FIG. 10), or through an implicit determination 206 at the application module 46, *e.g.* based upon the entered search parameters 108 and parameter string 106. If the subject matter of the query is not 210 determined, the application module 46 solicits 212 and receives 214 the subject matter from the user USR. If the subject matter of the query is 216 determined, the system 10 selects 212 one or more user-specified search parameters 148 based on the subject matter, either for refinement 22a, sorting 22b of search results 14, or for integration 210 of the selected user-specified search parameters 148 with a query 13.

Upon receipt 33 of source identification from a user USR, the application module 46 typically sends a data query 36 to any identified sources 16, if the available data 18 from an identified source 16 is not yet available. Upon a data query 36, the data 18 is sent from an identified source 16, either to be included in a search query 13, or to be used in the processing 22 of search results 14. The application module 46 produces 22 the enhanced results 40, which are then sent 41, such as by returning 41a the enhanced results 40 to the user terminal 32, and/or by sending 41b the enhanced results 40 to one or more recipients RCP.

The enhanced search system 10g seen in Figure 15 provides storage of a set of user-specified search parameters, and automatically integrates selected ones of the stored search parameters into a search query 13, based on the subject matter of that query 13. The subject matter of the query may be explicitly indicated by the user USR, or may be inferentially determined, based on user input.

In some embodiments of the enhanced search system 10g, general search parameters are typically solicited 184 from a user USR before a particularized search 13 is initiated. The solicited search parameters 148 are associated with the user USR, whereby the parameters 148 are available for automatic integration into future particularized searches initiated by the user USR.

The storage 150 of personal parameters may be located at a wide variety of locations within the system 10g, such as within a file stored on the user's computer 132 (FIG. 6). The general search parameters 148 may also be associated with a user's roaming profile, passport, or init packet, such that the system 10g may readily access personal parameters 148 for an identified user USR.

The determination and maintenance of generalized search parameters 148 can be provided by a wide variety of entities, such as but not limited to the application module 46, an independent enhancement module that works in conjunction with a conventional search engine 12 or as an application, utility, or application plug-in within a user terminal 32.

The enhanced system 10g is readily adapted to a wide variety of network structures, such as within an integrated search engine structure 44 (FIG. 6), or within an application module 46 which is retrofit to an existing, *i.e.* conventional, search engine 12. As well, the appropriate subject matter of a particularized search can be identified either at the application module 46, or even at a user terminal 32, e.g. such as for local storage or parameters 148 and refinement 22a and/or sorting 22b of search results 14.

In some system embodiments 10g, the user-specified search parameters are integrated into a search string and are therefore used to produce the number of resulting hits. In alternate embodiments of the enhanced search system 10g, user-specified parameters 148 are applied to the results of a search that is performed without their integration, such as within a system 10g which is retrofit to an existing search engine 12.

In some enhanced search system embodiments 10, at the time of a particularized search, an interface is presented to the user USR, which allows a customized search query, and enables the user USR to modify out-dated or unwanted search parameters.

Conventional search engines 12 typically operate in an objective manner, based upon search parameters 108 within a parameter string 106 input at the time of a search. As well, the search results of a conventional search engine are often sorted as a function of commercial or popular parameters.

The enhanced search system 10g and personal search parameters 148 allow the results of a search engine 12 to be enhanced 22, such as through refinement 22a and/or sorting 22b, to reflect the desired or intended focus of the user USR. Furthermore, the determined subject matter of the search allows the enhanced results 40 to reflect more subjective results than are provided in an objective search engine alone.

Although the enhanced search system and methods of use are described herein in connection with a user terminal, the apparatus and techniques can be implemented for a wide variety of electronic devices and systems, such as personal computers, mobile devices, and other microprocessor-based devices, such as portable digital assistants or network enabled cell phones, or any combination thereof, as desired.

As well, while the enhanced search system and methods of use are described herein in connection with interaction between a user terminal and an application module and search engine across a network such as the Internet, the enhanced search system and methods of use can be implemented for a wide variety of electronic devices and networks or any combination thereof, as desired.

Accordingly, although the invention has been described in detail with reference to a particular preferred embodiment, persons possessing ordinary skill in the art to which this invention pertains will appreciate that various modifications and enhancements may be made without departing from the spirit and scope of the claims that follow.

CLAIMS

What is claimed is:

1. A process, comprising the steps of:

storing a set of search parameters specified by a user; and

integrating selected ones of the stored search parameters into a search query based upon a subject matter of the search query.

2. The process of Claim 1, further comprising the step of:

returning results of the search query to the user.

3. The process of Claim 1, further comprising the step of:

sending results of the search query to a recipient.

4. The process of Claim 3, wherein the recipient is selected by the user.

5. The process of Claim 3, wherein the recipient is determined based on the results of the search query.

6. The process of Claim 3, wherein the recipient is determined based on information input by the user.

7. The process of Claim 1, further comprising the steps of:

receiving information from a recipient; and

selectively sending results of the search query to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user.

8. The process of Claim 7, wherein the received information comprises an interest in the results of the search query by the recipient.

9. The process of Claim 1, wherein the subject matter is explicitly indicated by the user.

10. The process of Claim 1, wherein the subject matter is inferentially determined based upon user input.

11. The process of Claim 1, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, further comprising the step of:

determining the selected ones of the stored search parameters based upon the selectable applicability.

12. A process, comprising the steps of:

receiving at least one search parameter specified by a user;

receiving a search query from the user;

determining the subject matter of the search query;

selecting one or more user-specified search parameters based on the determined subject matter; and

integrating the selected search parameters with the search query.

13. The process of Claim 12, further comprising the step of:

returning results of the search query to the user.

14. The process of Claim 12, further comprising the step of:

sending results of the search query to a recipient.

15. The process of Claim 14, wherein the recipient is selected by the user.

16. The process of Claim 14, wherein the recipient is determined based on the results of the search query.

17. The process of Claim 14, wherein the recipient is determined based on information input by the user.

18. The process of Claim 12, further comprising the steps of:

receiving information from a recipient; and

selectively sending results of the search query to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user.

19. The process of Claim 18, wherein the received information comprises an interest in the results of the search query by the recipient.

20. The process of Claim 12, wherein the subject matter is explicitly indicated by the user.

21. The process of Claim 12, wherein the subject matter is inferentially determined based upon user input.

22. The process of Claim 12, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, and further comprising the step of:

determining the selected ones of the stored search parameters based upon the selectable applicability.

23. A system, comprising:

at least one search parameter specified by a user;

an initial search query received from the user;

a determination of subject matter of the search query;

means for selection of one or more of the user-specified search parameters based on the determined subject matter; and

means for integrating the initial search query with the selected search parameters.

24. The system of Claim 23, wherein results of the integrated search query are sent to the user.

25. The system of Claim 23, wherein results of the integrated search query are sent to a recipient.

26. The system of Claim 25, wherein the recipient is selectable by the user.

27. The system of Claim 25, wherein the recipient is based on the results.

28. The system of Claim 25, wherein the recipient is based on information input by the user.

29. The system of Claim 23, further comprising:

information received from a recipient; and

a selective transmission of the results of the integrated search query to the recipient based upon any of the received information and a selection of the recipient by the user.

30. The system of Claim 29, wherein the received information comprises an interest in the enhanced results.

31. The system of Claim 23, wherein the determination of the subject matter is explicitly indicated by the user.

32. The system of Claim 23, wherein the determination of the subject matter is inferentially determined based upon user input.

33. The system of Claim 23, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, and wherein the means for selection is based upon the selectable applicability.

Search Enhancement System Having Personal S arch Paramet rs

ABSTRACT OF THE DISCLOSURE

The system and method comprises enhancement of results for a search engine, wherein the results from the search engine are refined or reorganized, based upon information from an identified secondary source. The results obtained using a conventional search are compared against the identified secondary source, *e.g.* a ratings service, and are filtered and/or sorted appropriately. In some embodiments, identification of the secondary source, such as a ratings service comprising information which may supplement the subject of a search query, is based upon information entered by the user. In alternate embodiments, the secondary source is associated with a user, as part of general user-specified search parameters, wherein one or more parameters are consulted automatically for searches for appropriate subject matter.

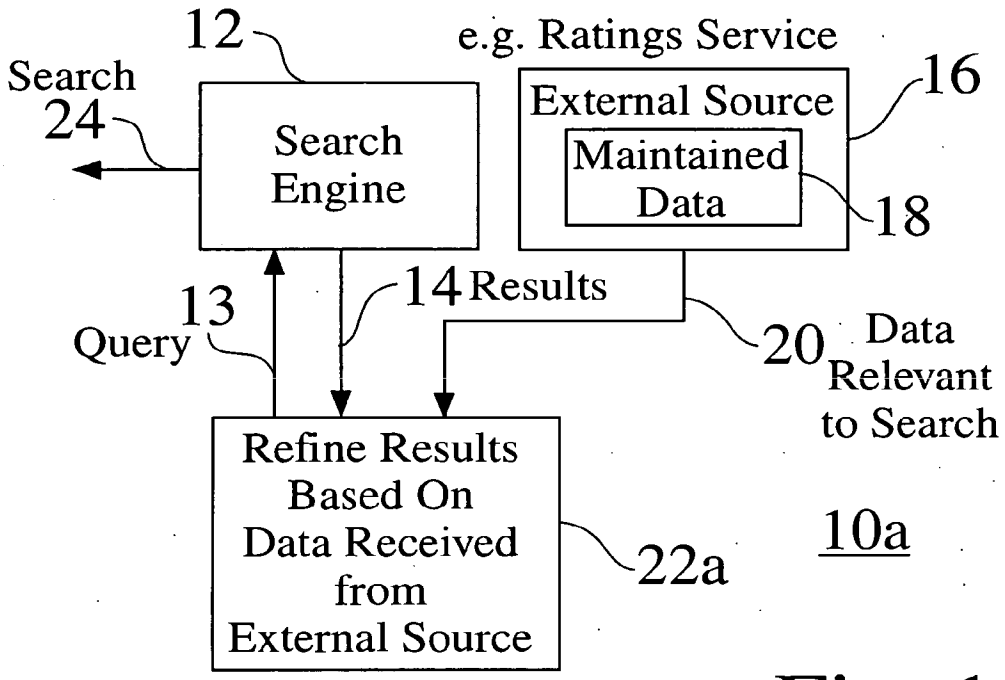

12
e.g. Ratings Service
Search
24
Search Engine
External Source
16
Maintained Data
18
13
Query
14
Results
20
Data Relevant to Search
Refine Results Based On Data Received from External Source
22a
10a

Fig. 1

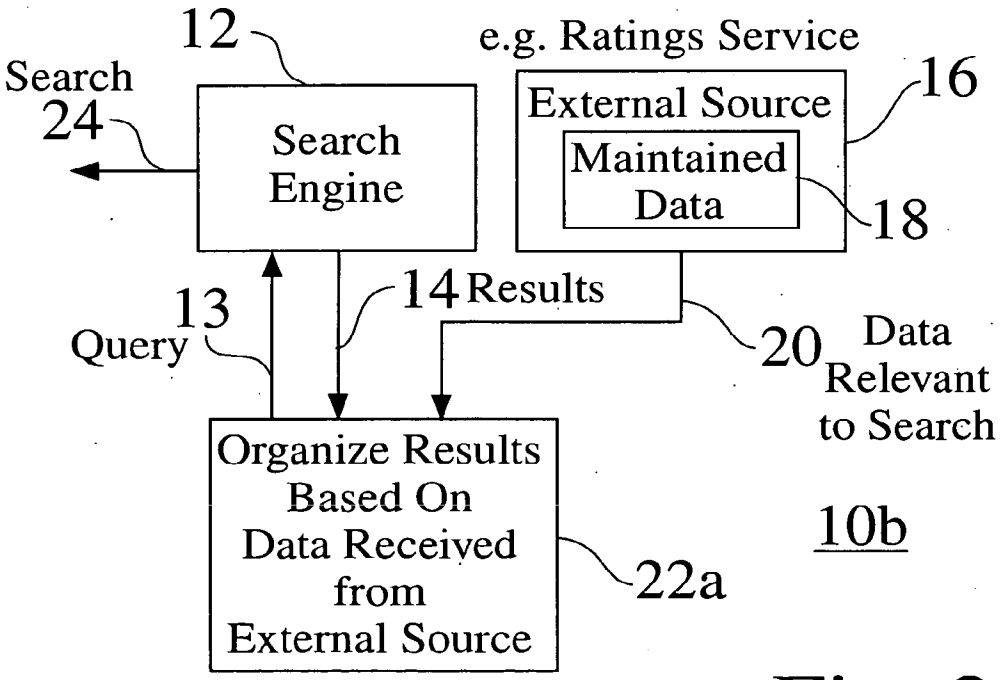

12
e.g. Ratings Service
Search
24
Search Engine
External Source
16
Maintained Data
18
13
Query
14
Results
20
Data Relevant to Search
Organize Results Based On Data Received from External Source
22a
10b

Fig. 2

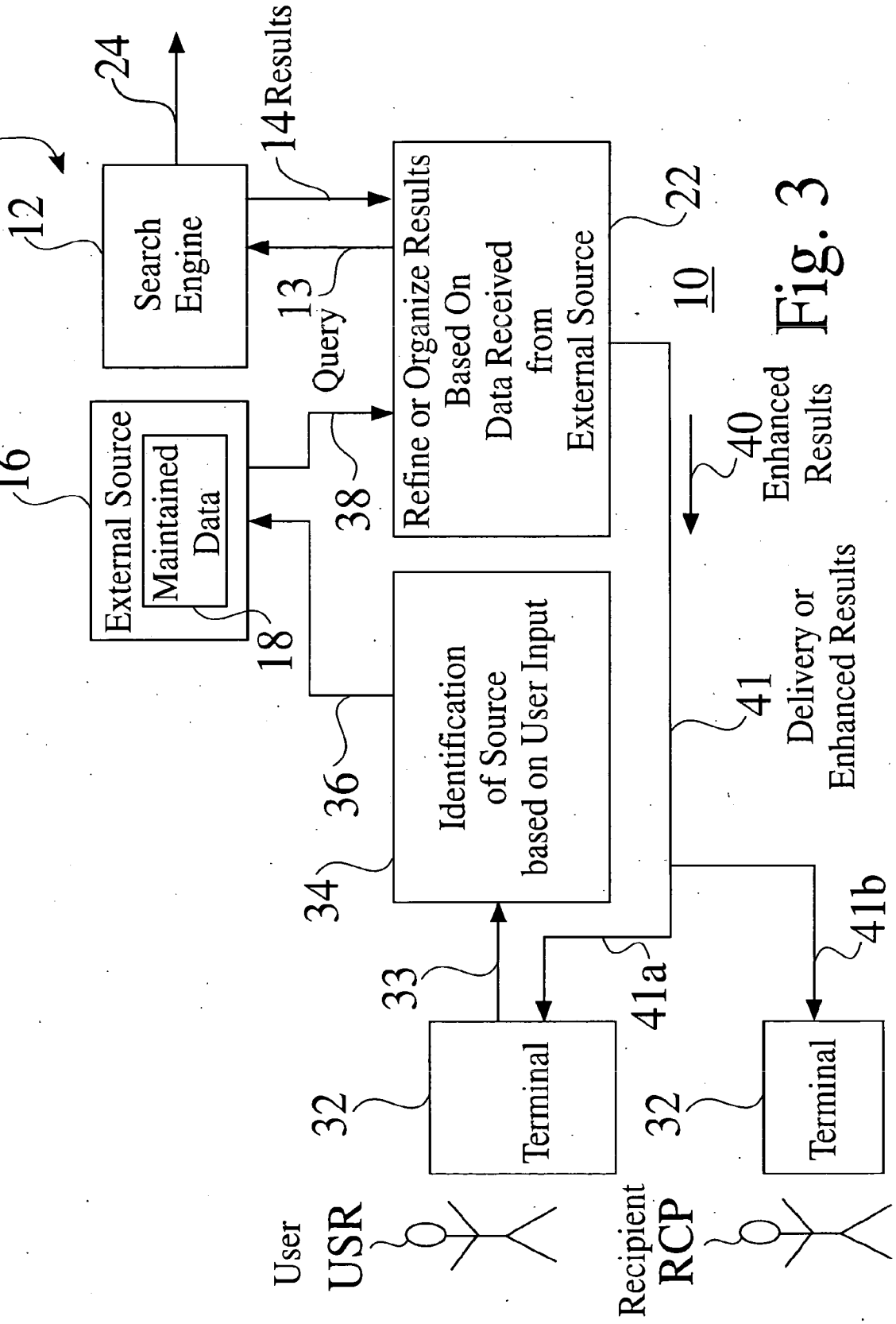

30
12
Search Engine
24
14 Results
13 Query
16
External Source
Maintained Data
18
38
Refine or Organize Results Based On Data Received from External Source
22
10
40 Enhanced Results
34
36
Identification of Source based on User Input
41 Delivery or Enhanced Results
33
41a
41b
32
Terminal
32
Terminal
User
USR
Recipient
RCP

Fig. 3

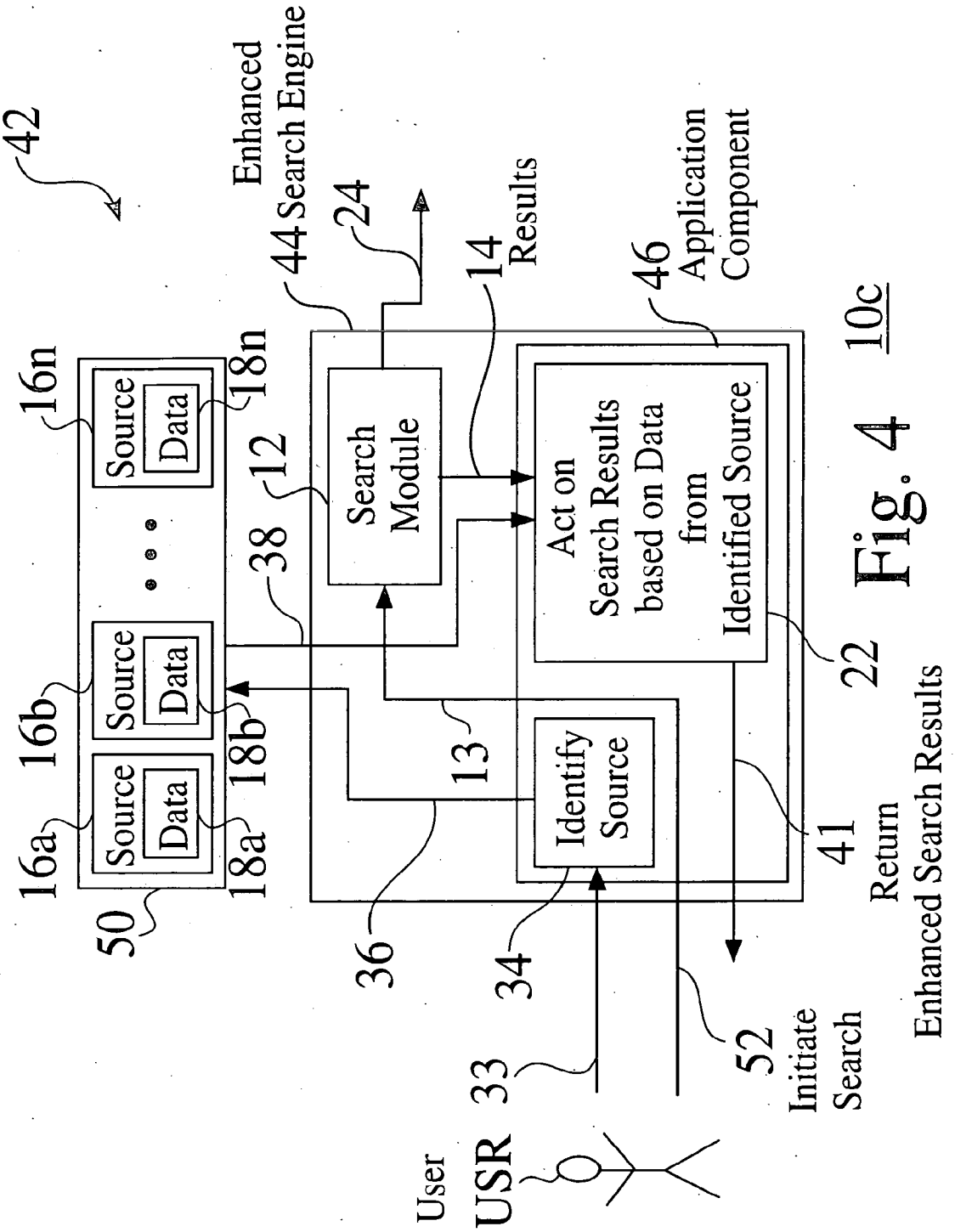

42
Enhanced
44 Search Engine
24
14
Results
46
Application
Component
16n
18n
Source
Data
12
Search
Module
38
16b
18b
Source
Data
13
Act on
Search Results
based on Data
from
Identified Source
16a
18a
Source
Data
Identify
Source
50
36
34
22
41
Return
Enhanced Search Results
User
USR
33
52
Initiate
Search

Fig. 4 10c

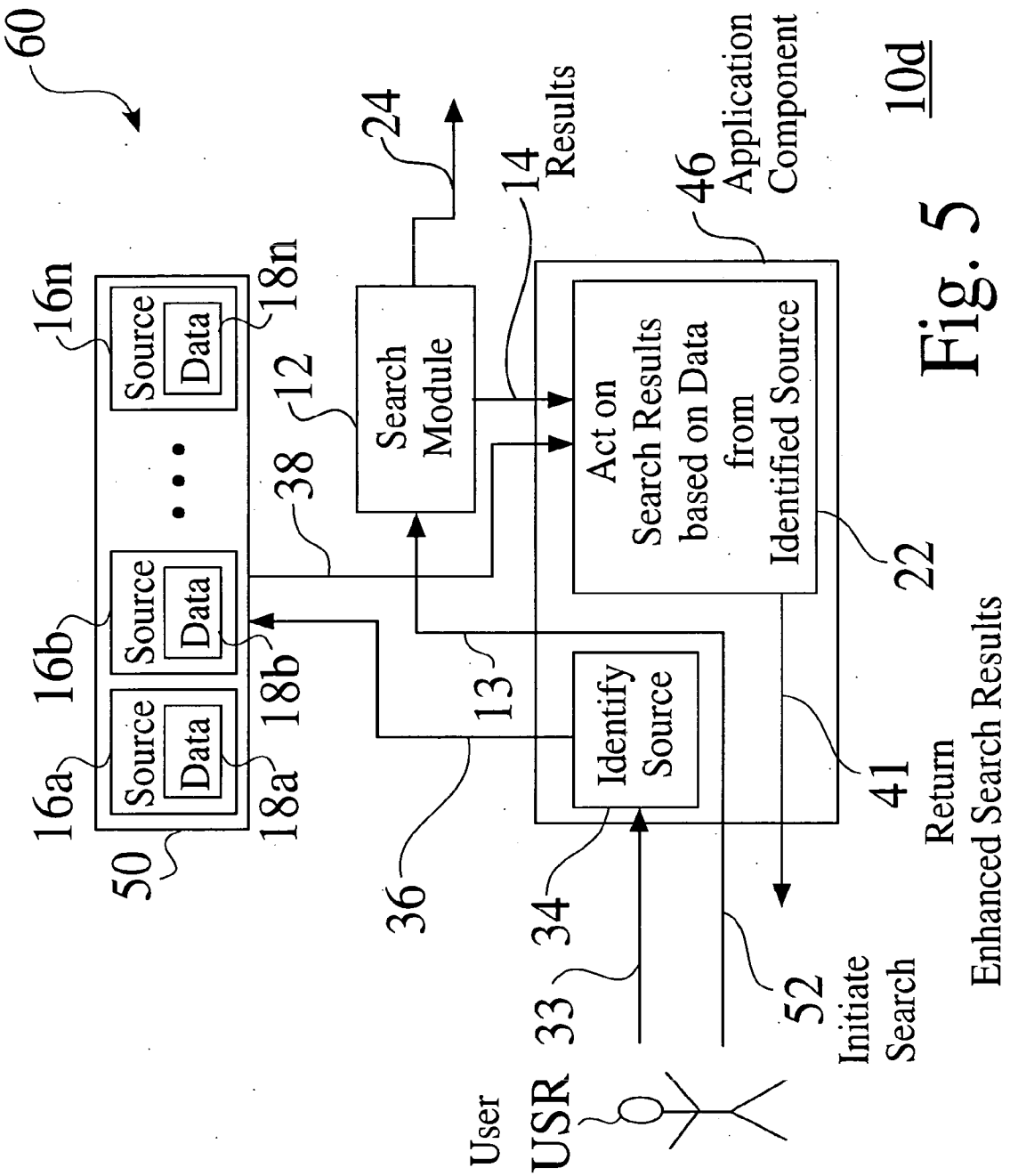

60
16n
Source
Data
18n
16b
Source
Data
18b
16a
Source
Data
18a
50
38
12
Search Module
24
14
Results
46
Application Component
Act on Search Results based on Data from Identified Source
22
13
Identify Source
36
34
41
Return Enhanced Search Results
User
USR
33
52
Initiate Search
10d

Fig. 5

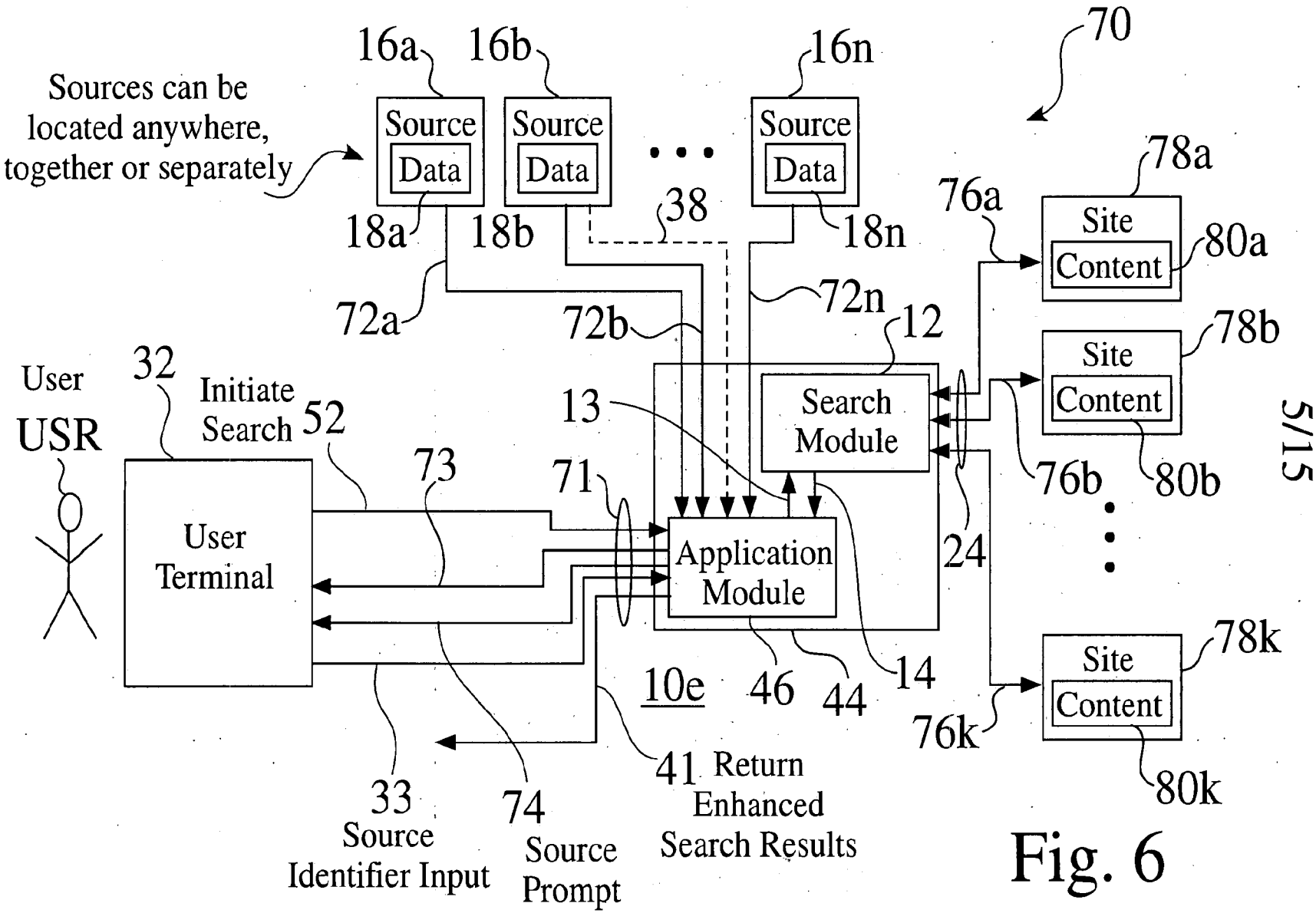


Fig. 6

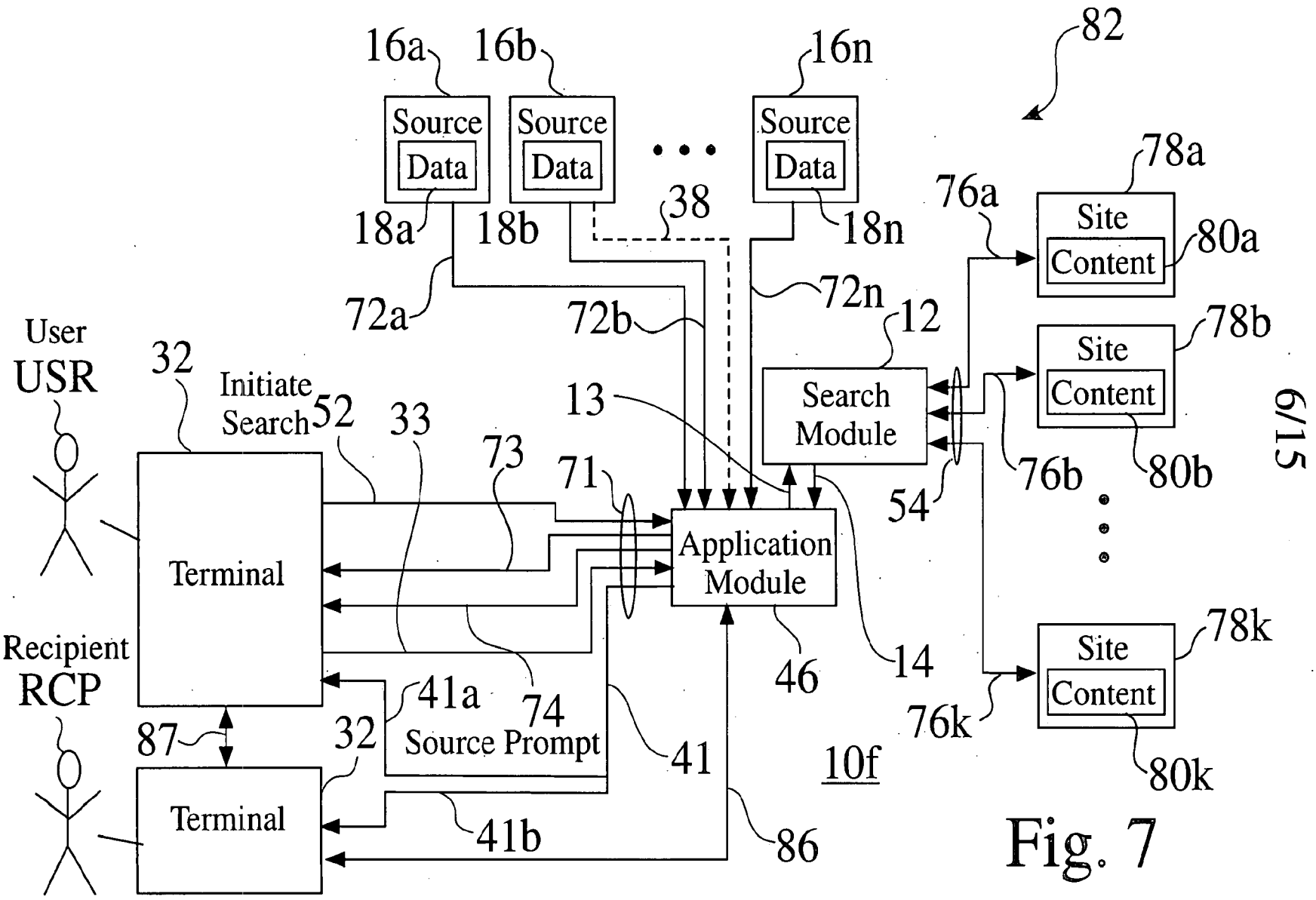


Fig. 7

6/15

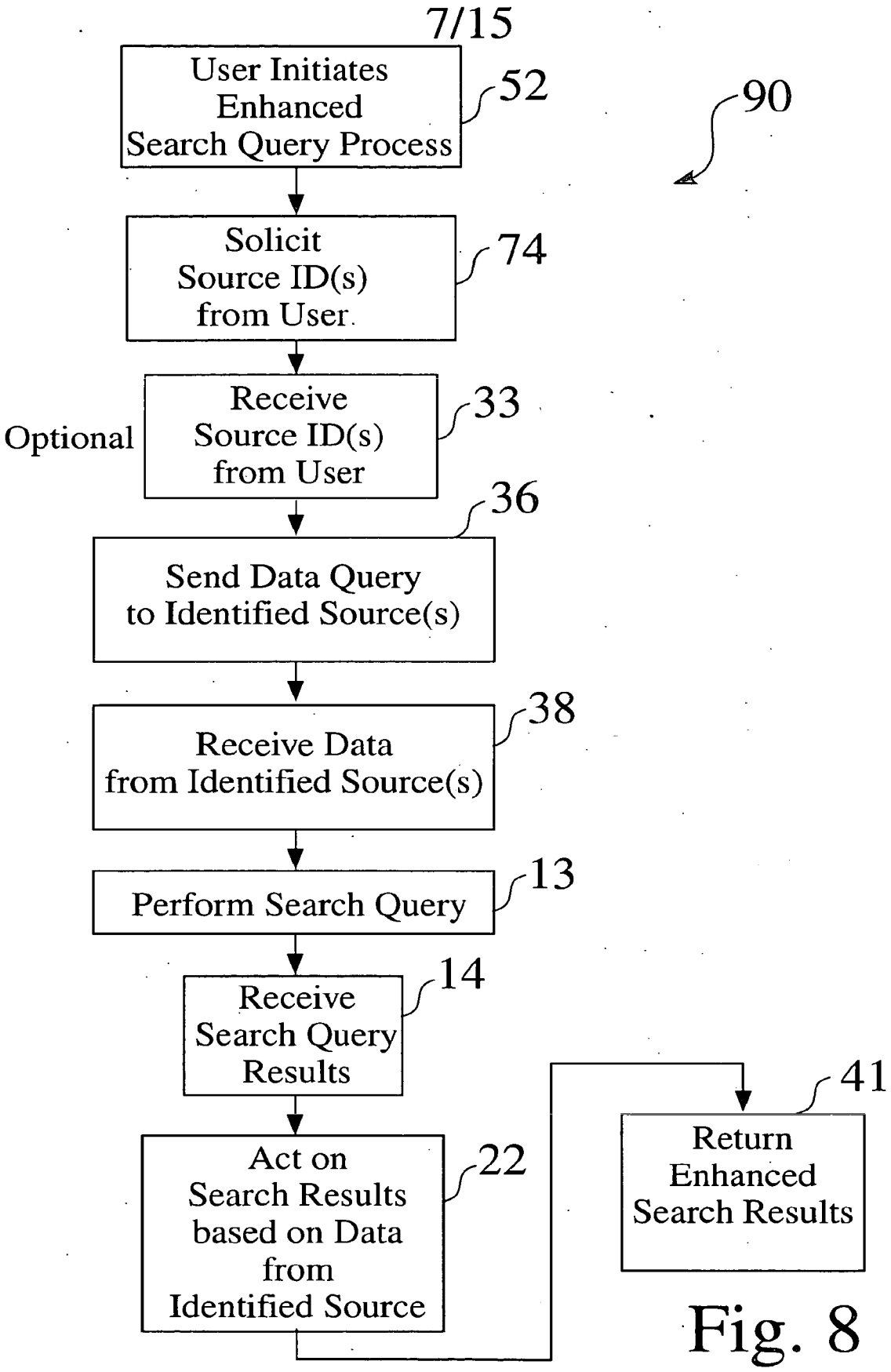

User Initiates Enhanced Search Query Process
52
90
Solicit Source ID(s) from User
74
Optional
Receive Source ID(s) from User
33
36
Send Data Query to Identified Source(s)
Receive Data from Identified Source(s)
38
Perform Search Query
13
Receive Search Query Results
14
Act on Search Results based on Data from Identified Source
22
Return Enhanced Search Results
41

Fig. 8

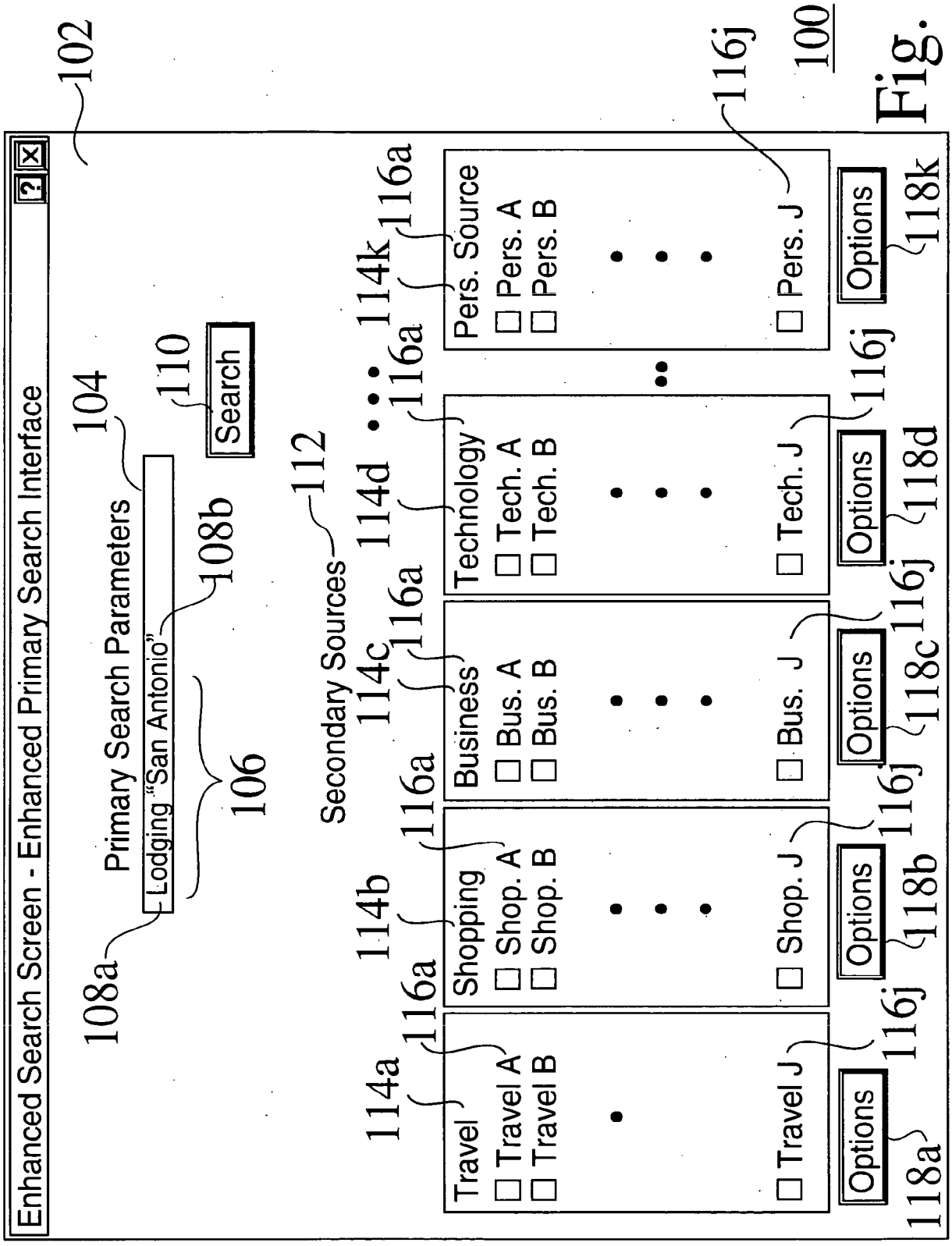

Enhanced Search Screen - Enhanced Primary Search Interface
102
Primary Search Parameters
104
Lodging "San Antonio"
108a
108b
106
Search
110
Secondary Sources
112
Travel
Travel A
Travel B
Travel J
114a
116a
116j
118a
Shopping
Shop. A
Shop. B
Shop. J
114b
116a
116j
118b
Business
Bus. A
Bus. B
Bus. J
114c
116a
116j
118c
Technology
Tech. A
Tech. B
Tech. J
114d
116a
116j
118d
Pers. Source
Pers. A
Pers. B
Pers. J
114k
116a
116j
118k
Options
100

Fig. 9

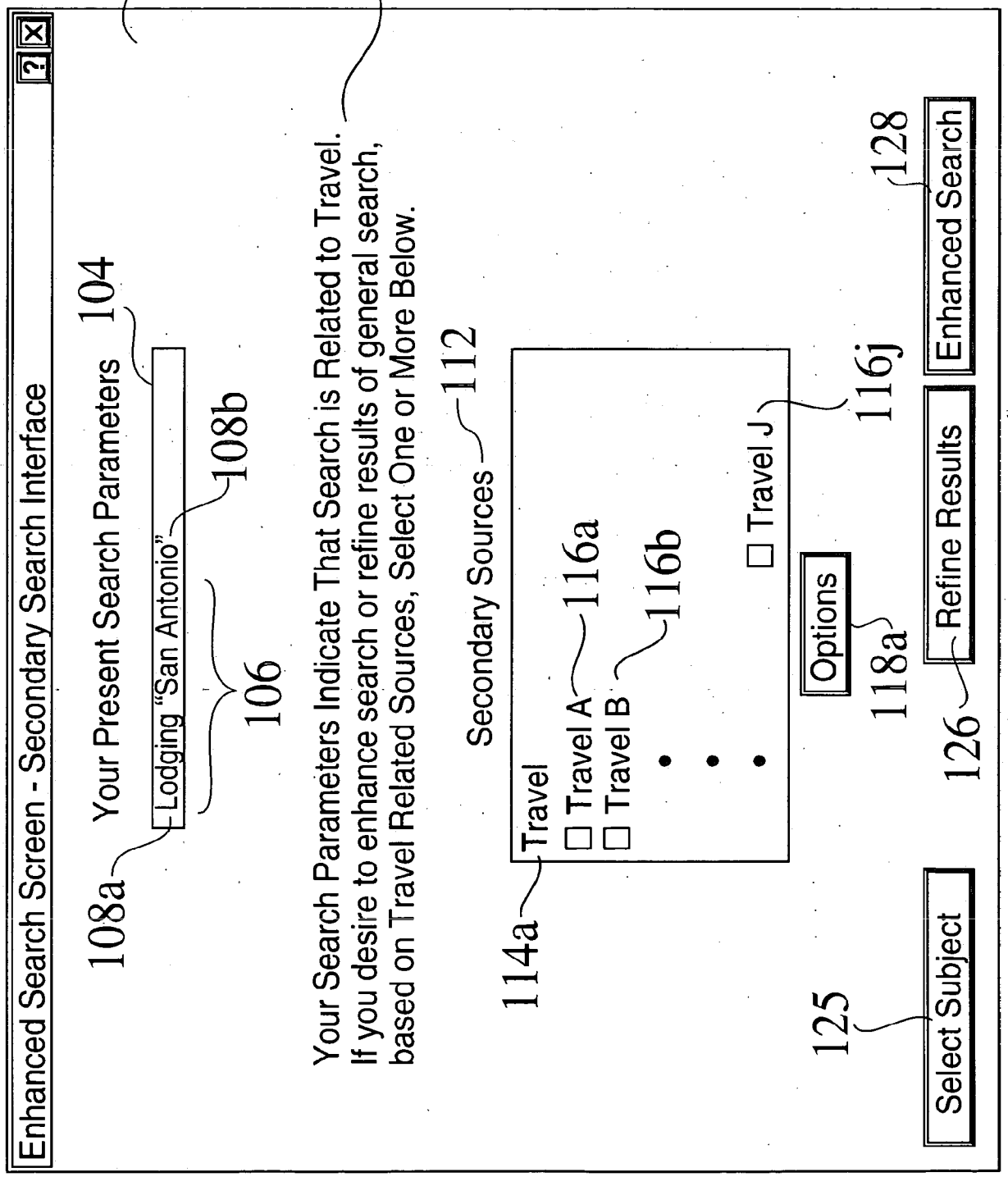

Enhanced Search Screen - Secondary Search Interface
122
Your Present Search Parameters
104
108a
Lodging "San Antonio"
108b
106
124
Your Search Parameters Indicate That Search is Related to Travel.
If you desire to enhance search or refine results of general search,
based on Travel Related Sources, Select One or More Below.
Secondary Sources
112
114a
Travel
Travel A
116a
Travel B
116b
Travel J
116j
Options
118a
125
Select Subject
126
Refine Results
128
Enhanced Search
120

Fig. 10

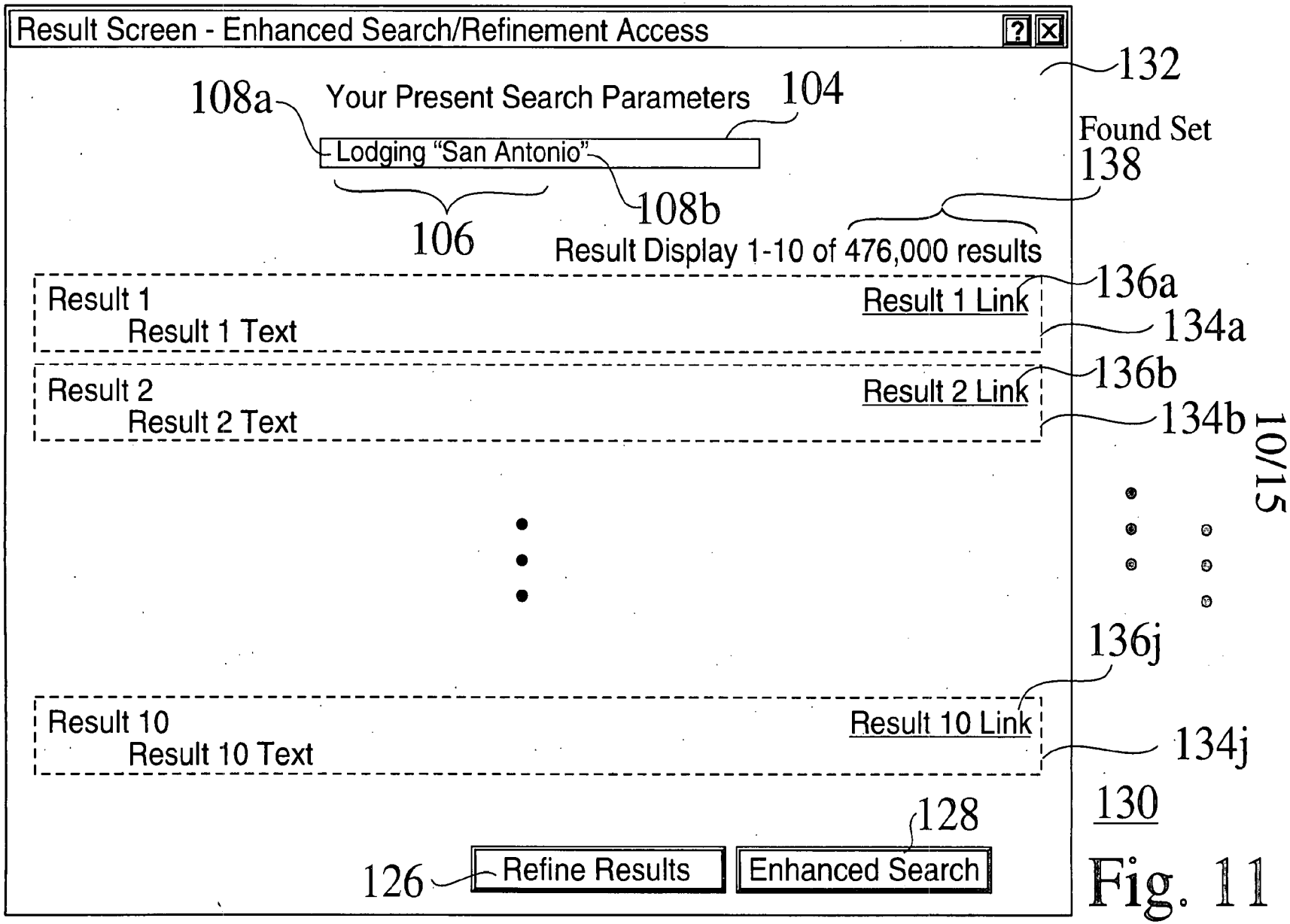


Fig. 11

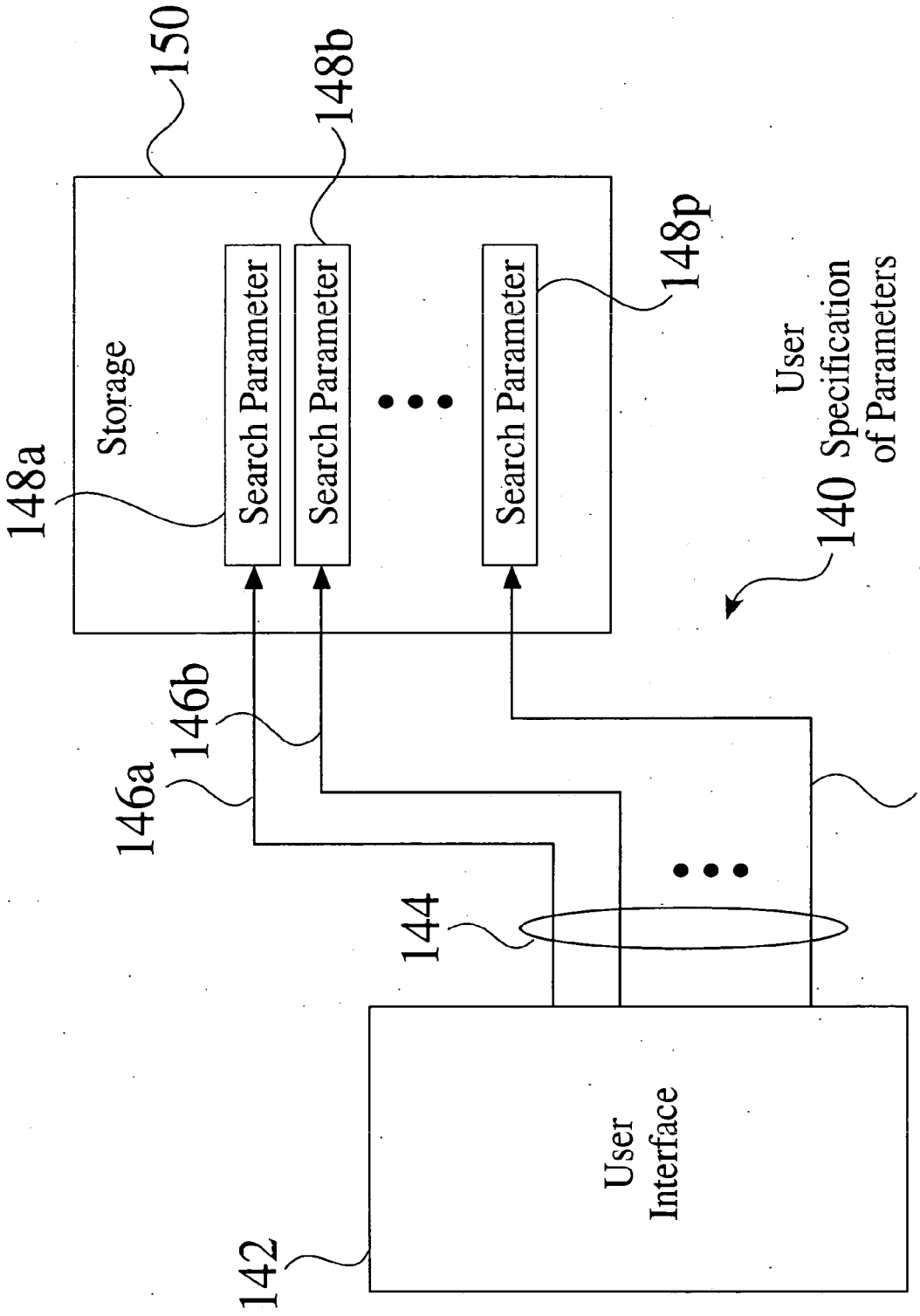

150
148b
148p
148a
Storage
Search Parameter
Search Parameter
Search Parameter
User
140 Specification
of Parameters
146b
146a
146p
144
142
User
Interface

Fig. 12

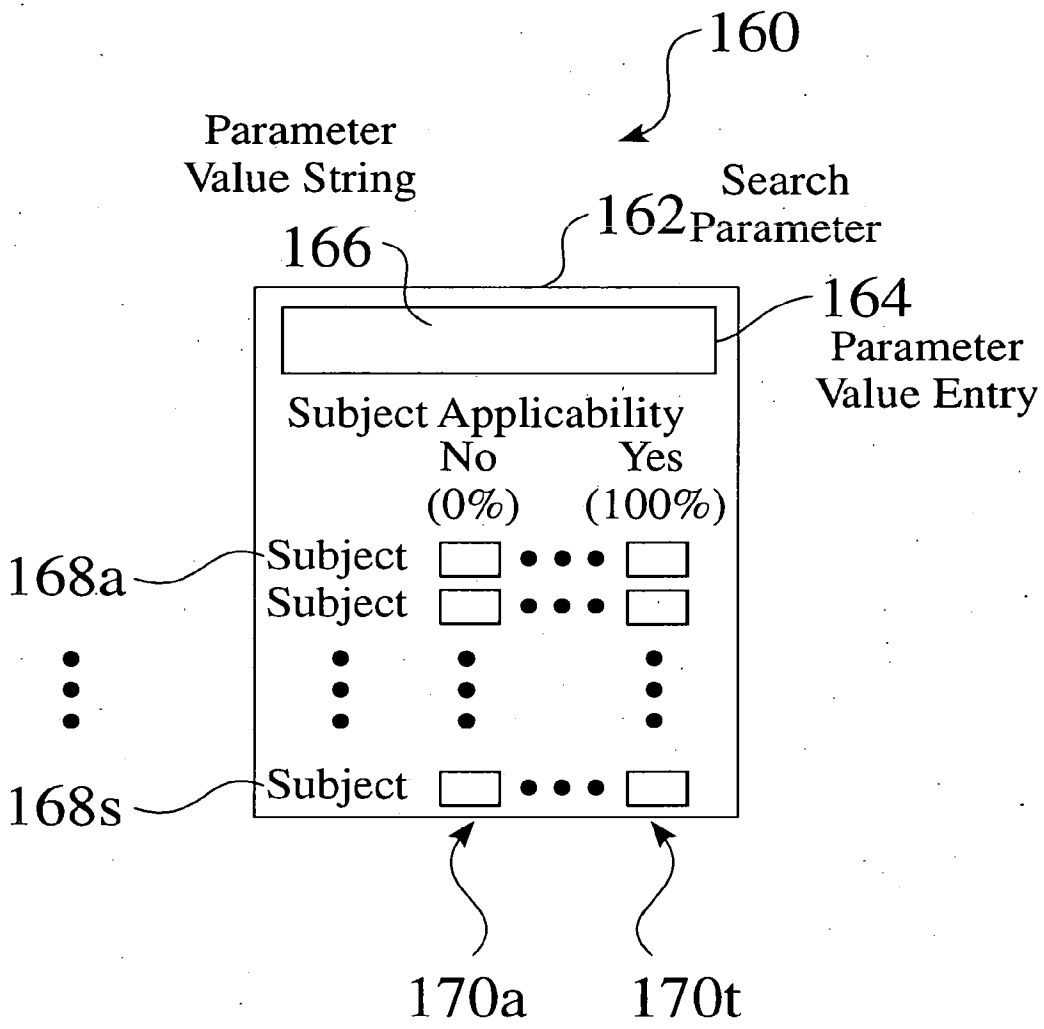

160
Parameter
Value String
Search
Parameter
162
166
164
Parameter
Value Entry
Subject Applicability
No
(0%)
Yes
(100%)
168a
Subject
Subject
168s
Subject
170a
170t

Fig. 13

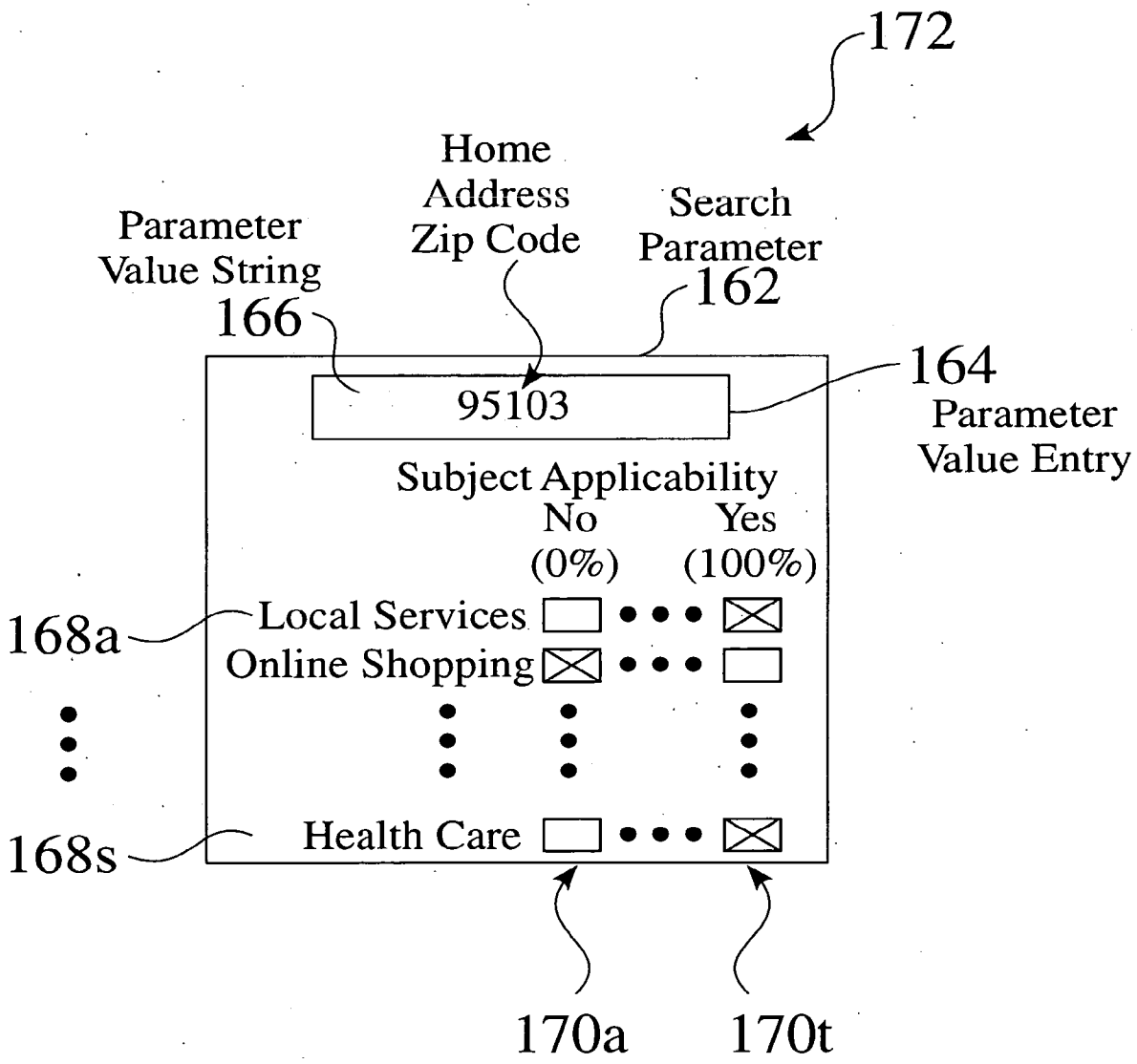

172
Home
Address
Zip Code
Search
Parameter
Parameter
Value String
166
162
164
Parameter
Value Entry
95103
Subject Applicability
No
(0%)
Yes
(100%)
168a
Local Services
Online Shopping
168s
Health Care
170a
170t

Fig. 14

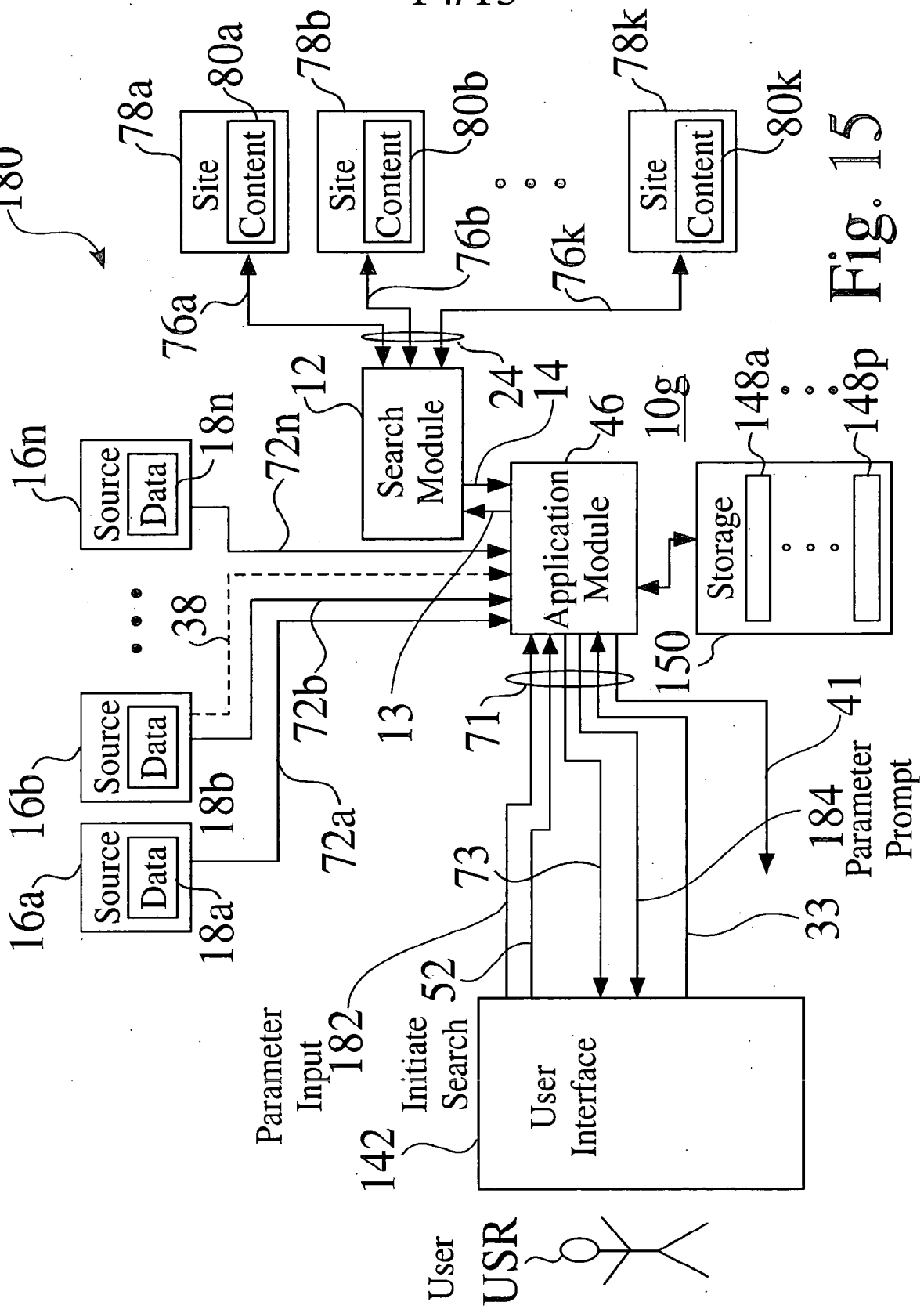

180
78a
80a
Site
Content
78b
80b
78k
80k
76a
76b
76k
12
Search Module
24
14
46
10g
Application Module
148a
148p
Storage
150
16n
Source
Data
18n
72n
16b
18b
72b
16a
18a
72a
38
13
71
41
Parameter Prompt
184
73
33
52
Initiate Search
142
Parameter Input
182
User Interface
User
USR

Fig. 15

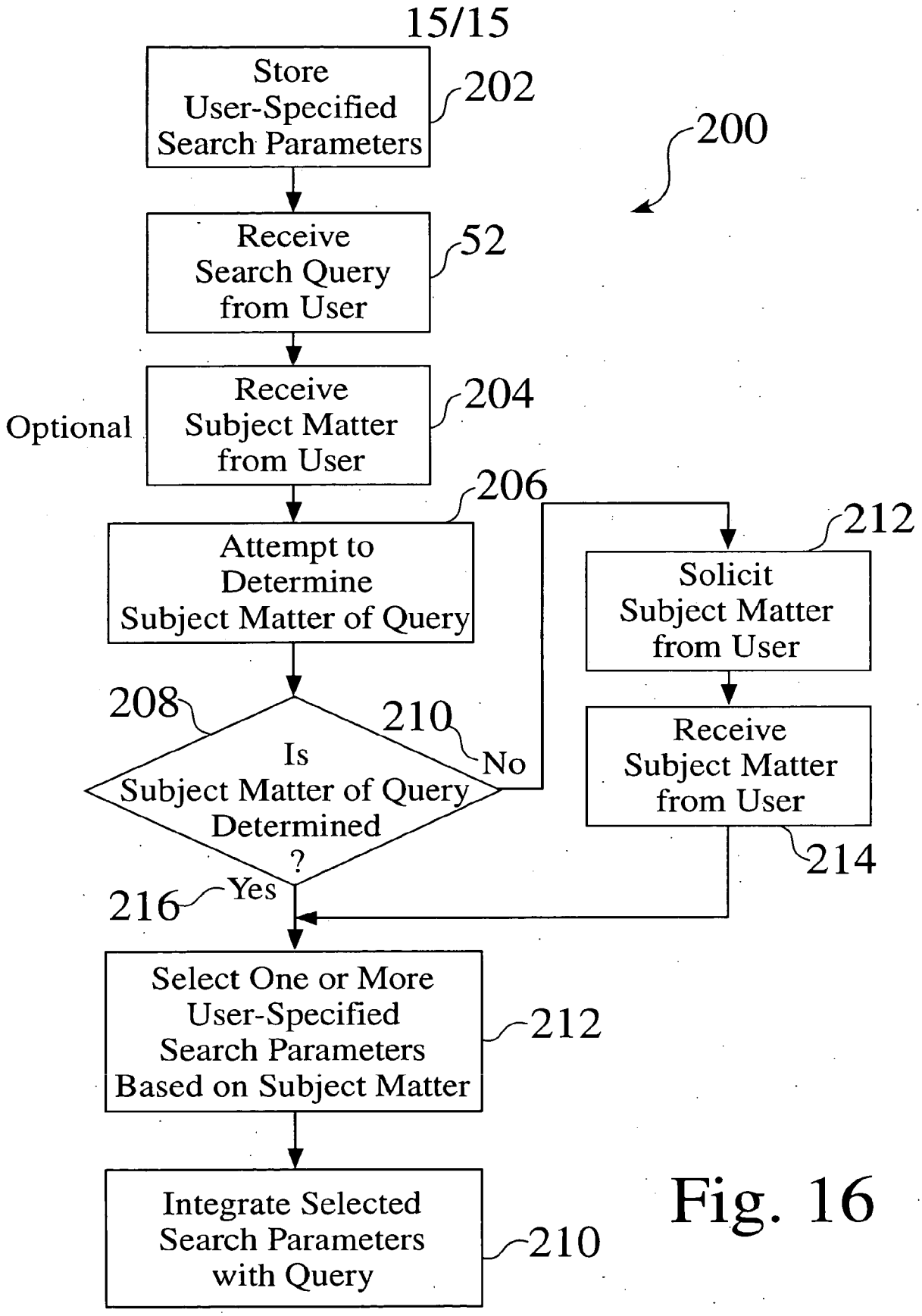

Store User-Specified Search Parameters
202
200
Receive Search Query from User
52
Optional
Receive Subject Matter from User
204
206
Attempt to Determine Subject Matter of Query
212
Solicit Subject Matter from User
208
210
No
Is Subject Matter of Query Determined ?
Receive Subject Matter from User
214
216
Yes
Select One or More User-Specified Search Parameters Based on Subject Matter
212
Integrate Selected Search Parameters with Query
210

Fig. 16

Attorney Docket No. AOL0142

DECLARATION FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name;

I believe I am the original, first, and sole inventor (if only one name is listed below) or an original, first, and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

SEARCH ENHANCEMENT SYSTEM HAVING PERSONAL SEARCH PARAMETERS

the specification of which (check one) __X__ is attached hereto, or __ was filed on_____as Application Serial No. _____and was amended on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

==========

I hereby claim foreign priority benefits under Title 35, United Sates Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)	Priority Claimed Yes	No
_____ Number Country Day/Month/Year Filed	____	____
_____ Number Country Day/Month/Year Filed	____	____

==========

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith:

MICHAEL A. GLENN, Reg. No. 30,176
JAMES R. BRAMSON, Reg. No. 41,632
DONALD M. HENDRICKS, Reg. No. 40,355
CHRISTOPHER PEIL, Reg. No. 45,005
IVY LEE MAY, Reg.No. 46,925
JULIA THOMAS, Reg. No. 52,283

SEND CORRESPONDENCE TO:

GLENN PATENT GROUP, 3475 Edison Way, Suite L, Menlo Park, CA 94025

==========

Attorney Docket No. AOL0142

I hereby claim the benefit under Title 35, United States code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Ser. No.	Filing Date	Status: Patented, Pending, Abandoned

==========

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor: Edmund J. FISH

Inventor's signature _____ Date _____

Residence 22260 Pacific Blvd., Dulles, Virginia 20166

Post Office Address Same

Citizenship Unjited States of America

PATENT APPLICATION SERIAL NO. _____

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

10/21/2003 HLE333 00000094 071445 10685749

01 FC:1001 770.00 DA
02 FC:1202 234.00 DA

PTO-1556
(5/87)

*U.S. Government Printing Office: 2002 — 489-267/69033

PATENT APPLICATION FEE DETERMINATION RECORD
Effective October 1, 2003

Application or Docket Number: 10685749

CLAIMS AS FILED - PART I

	(Column 1)	(Column 2)
TOTAL CLAIMS	33	
FOR	NUMBER FILED	NUMBER EXTRA
TOTAL CHARGEABLE CLAIMS	33 minus 20=	* 13
INDEPENDENT CLAIMS	3 minus 3 =	* 0
MULTIPLE DEPENDENT CLAIM PRESENT		☐

* If the difference in column 1 is less than zero, enter "0" in column 2

SMALL ENTITY TYPE ☐		OR	OTHER THAN SMALL ENTITY	
RATE	FEE		RATE	FEE
BASIC FEE	385.00	OR	BASIC FEE	770.00
X$ 9=		OR	X$18=	234
X43=		OR	X86=	
+145=		OR	+290=	
TOTAL		OR	TOTAL	1004

CLAIMS AS AMENDED - PART II

AMENDMENT A

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				☐

SMALL ENTITY		OR	OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE		RATE	ADDITIONAL FEE
X$ 9=		OR	X$18=	
X43=		OR	X86=	
+145=		OR	+290=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

AMENDMENT B

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				☐

RATE	ADDITIONAL FEE		RATE	ADDITIONAL FEE
X$ 9=		OR	X$18=	
X43=		OR	X86=	
+145=		OR	+290=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

AMENDMENT C

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				☐

RATE	ADDITIONAL FEE		RATE	ADDITIONAL FEE
X$ 9=		OR	X$18=	
X43=		OR	X86=	
+145=		OR	+290=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20."
***If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3."
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371 (c) DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
10/685,749	10/14/2003	Edmund J. Fish	AOL0142

CONFIRMATION NO. 6161

22862
GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

FORMALITIES LETTER

OC000000011746522

Date Mailed: 01/22/2004

NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

FILED UNDER 37 CFR 1.53(b)

Filing Date Granted

Items Required To Avoid Abandonment:

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given **TWO MONTHS** from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The oath or declaration is unsigned.
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of $130 for a non-small entity, must be submitted with the missing items identified in this letter.

SUMMARY OF FEES DUE:

Total additional fee(s) required for this application is **$130** for a Large Entity

- **$130** Late oath or declaration Surcharge.

Replies should be mailed to: Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

A copy of this notice MUST be returned with the reply.

Customer Service Center
Initial Patent Examination Division (703) 308-1202
PART 3 - OFFICE COPY

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 072 982 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
31.01.2001 Bulletin 2001/05

(51) Int. Cl.[7]: **G06F 17/30**

(21) Application number: **00116441.7**

(22) Date of filing: **28.07.2000**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **30.07.1999 JP 21661799**

(71) Applicant:
MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.
Kadoma-shi, Osaka 571-8501 (JP)

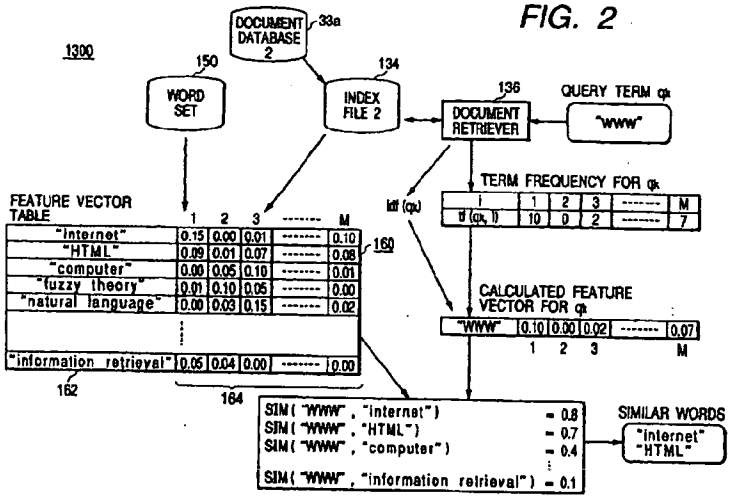
(72) Inventors:
- **Sato, Mitsuhiro**
  Atsugi-shi, Kanagawa-ken 243-0216 (JP)
- **Ito, Hayashi**
  Matsudo-shi, Chiba-ken 271-0045 (JP)

(74) Representative:
Manitz, Finsterwald & Partner
Postfach 22 16 11
80506 München (DE)

(54) Method and system for similar word extraction and document retrieval

(57) Similar words for a given sequence of letters are extracted from a document database of documents is provided. The occurrence pattern of the similar words is similar to that of the given sequence. A feature vector table which contains a first feature vector for each of words in the document database is prepared. Statistical information on the given sequence is obtained through a retrieval of the document database with respect to the given sequence. From the statistical information, a second feature vector with respect to the given sequence is calculated. A degree of similarity between the second feature vector and each of the first feature vectors is calculated Such words as yield degrees of similarity higher than a predetermined value is selected.

FIG. 2



EP 1 072 982 A2

Printed by Xerox (UK) Business Services
2.16.7 (HRS)/3.6

Description

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The invention generally relates to a document retrieval system and, more specifically, to an occurrence pattern-based similar word extractor and its applications to the query expansion and the multilingual document retrieval.

2. Description of the Prior Art

[0002] As numerous documents have come to be computerized and distributed due to the spread of the Internet and personal computers, electronic search for documents is getting popular. Conventional document retrieval systems usually make a reference for query key words input by a user and provide documents that include the key words as the result of retrieval. However, since there may be possibly a lot of documents that describe the same subject matter in different expressions, it is difficult for a user to find all such documents as meet his or her desire by using only key words the user hits on. For this purpose, various query expansion techniques have been proposed so far which expand a given query for retrieval.

[0003] Also, requests for the retrieval of information in languages other than user's mother tongue are increasing with an improvement of circumstance that enables the access to information in the world through spreading of Internet. In order to meet such requests, studies on multilingual document retrieval methods are increasing recently.

[0004] An example of query expansion techniques is proposed by Akamine et al., "Information Retrieval System with Query Expansion Using WordNet" in the proceedings of the fifty second national convention of Information Processing Society of Japan (IPSJ), 1996, pp. 4-201 — 4-202. In this system, synonyms for query terms are found by using a fixed thesaurus and the query is expanded with the found synonyms. However, since the found synonyms are dependent on the vocabulary of the fixed thesaurus, it may not be suitable for retrieval in a special field.

[0005] Another query expansion technique is proposed by K. Saito et al., "Concept-Based Query Expansion", IPSJ Study Group Report, Information Study Fundamentals 47-10, 1997, pp. 67-74. In this system, each of words that occurs in a document database to be searched, i.e., each word of a dictionary, is mapped on a vector space the dimensions of which correspond to the documents constituting the document database. The degree of similarity between two words is determined by the inner product of vectors defined for the two words in the vector space. Similar words relevant to desired documents are obtained and used for the expansion of the query.

[0006] However, the degree of similarity can not obtained for a word that is not included in the vocabulary of the dictionary in just mentioned query expansion technique. For this reason, the technique is not applicable to queries containing any query term which is not included in the dictionary.

[0007] A multilingual document retrieval system is described by L. Ballesteros et al., "Resolving Ambiguity for Cross-language Retrieval" ACM-SIGIR98, 1998, pp. 64-71. In reference 3, all possible target language translations are listed for a given source language query by using parallel corpora which contain a set of documents and their translations in one or more languages. A target language query is generated by narrowing the listed translations down by using co-occurrence statistics.

[0008] Since the use of one or more translation dictionary is assumed in this multilingual system, selected translations depend on the vocabulary of the translation dictionary, which means that no translation is obtained for terms that are not included in the vocabulary. If the system is applied to a document database in a language in which a space is not left between words as in case of Japanese and if a given query includes a sequence of letters (usually Chinese characters) which is not included in the vocabulary, then no translation is obtained for the letter sequence.

SUMMARY OF THE INVENTION

[0009] The present invention is intended for solving above and other problems, and makes it an object to provide a similar word extractor capable of extracting, from a predetermined document database, similar words for a given query term which has not yet been widely accepted.

[0010] It is another object of the invention to provide query expansion techniques incorporating the similar word extractor.

[0011] It is further object of the invention to provide a multilingual document retrieval system incorporating the similar word extractor.

[0012] According to an aspect of the invention, a method of and a system for extracting similar words for a given sequence of letters from a document database of documents is provided. The occurrence pattern of the similar words is similar to that of the given sequence. A feature vector table which contains a first feature vector for each of words in

the document database is prepared. Statistical information on the given sequence is obtained through a retrieval of the document database with respect to the given sequence. From the statistical information, a second feature vector with respect to the given sequence is calculated. A degree of similarity between the second feature vector and each of the first feature vectors is calculated. Such words as yield degrees of similarity higher than a predetermined value is selected.

[0013] The feature vector table is prepared by generating a word set comprising the words in the document database; generating an index file from the document database, the index file being such that statistical information for each of words used in the document database can be obtained from the index file; and calculating each first feature vector from the index file. In order to obtain statistical information, the index file is searched for the given sequence.

[0014] According to an aspect of the invention, a query expansion method and system for use in a document retrieval apparatus comprising a first document database and a retriever is provided. A given query is expanded to feed the retriever. For this purpose, similar words of a query word in the given query are utilized. The similar words of the query word are extracted from a second document database of documents. The occurrence pattern of the similar words is similar to that of the query word. The extraction of the similar words is achieved by preparing a feature vector table which contains a first feature vector for each of words in the second document database; obtaining statistical information on the query word through a retrieval of the second document database with respect to the query word; calculating a second feature vector with respect to the query word from the statistical information; calculating a degree of similarity between the second feature vector and each of the first feature vectors; and selecting, as the similar words, such words as yield degrees of similarity higher than a predetermined value.

[0015] The first and second document databases are preferably the same thing.

[0016] According to another aspect of the invention, a translation extracting method and system are provided. For a given word in a source language, translations in a target language are provided such that an occurrence pattern of the given word in a source-language document database being the same as that of translations in a target-language document database if one of the two document databases is a translation of the other. The target-language translations are obtained by preparing a feature vector table which contains a first feature vector for each of words in the target-language document database; obtaining statistical information on the query word through a retrieval of the source-language document database with respect to the query word; calculating a second feature vector with respect to the query word from the statistical information; calculating a degree of similarity between the second feature vector and each of the first feature vectors; and selecting, as the translations, such target-language words as yield degrees of similarity higher than a predetermined value.

[0017] The feature vector table is prepared by generating an index file from the target-language document database, the index file being such that statistical information for each of words used in the target-language document database can be obtained from the index file; and calculating a first feature vector for each word in a given dictionary (or a given set of words) from the index file. In order to obtain statistical information, the index file is searched for the query word.

[0018] According to further aspect of the invention, a query expanding method and system for use in a document retrieval system comprising a first document database and a retriever are provided. The query expansion means converting a source-language query into a target-language query to feed the retriever. The query expansion is achieved by extracting a translation of each of query words in the source-language query from a second document database of target-language documents and replacing each query word in the source-language query with the translation of the query word. The occurrence pattern of the translation is similar to that of the query word. The extraction of the translation is achieved by preparing a feature vector table which contains a first feature vector for each of words in the target-language document database; obtaining statistical information on the query word through a retrieval of the source-language document database with respect to the query word; calculating a second feature vector with respect to the query word from the statistical information; calculating a degree of similarity between the second feature vector and each of the first feature vectors; and selecting, as the translations, such target-language words as yield degrees of similarity higher than a predetermined value.

[0019] According to an aspect of the invention, a method of receiving a given letter sequence and providing bisected letter sequences with a sufficient propriety of division is provided. The method comprises the steps of forming a retrieval expression for retrieving documents which contains all of the letters of the letter sequence but does not contain the letter sequence itself; obtaining statistical information by executing a retrieval based on the retrieval expression; generates a feature vector for the retrieval expression from the statistical information, the feature vector being referred to as "mask feature vector"; dividing the letter sequence into two subsequences shifting the division point one by one to obtain N-1 sets of bisected letter sequences, where N is the number of letters in the letter sequence; calculate a feature vector for each of the bisected letter sequences of each of the N-1 sets; masks each of the calculated feature vectors with the mask feature vector and normalizing the masked calculated feature vector; calculating a degree of similarity for each of the N-1 sets, the degree of similarity being defined as an inner product of the bisected letter sequences of each of the N-1 sets; and outputting such bisected letter sequences as have a degree of similarity larger

than a threshold.

[0020] According to another aspect of the invention, for a given word, at least one variant including the given word can be obtained by using the above-described translation extracting system. In order to achieve this, the translations are first obtained for the given word by using the translation extracting system. The translation extracting system is configured by replacing the feature vector table with a feature vector table which contains a third feature vector for each of words in the first-language document database and by exchanging the first-language document database and the second-language document database. Second translations for each of the translations are obtained as the variant(s) by using reconfigured translation extracting system.

BRIEF DESCRIPTION OF THE DRAWING

[0021] The features and advantages of the present invention will be apparent from the following description of an exemplary embodiment of the invention and the accompanying drawings, in which:

FIG. 1 is a schematic block diagram showing an exemplary arrangement of a document retrieval computer system in which the present invention can be embodied;
FIG. 2 is a diagram showing an exemplary arrangement of a similar word extractor according to the invention;
FIG. 3 is a flowchart showing the operation executed by the processor 20 under the control of the similar word extractor 130 according to the invention;
FIG. 4 is a diagram showing an exemplary structure of the feature vector table 160;
FIG. 5 is a flowchart showing an operation of a query expander using the similar word extractor 130 of FIG. 3;
FIG. 6 is a diagram showing an arrangement of a translation extractor system 1300a for use in a query expander for cross-language retrieval;
FIG. 7 is a flowchart of a query expander program 31b executed in combination with the extractor 130;
FIG. 8 is a flowchart showing an exemplary operation of a letter sequence divider according to the invention;
FIG. 9 is a diagram showing an example of a retrieval expression for a letter sequence

"自律移動ロボット";

FIG. 10 is a diagram showing seven sets of divided sequences for the letter sequence

"自律移動ロボット";

FIG. 11 is diagram showing an example of the processing of step 212;
FIG. 12 is a diagram showing the way of judging whether each of divided sequence sets has a sufficient division propriety; and
FIG. 13 is a flowchart showing an exemplary operation of a variant finder 300 according to the invention.

[0022] Throughout the drawing, the same elements when shown in more than one figure are designated by the same reference numerals.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0023] FIG. 1 is a schematic block diagram showing an exemplary arrangement of a document retrieval computer system in which the present invention can be embodied. The document retrieval computer system 1 typically comprises a server 10, communication media 12 and numerous clients 14. In the system 1, the server receives queries from (typically remote) clients 14 and responsively returns retrieval results to the clients. The communication media 12 generally includes various communication networks such as the internet, LAN, etc. Though a client-server configuration is shown in FIG. 1, the invention may be implemented as a standalone computer system, in which case the communication media 12 and the clients 14 would be eliminated from the figure.

[0024] The server 10, which is typically a computer with a capacity and a speed necessary for providing document retrieval services to some clients 14 at a time, comprises one or more processor(s) and memories 20 which serves as

the core of the computer; mass storage 30 for storing various programs and data; communication interfaces 40 for communication with the clients 14; a user interface 50; and input and output devices 60 for permitting an administrator to operate the server 10. The mass storage 30 preferably includes a hard disc (not shown) for storing programs and data necessary for document retrieval. The mass storage 30 may further include, e.g., one or more optical disc(s) (not shown) for storing a document database 33. The above-described server 10 and/or the components 20, 30 and 40 through 60 may be any suitable standard ones.

[0025] Each client 14 may be any suitable terminal or personal computer with a communication interface (not shown) such as a modem or a network interface.

[0026] The mass storage 30 typically stores a document database/retrieval engine 32 and a query expander 31 for receiving a query from a client 14 and providing an expanded version of the query in accordance with the principles of the invention as detailed later. The document database/retrieval engine 32 and a query expander 31 may be a combination of any suitable document database and any suitable retrieval engine, and typically comprises a document database 33 which contains a lot of documents, an index generator 34, an index file 35 generated by the index generator 34 from the database 33, and a document retriever 36 for retrieving documents most relevant to the query by using the index file 35.

[0027] It will be helpful for better understanding the invention to define some terms used wherein prior to the description of the following illustrative embodiments.

[0028] Unless otherwise stated, the term "document" should be taken to mean an entire document or any portion thereof, such as a title, an abstract, or one or more clauses, sentences, or paragraphs.

[0029] Unless otherwise stated, the term "query" should be taken to mean text that is input for the purpose of selecting a subset of documents from a document database 33. A query typically contains a plurality of query terms in the form of a logical expression or a natural language. Each query term is a sequence of letters in a supported language, which sequence typically includes one or more words of the supported language. That is, each of the letter sequences (or query terms) that constitutes a query may be single words, compound words, phrases, and other multi-word constructs. A space may or may not be left between single words in a letter sequence depending on the language used.

[0030] Unless otherwise stated, the term "word" should be taken to include single words, compound words, phrases, and other multi-word constructs. Furthermore, the terms "word" and "term" are often used interchangeably. Terms and words include, for example, nouns, proper nouns, complex nominals, noun phrases, verbs, adverbs, numeric expressions, and adjectives.

[0031] In the following, we disclose:

(1) a similar word extractor;
(2) a query expander (31a) using the similar word extractor;
(3) a query expander (31b) which is adapted to a document database/retrieval engine 32 that supports a language different from the one used for the query (i.e., a query expander for cross-language retrieval);
(4) a letter sequence divider;
(5) a variant (or equivalent word) finder; and
(6) a preferred query expander (31c) based on the similar word extractor and incorporating the letter sequence divider and the variant finder.

[0032] The similar word extractor, the letter sequence divider and the variant finder are powerful tools for use in the query expander 31. The letter sequence divider and the variant finder are implemented with the similar word extractor as will be seen from the following description.

Similar Word Extractor

[0033] FIG. 2 is a diagram showing an exemplary arrangement of a similar word extractor system 1300 according to the invention. In FIG. 2, a one-way arrow drawn from a source to a destination indicates that data of the destination is obtained from the source, and a two-way arrow drawn between a block indicative of a program and a block indicative of a data structure indicates that the program refers to the data structure.

[0034] Prior to an actual operation of similar word extraction, it is preferable to prepare a second index file 134 and a feature vector table 160 from a second document database 33a for use in a similar word extraction operation.

[0035] Though the second document database 33a may be different from the document database 33 of the document database/retrieval engine 32, it is strongly recommended that the second document database 33a should be identical to the document database 33.Even if the second document database 33a is different from the document database 33, the second document database 33a preferably has the same characteristics that the document database 33 has in order to raise the accuracy (i.e., relevance) of extracted words. (For this reason, the second document database 33a is

hereinafter referred to as "the document database 33a" or simply "the database 33a".)

[0036] On the other hand, the second index file 134 has not necessarily to be the same as the index file 34 of the document database/retrieval engine 32. However, it is noted that the second index file 134 is preferably so arranged that when the second index file 134 is searched by a document retriever 136 for a query term contained in the query from a client 14, the document retriever 136 can obtain statistical information such as a frequency of occurrences of the query term in each of the documents in the database 33a (referred to as "document-based term frequency") and the number of documents that contain the query term even if the query term is any sequence of letters, i.e., even if the query term is such a word as can not found in a dictionary. Such index file 134 can be realized in accordance with the principles of Japanese unexamined patent publication No. Hei08-249354 by the present applicant.

[0037] The feature vector table 160 is generated as follows. It is assumed that the number of documents {Di Ii = 1, 2,..., M} in the database 33a is M; and a suitable set of words (WORD SET) 150 is given that contains terms (or words) {Tj I J = 1, 2,..., N}, where N is the total number of kinds of words in the database 33a. For each word Tj in the word set 150, a M-dimensional feature vector tj is calculated. Each feature vector tj is defined as:

$$tj = (w(j,1), w(j,2), ...w(j, i),...,w(j, M)) \quad (1)$$

[0038] As seen from expression (1), the elements of the feature vector tj, w(j,1), w(j,2), ...w(j, i),...,w(j, M) correspond to the documents D1, D2, ..., Di, ...DM. The i-th element w(j, i) of the feature vector tj is calculated as a weight based on a tf(j,i)*idf(j) score well-known in the art. "tf(j,i)" is the number of occurrences of a term Tj within a given document Di, and is referred to as "the term frequency" of term Tj in document Di. "idf(j)" is referred to as "the inverse document frequency" for term Tj, and is defined as

$$idf(i,j) = \log(M/dj).$$

where dj is the number of documents in which the term Tj occurs. (For further details of the tf(j,i)*idf(j) score, refer to G. Saltion and M.J. McGill, "Introduction to Modern Information Retrieval", McGraw-Hill Publishing Company, 1983.) In this case, the feature vector tj may be so normalized as to be 1 in magnitude or length. In this way, the feature vector table 160 is generated, and is now ready for a similar word extraction operation.

[0039] FIG. 3 is a flowchart showing the operation executed by the processor 20 under the control of the similar word extractor 130 according to the invention. The similar word extractor 130 is invoked in response to a reception of a query from the remote client 14. Though similar words are obtained for each query term in a given query in the following specific example, it should be noted that similarly words may be obtained for a given query itself in the same manner as described below.

[0040] If a k-th query term q_k in the received query is received, then the processor 20 searches the second index file 134 on the basis of the received query term q_k and outputs the term frequency of the query term, $tf(q_k,i)$, for each document Di and the inverse document frequency for the query term, $idf(q_k)$ in step 136. The processor 20 calculate a feature vector q_k for a given query term q_k from the inverse document frequency $idf(q_k)$ and the term frequencies of the query term, $tf(q_k,1)$, $tf(q_k,2)$,...., $tf(q_k,M)$ in step 138.

[0041] Step 140 marks the records which should be subjected to the following process in the feature vector table 160. In order to facilitate this step, it is preferable to configure the structure of the feature vector table 160 as shown in FIG. 4. Each record of the feature vector table 160 may contain one or more category fields (CATE 1, CATE 2,...) 168 for containing category code, subcategory code and so on and a mark field 166 for containing a special code indicating that the record is marked as well as the word field 162 and the feature vector fields 164. If it is desired to calculate the degrees of similarity only for the words of one or more specific field in the next step, then the special code is written in the mark fields 166 of the records in which the value of the category field 168 falls on the specific field. Also, if it is desired to exclude the query term itself from the following process, then the special code has only to be eliminated from the mark field 166 of the record for the query term q_k if the term q_k exists in the feature vector table 160, i.e., in the word set 150.

[0042] However, it should be noted that the query term q_k can not be always found in the word set 150 or the feature vector table 160 only because the inverse document frequency and the term frequencies have been successfully obtained for the query term q_k through the document retriever 136. This is because the second index file 134 is so arranged as to cause the inverse document frequency and the term frequencies to be obtained for any query term q_k as described above. In other words, such statistical information can be obtained even for a query term (or even for a query itself) which is not found in the word set 150 through a retrieval using the second index file 134.

[0043] Then, step 142 calculates the degree of similarity between the calculated feature vector q_k and each of the marked-record feature vectors in the feature vector table 160. The degree of similarity between feature vectors is calculated as an inner product of the feature vectors. For example, the degree of similarity between the query term "www" and the word "internet", i.e., SIM(www, internet) is calculated as follows:

SIM(www, internet) = 1.10*0.15 + 0.00*0.00 + 0.12*0.01 +.....+ 0.07*0.10
= 0.9

[0044] Alternatively, the degree of similarity between vectors cam be calculated by using statistic information such as the mutual information, t-score. For further details for the mutual information and t-score, refer to K. W. Church and R. L. Mercer, "Introduction to the Special Issue on Computational Linguistics Using Large Corpora", Computational Linguistics, Vol. 19, No. 1, 1993, pp. 1-24.

[0045] Step 144 ranks the marked-record words in the order of the calculated degrees of similarities. Step 146 outputs the words with a similarity higher than a predetermined value as the similar words, and end the operation. In FIG. 2, words "internet" and "HTML" are output as the similar words for the query term "www". Thus, Similar words can be obtained for any query term.

[0046] In the above embodiment, all of the M documents in the document database 33a has been used. However, the database 33a may be divided by the fields into document subsets for respective fields; an second index file 134 and a feature vector table 160 are prepared for each document subset; and, for a given field, the second index file 134 and the feature vector table 160 are used which are associated with the given field. Alternatively, a document database 33a, a second index file 134 and a feature vector table 160 may be prepared for each of the desired fields.

[0047] In the above embodiment, the second index file 134 and the feature vector table 160 have been prepared prior to an actual similar word extraction. However, if the word sets 150 contains very few words, then the second index file 134 and the feature vector table 160 may be generated after the calculation of feature vector in step 138.

[0048] Though the word set 150 has been used for generating the feature vector table 160, the feature vector table 160 may be generated directly from the database 33a.

[0049] The word set 150 may contain either all of the words that occurs in the document database 33a or only the words in one or more specific field. By limiting the vocabulary of the word set 150, the records in the feature vector table 160 can be limited accordingly.

Query Expander Using the Similar Word Extractor

[0050] FIG. 5 is a flowchart showing an operation of a query expander using the similar word extractor 130 of FIG. 3, which query expander 31a is an illustrative embodiment of the query expander 31 of FIG. 1. In FIG. 5, step 102 selects a first query term from a given query. Step 104 finds similar words for the selected query term by using the similar word extractor 130. Step 106 replaces the selected query term in the query with, for example, a logical sum of the found similar words. Decision step 108 makes a test to see if the terms have been exhausted (or replaced) in the query. If not, then the control is passed to step 110, where a next term is selected in the given query and the control is returned to step 104. If the test result is YES in step 108, then the control is passed to step 112, where the expanded query is passed to the document retriever 36, and the operation is terminated.

[0051] On receiving the expanded query, the document retriever 36 searches the index file 35 on the basis of the received expanded query in a conventional manner.

[0052] For example, if a user at a client enters (www (and) "search engine") as a query in order to know about search engines on the WWW (world wide web), then the similar word extractor 130 will provide "internet" as a similar word for the query term "www" and "retrieval" as a similar word for the query term "search engine", and accordingly the query expander 31a expands the query (www (and) "search engine") to provide an expanded query ((www (or) internet) (and) ("search engine" (or) retrieval)). This enables an inclusion of a document with an expression "retrieval services on the internet" in the retrieval results.

[0053] In the query expander 31a, all of the query terms have been expanded with the similar word extractor 130. Alternatively, only terms of the query terms that are not found in the word set 150 may be expanded with the word extractor 130.

[0054] The similar words found for a selected query term in step 104 have been used in place of the selected (or original) term in the query in step 106. Alternatively, the found similar words may be added to the original term.

[0055] Though the query expander 31a has expanded each of the query word in the given query, the query expander 31a may be so configured as to generate a single retrieval condition for similar word extraction, pass the generated retrieval condition to the similar word extractor 130 to obtain the similar words, and pass the obtained similar words to the document retriever 36.

Query Expander for Cross-language Retrieval

[0056] FIG. 6 is a diagram showing an arrangement of a translation extractor system 1300a for use in a query expander for cross-language retrieval. As seen from FIG. 6, the translation extractor system 1300a is closely similar to the similar word extractor system 1300 of FIG. 2.

[0057] The two extractor systems 1300a and 1300 are identical in an actual extraction operation except that the translation extractor system 1300a uses a source-language index file 235 instead of the second index file 134. The source-language (SL) index file 235 is generated from a SL document database 233. For this reason, the similar word extractor 130 shown as the flowchart in FIG. 3 can be used for the translation extraction operation by replacing the second index file 134 with the SL index file 235. Also, in this case, if more than one translation is to be output in step 146, then the translations are preferably output in the form of logical product.

[0058] Also, what differs from the similar word extractor system 1300 is that the feature vector table 160 is generated from a given target-language (TL) word set 350 and a TL index file 335 which is generated from a TL document database 333. The TL document database 333 of FIG. 6 is preferably an identical to the document database 33 of FIG. 1. However, though the TL document database 333 may be different from the document database 33 of FIG. 1, if so, then the two databases 333 and 33 should be of the same language and the same field.

[0059] The SL document database 233 and the TL document database 333 has to be translations of each other. If a translated version of one document database can not obtained, one document database 233 or 333 may be obtained through a machine translation of the other document database 333 or 233, respectively.

[0060] FIG. 7 is a flowchart of a program 31b which operates as a query expander for cross-language retrieval when executed in combination with the translation extractor 130 in the translation extractor system 1300a. The query expander program 31b is executed in response to a reception of a query. In FIG. 7, the processor 20 convert the given query into a sum-of-product form in step 122. Step 124 passes each of the products in the sum to the extractor 130 to obtain the results (translations in this case). Then, step 126 passes the logical sum of the results to the document retriever 36.

[0061] In this way, the query expander 31b is adapted to a document database/retrieval engine 32 that supports a language (i.e., the target language) different from the one used for the query (i.e., the source language). It is noted that the query expanders 31a and 31b can be used interchangeably and can be used for both the query expansion in a monolingual retrieval and the query expansion for cross-language retrieval. Though the given query is converted into a sum-of-product form in step 122, the given query may be passed as it is to the query expander 130.

[0062] As seen from FIG. 6, if a use at a client 14 enters

(情報検索 <or> 情報

抽出)

in Japanese, which corresponds to ("information retrieval" (or) "information extraction") in English, then the extractor 130 in the translation extractor system 1300a provides translations "information" and "retrieval" for the query term

"情報検索"

and also provides translations "information" and "extraction" for the query term

"情報抽出".

If more than one translation is extracted for a query term, the extractor 130 of the translation extractor system 1300a outputs the translations in the form of logical product as described above. Accordingly, a retrieval condition

((information (and) retrieval) (or) (information (and) extraction)) is passed to the document retriever 36.

[0063] As seen from FIG. 6, once the feature vector table 160 and the SL index file 235 have been generated, the SL and TL databases 233 and 333, the TL word set 350 and the TL index file 335 are no longer necessary. Preparing feature vector tables 160 for various language by machine translating a document database in one language, a query expander adapted to multilingual document retrieval.

Letter Sequence Divider

[0064] FIG. 8 is a flowchart showing an exemplary operation of a letter sequence divider 200 according to the

invention. If the letter sequence divider 200 receives a query term, i.e., a sequence of letters (e.g., "L1L2L3.....Ls" where s is the number of letters), then step 202 forms a retrieval expression

(L1 (and) L2 (and) L3.....(and)Ls) (not)L1L2L3.....Ls.

[0065] FIG. 9 shows an example of such retrieval expression for a letter sequence

"自律移動ロボット",

which is a Japanese term that means "autonomous mobile robot".

[0066] Step 204 retrieves the documents that satisfy the retrieval expression and obtain statistical information such as described in step 136. Step 206 generates a feature vector for the retrieval expression (referred to as "mask feature vector"). The documents that satisfy the retrieval expression are such documents as include any of the letters of the letter sequence but does not include the letter sequence itself. For example, a document containing only an expression

"自律的に移動するロボット"

is selected, while a document containing only an expression

"自律移動ロボットについて"

is not selected.

[0067] Step 208 divides the letter sequence into two parts shifting the division point one by one to obtain s-1 sets of divided sequences. The division may be achieved by means of morpheme analysis, if it yields divided subsequences. FIG. 10 shows the seven sets of divided sequences for the letter sequence

"自律移動ロボット".

For each divided sequence in each set, step 210 calculates a feature vector. For a set of divided sequences

(自、律移動ロボット)

for example, step 210 calculates a feature vector for a subsequence

"自"

and a feature vector for a subsequence

"律移動ロボット".

The feature vector for a subsequence

"自"

have positive values in elements thereof corresponding to dimensions for the documents that include the subsequence

"自".

For example, the feature vector for the subsequence

"自"

has a positive value in a dimension for the document that contains only

"自由の追求".

[0068] Then, step 212 masks each of the calculated feature vectors with the mask feature vector obtained in step 206, and normalizes the masked vector so as to be 1 in length. Masking a feature vector with the mask feature vector means leaving, as they are, elements of the feature vector that correspond to such elements of the feature vector as have positive values and making the other elements of the feature vector zero. FIG. 11 is diagram showing an example of the processing of step 212. In FIG. 11, the feature vectors for a set

(自、律移動ロボット)

are masked with the mask feature vector. In the figure, the elements or dimensions with a value of 0 is denoted with "o", and the elements or dimensions with a positive value is denoted with "x". In the resultant feature vectors in this example, dimensions for a document that contains only

"自由の

追求",

a document that contains only

"自律移動ロボット"

and so on have a value of 0; and only the dimensions for the documents in which both of the letter subsequences

"自" and "律移動ロボット"

occur independently without succeeding each other have positive values.

[0069] Step 214 finds the degree of similarity for each set by calculating an inner product of the feature vectors of the set. FIG. 12 is a diagram showing the divided sequence sets, the respective degrees of similarity, and respective division proprieties judged by a threshold of 0.35. According to this embodiment, if the divided sequences of a letter sequence set occur in common more frequently in a document, the degree of similarity between the divided sequences of the letter sequence set becomes the higher. Since the documents that contain a letter sequence as it is are excluded in steps 202 and 212, frequent independent occurrences of the divided sequences yields a higher degree of similarity. Independently occurring letter sequences are thought to be ones that make sense by themselves.

[0070] Step 216 outputs divided sequences of the letter sequences whose degrees of similarity is equal to or more than the threshold. Since the threshold is set to 0.35 in the example of FIG. 12, two of the divided sequence sets, i.e.,

(自

律、移動ロボット) and (自律移動、ロボット)

are judged to be of good propriety in division and accordingly output.

[0071] As described above, a letter sequence divider 200 according to the invention tries all possible division of a given letter sequence to yield bisected letter sequences with a sufficient propriety of division.

[0072] In the path with a dot labeled "A" in FIGS. 5 and 7, the letter sequence divider 200 can be utilized. In FIG. 5 for example, if the term is not found in the word set 150, then bisected letter sequences are obtained from the term by using the letter sequence divider 200 and used in place of the term.

Variant Finder

[0073] FIG. 13 is a flowchart showing an exemplary operation of a variant finder 300 according to the invention. In FIG. 13, step 302 attaches a first-language index file and a second-language feature vector table 160a to the translation extractor 130 of FIG. 3. Step 304 finds second-language translation(s) of an input first-language query term by using the translation extractor 130. If the input first-language query term is, e.g., an Japanese term

"

ギリシャ",

which corresponds to an English word "Greek", then it is assumed that a second-language (say, English) translations "GREECE" is obtained as the translation.

[0074] Then, step 306 replaces the first-language index file and the second-language feature vector table with a second-language index file and a first-language feature vector table, respectively in the translation extractor 130. Step 308 finds first-language translation(s) of each the second-language translation(s) by using the translation extractor 130. In this step, an input term "GREECE" will cause three translations

"ギリシャ", "ギリシヤ" and "ギリ

シア"

to be output. In this case, these three variants are obtained for the input term

"ギリシャ".

[0075] If a word has variant(s), then an identical expression is usually used in a document. Since variants are hardly used in a single document, though the variants are closely similar to each other, the similar word extractor 130 can not usually extract such a variant.

[0076] According to the variant finder 300 of the invention, a letter sequence the second-language translation of which is identical to the second-language translation of an input term can be obtained as a variant of the input term. In other words, if a word has one or more variant in a first language and a second language has only one expression that corresponds to the word, then, for a member of such a word group, the other variant(s) can be obtained by using the variant finder 300.

[0077] In the path with a dot labeled "B" in FIGS. 5 and 7, the variant finder 300 can be utilized. Specifically, the

obtained query is searched for words written in Japanese "katakana" letters which words tend to have variant(s). For each of the found words, the variant(s) are found by using the variant finder 300, and a logical sum of the found variant(s) and the original word are used in place of the original word. Doing this enables a further expansion of the query, resulting in an effective document retrieval.

[0078] The disclosures of all articles and references, including patent documents, mentioned in this application are incorporated herein by reference as if set out in full.

[0079] Many widely different embodiments of the present invention may be constructed without departing from the spirit and scope of the present invention. It should be understood that the present invention is not limited to the specific embodiments described in the specification, except as defined in the appended claims.

Claims

1. A method of extracting similar words for a given queryfrom a document database of documents, the occurrence pattern of the similar words being similar to that of the given query, the method comprising the steps of:

   preparing a feature vector table which contains a first feature vector for each of words in said document database;
   obtaining statistical information on said given query through a retrieval of said document database with respect to said given query;
   calculating a second feature vector with respect to said given query from said statistical information;
   calculating a degree of similarity between said second feature vector and each of said first feature vectors; and
   selecting such words as yield degrees of similarity higher than a predetermined value.

2. A method of claim 1, wherein said step of preparing a feature vector table comprises the steps of:

   generating an index file from said document database, said index file being such that statistical information for each of words used in said document database can be obtained from said index file; and
   calculating each first feature vector from said index file, and
   wherein said step of obtaining statistical information comprising the step of searching said index file for said given query.

3. A method of extracting similar words for a given sequence of letters from a document database of documents, the occurrence pattern of the similar words being similar to that of the given sequence, the method comprising the steps of:

   preparing a feature vector table which contains a first feature vector for each of words in said document database;
   obtaining statistical information on said given sequence through a retrieval of said document database with respect to said given sequence;
   calculating a second feature vector with respect to said given sequence from said statistical information;
   calculating a degree of similarity between said second feature vector and each of said first feature vectors; and
   selecting such words as yield degrees of similarity higher than a predetermined value.

4. In a document retrieval system comprising a first document database and a retriever, a method of expanding a given query to supply a expanded query to the retriever, the method including the steps of:

   extracting similar words of said given query or a query word in said given query from a second document database of documents, the occurrence pattern of said similar words being similar to that of said given query or said query word; and
   utilizing said similar words of said given query or said query word for said given query, wherein said step of extracting said similar words comprises the steps of:
   preparing a feature vector table which contains a first feature vector for each of words in said second document database;
   obtaining statistical information on said query or said query word through a retrieval of said second document database with respect to said query or said query word;
   calculating a second feature vector with respect to said query or said query word from said statistical information;
   calculating a degree of similarity between said second feature vector and each of said first feature vectors; and

selecting, as said similar words, such words as yield degrees of similarity higher than a predetermined value.

5. A method of claim 4, wherein said first and second document databases are the same thing.

6. A method of receiving any given sequence of letters in a source language and providing translations in a target language such that an occurrence pattern of said given sequence in a source-language document database being the same as that of translations in a target-language document database wherein one of said two document databases is a translation of the other, the method comprising the step of:

preparing a feature vector table which contains a first feature vector for each of words in said target-language document database;
obtaining statistical information on said given sequence through a retrieval of said source-language document database with respect to said given sequence;
calculating a second feature vector with respect to said given sequence from said statistical information;
calculating a degree of similarity between said second feature vector and each of said first feature vectors; and
selecting, as said translations, such target-language words as yield degrees of similarity higher than a predetermined value.

7. A method of claim 6, wherein said step of preparing a feature vector table comprises the steps of:

generating an index file from said target-language document database, said index file being such that statistical information for each of words used in said target-language document database can be obtained from said index file; and
calculating each first feature vector from said index file, and
wherein said step of obtaining statistical information comprising the step of searching said index file for said given sequence.

8. In a document retrieval system comprising a first document database and a retriever, a query expanding method of converting a source-language query into a target-language query to feed the retriever, the method including the steps of:

extracting a translation of each of query words in said source-language query from a second document database of target-language documents, the occurrence pattern of said translation being similar to that of each query word; and
replacing each query word in said source-language query with said translation of said query word, wherein said step of extracting said translation comprises the steps of:
preparing a feature vector table which contains a first feature vector for each of words in said target-language document database;
obtaining statistical information on said query word through a retrieval of said source-language document database with respect to said query word;
calculating a second feature vector with respect to said query word from said statistical information;
calculating a degree of similarity between said second feature vector and each of said first feature vectors; and
selecting, as said translations, such target-language words as yield degrees of similarity higher than a predetermined value.

9. A method of receiving a given letter sequence and providing bisected letter sequences with a sufficient propriety of division, the method comprising the steps of:

forming a retrieval expression for retrieving documents which contains all of the letters of said letter sequence but does not contain said letter sequence itself;
obtaining statistical information by executing a retrieval based on said retrieval expression;
generates a feature vector for said retrieval expression from said statistical information, said feature vector being referred to as "mask feature vector";
dividing said letter sequence into two subsequences shifting the division point one by one to obtain N-1 sets of bisected letter sequences, where N is the number of letters in said letter sequence;
calculate a feature vector for each of said bisected letter sequences of each of said N-1 sets;
masks each of said calculated feature vectors with the mask feature vector and normalizing the masked calculated feature vector;

calculating a degree of similarity for each of said N-1 sets, said degree of similarity being defined as an inner product of said bisected letter sequences of each of said N-1 sets; and

outputting such bisected letter sequences as have a degree of similarity larger than a threshold.

10. In a system provided with translation means for receiving a given word in a first language and providing translations in a second language such that an occurrence pattern of said given word in a first-language document database being the same as that of translations in a second-language document database wherein one of said two document databases is a translation of the other, wherein said translation means comprise a feature vector table which contains a first feature vector for each of words in said second-language document database; means for obtaining statistical information on said query word through a retrieval of said first- language document database with respect to said query word; means for calculating a second feature vector with respect to said query word from said statistical information; means for calculating a degree of similarity between said second feature vector and each of said first feature vectors; and means for selecting, as said translations, such second-language words as yield degrees of similarity higher than a predetermined value, a method of finding, for a given word, at least one variant including the given word, the method comprising the steps of:

obtaining said translations for said given word by using said translation means;
reconfiguring said translation means by replacing said feature vector table with a feature vector table which contains a third feature vector for each of words in said first-language document database and by exchanging said first-language document database and said second-language document database; and
obtaining, as said at least one variant, second translations for each of said translations by using reconfigured translation means.

11. A system of extracting similar words for a given query from a document database of documents, the occurrence pattern of the similar words being similar to that of the given query, the system comprising:

means for preparing a feature vector table which contains a first feature vector for each of words in said document database;
means for obtaining statistical information on said given query through a retrieval of said document database with respect to said given query;
means for calculating a second feature vector with respect to said given query from said statistical information;
means for calculating a degree of similarity between said second feature vector and each of said first feature vectors; and
means for selecting such words as yield degrees of similarity higher than a predetermined value.

12. In a document retrieval system comprising a first document database and a retriever, a system for expanding a given query to supply a expanded query to the retriever, the system comprising:

means for extracting similar words of said given query or a query word in said given query from a second document database of documents, the occurrence pattern of said similar words being similar to that of said given query or said query word; and
means for utilizing said similar words of said given query or said query word for said given query, wherein said means for extracting said similar words comprises:
means for preparing a feature vector table which contains a first feature vector for each of words in said second document database;
means for obtaining statistical information on said query or said query word through a retrieval of said second document database with respect to said query or said query word;
means for calculating a second feature vector with respect to said query or said query word from said statistical information;
means for calculating a degree of similarity between said second feature vector and each of said first feature vectors; and
means for selecting, as said similar words, such words as yield degrees of similarity higher than a predetermined value.

13. A method of claim 12, wherein said first and second document databases are the same thing.

FIG. 1

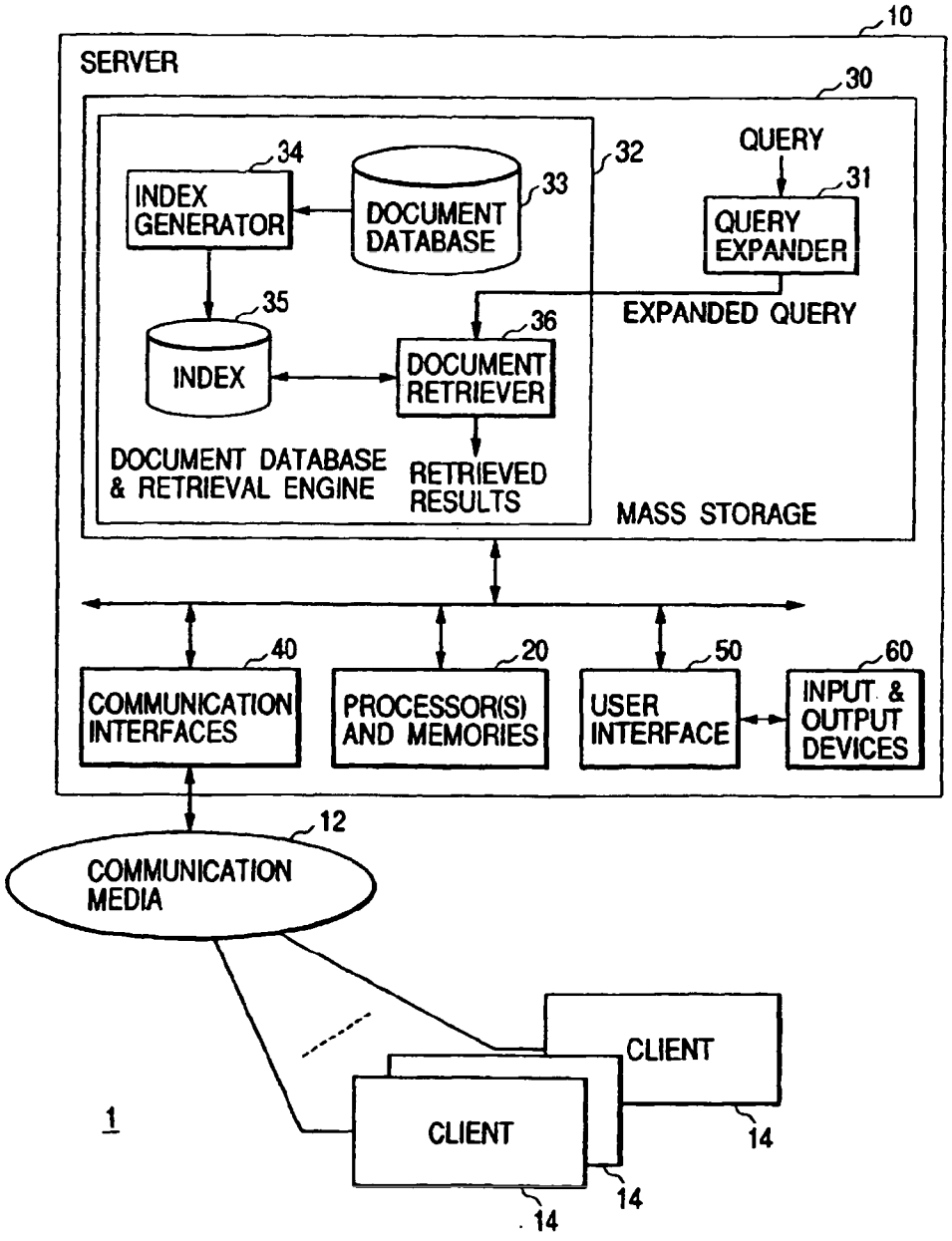


FIG. 2

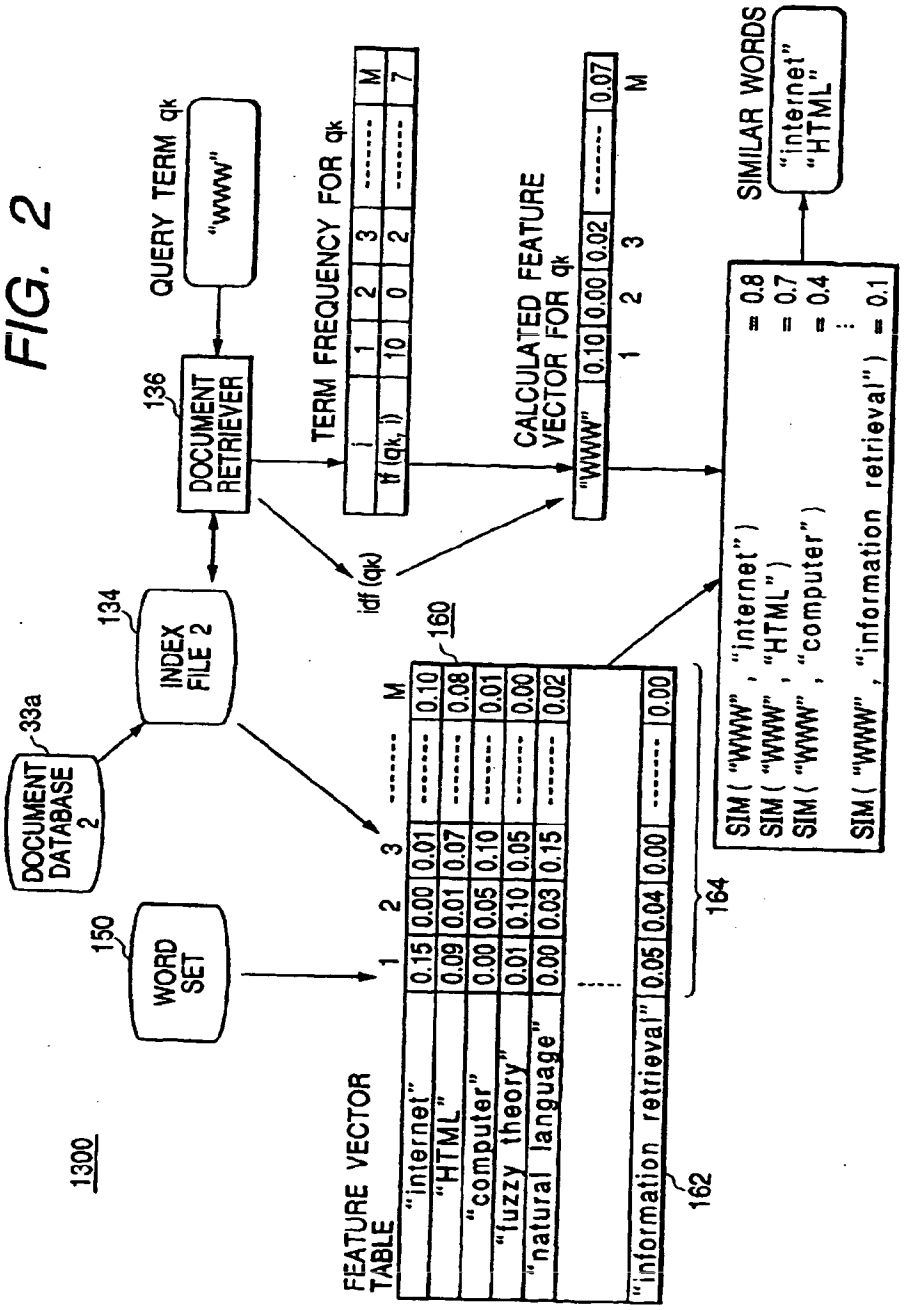


FIG. 3

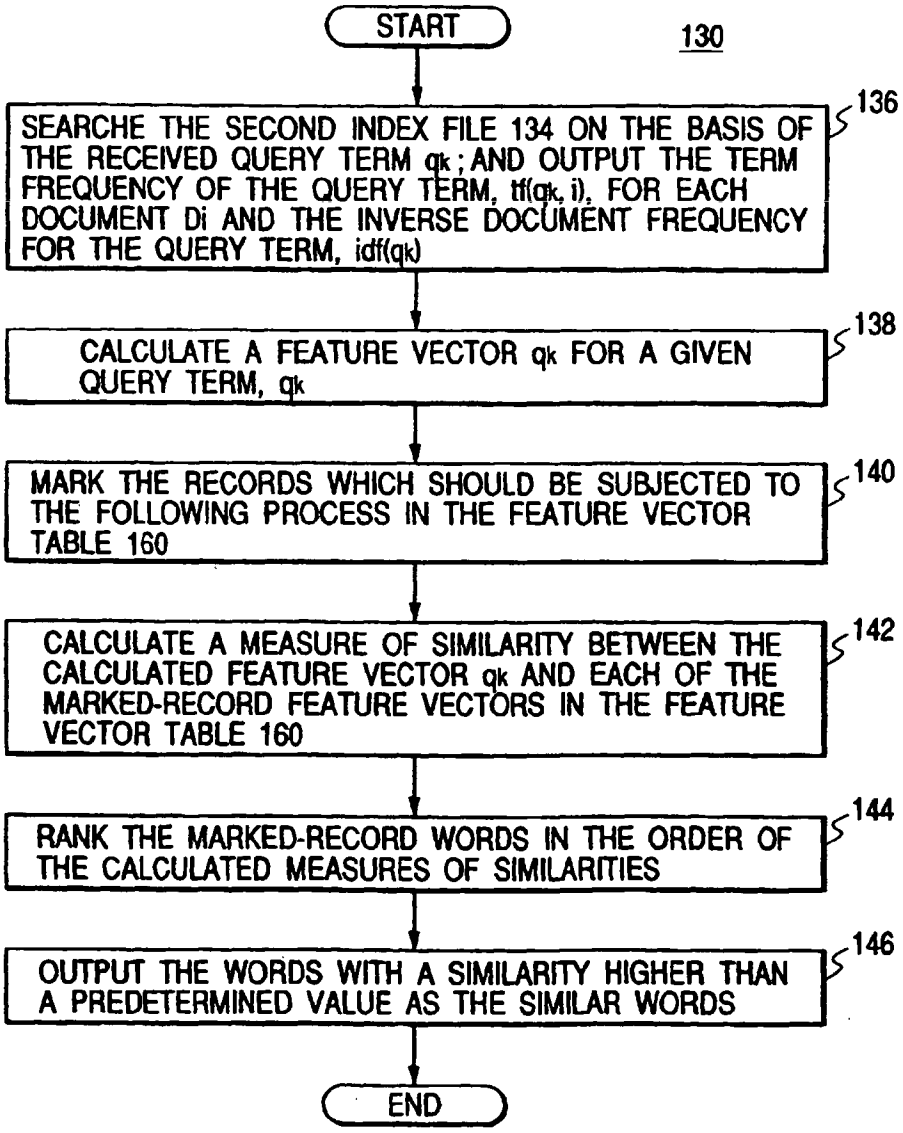


FIG. 4

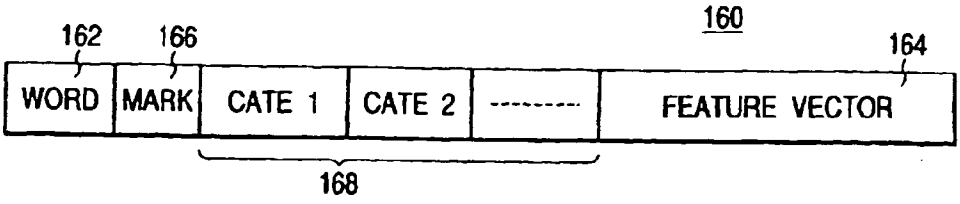


FIG. 5

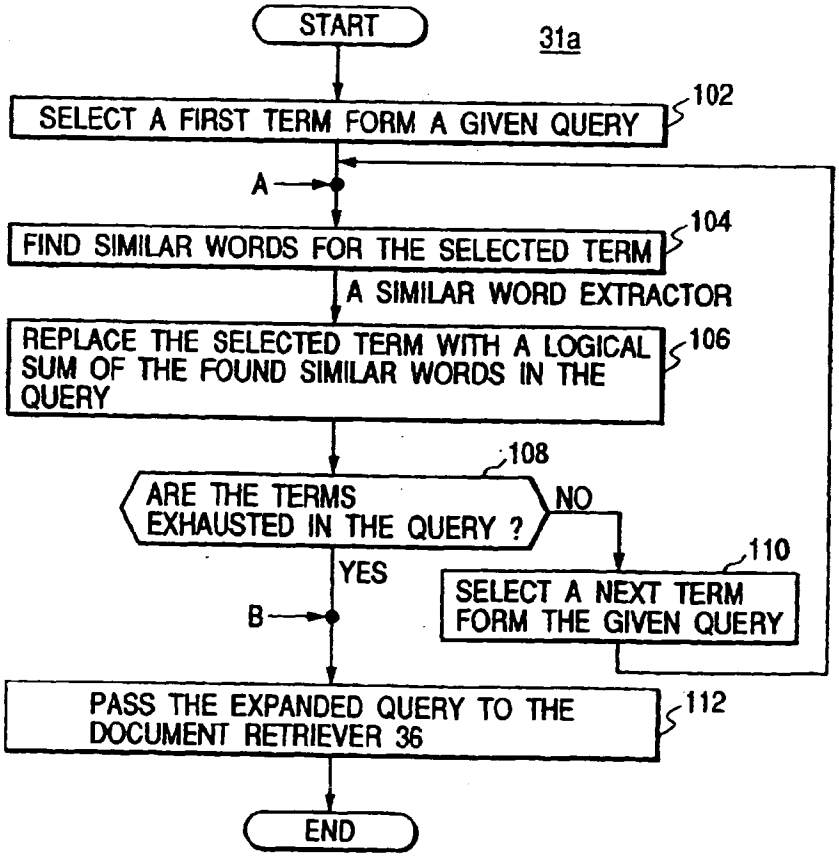


FIG. 6

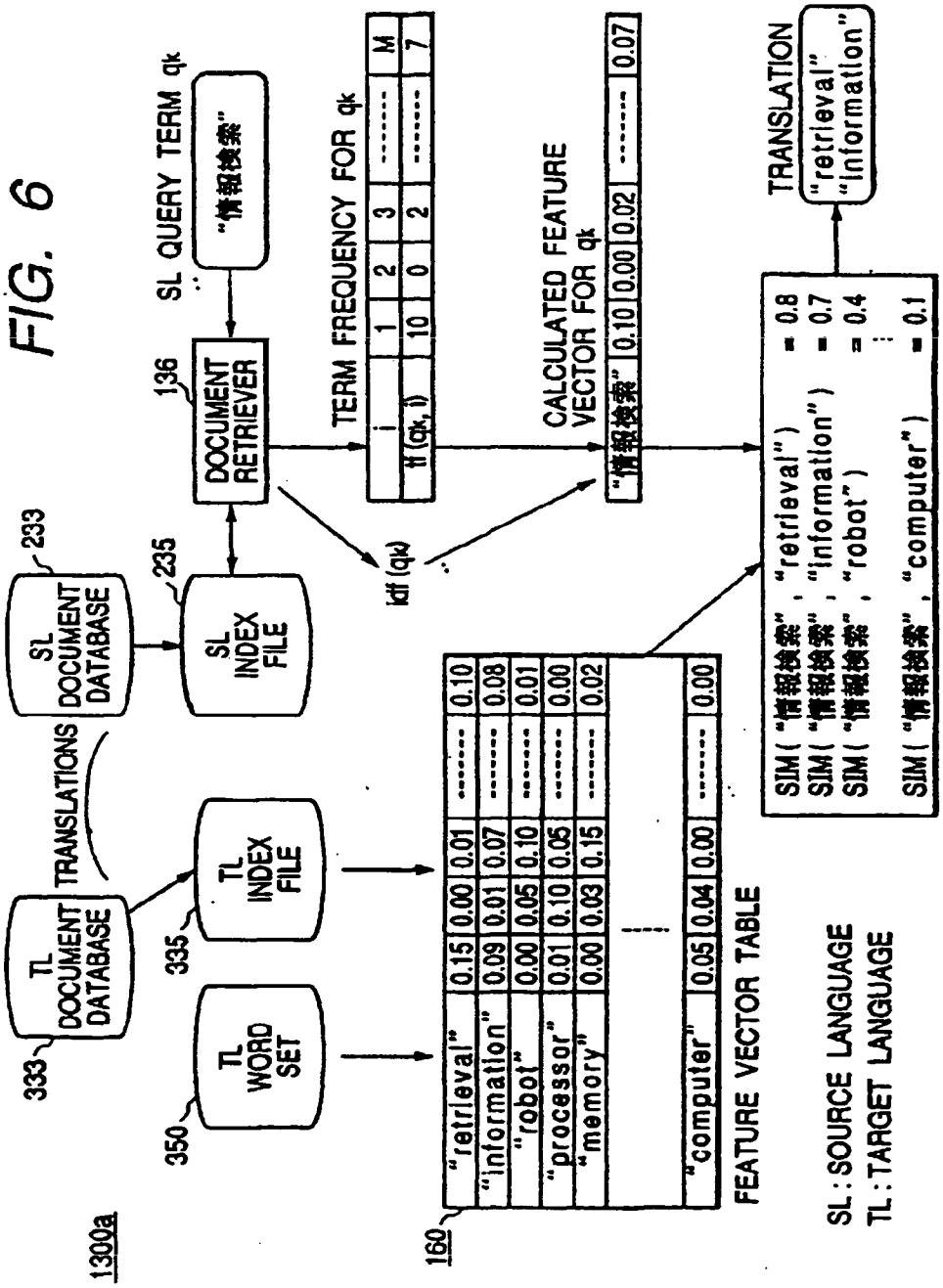


FIG. 7

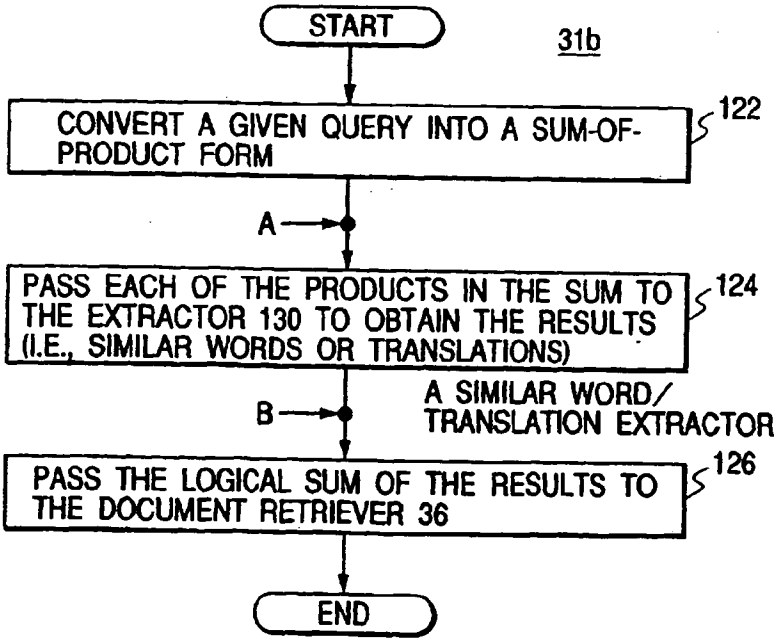


FIG. 8

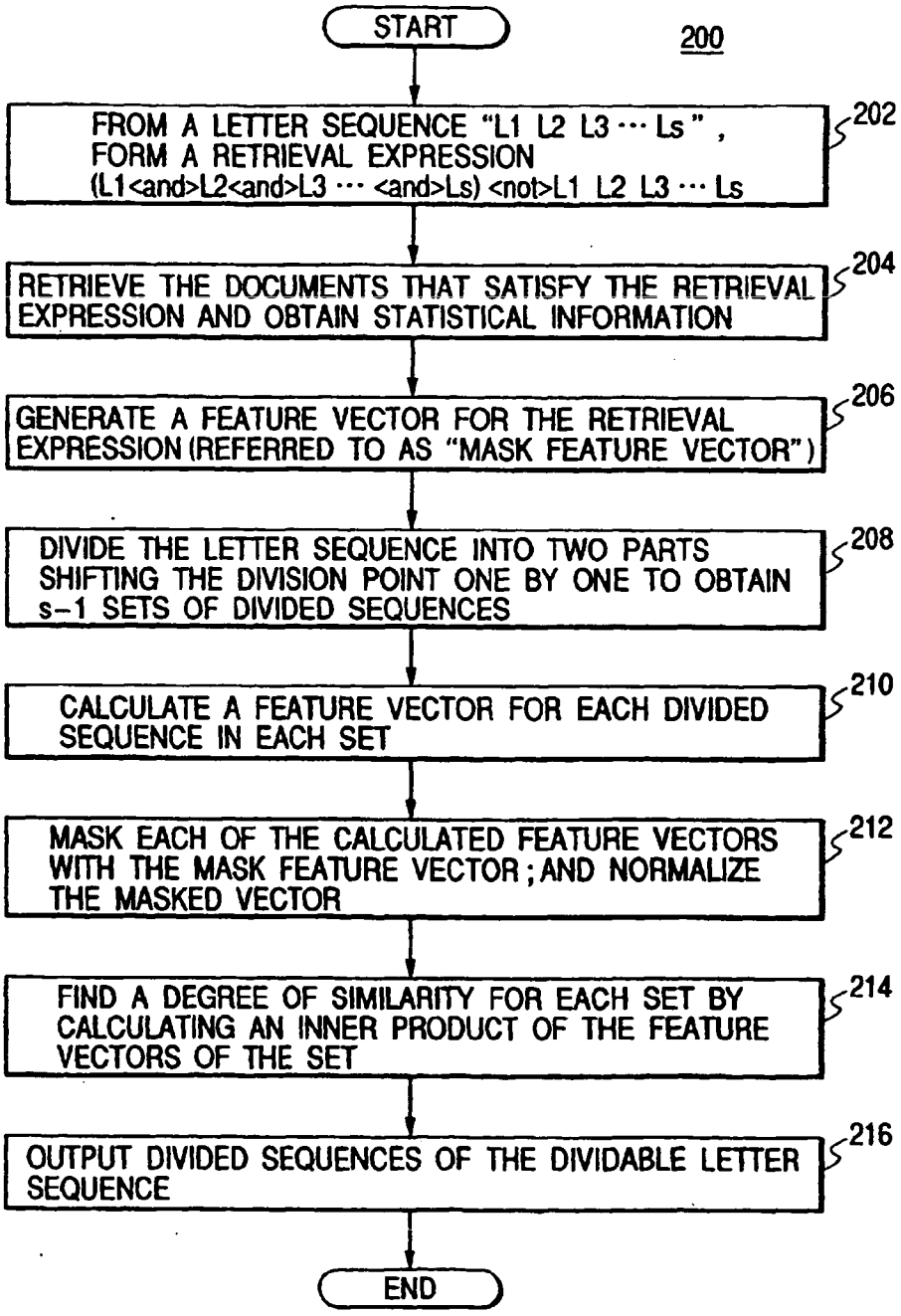


FIG. 9

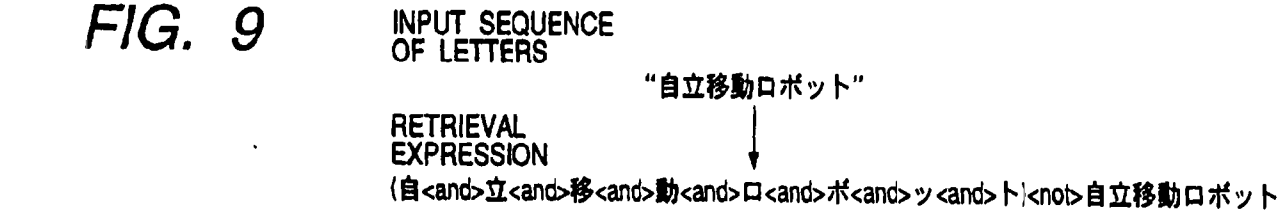


FIG. 10

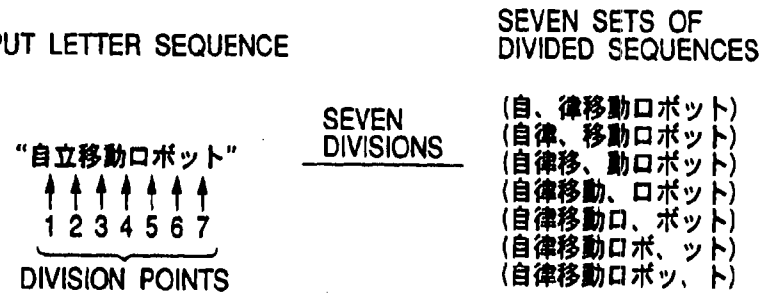
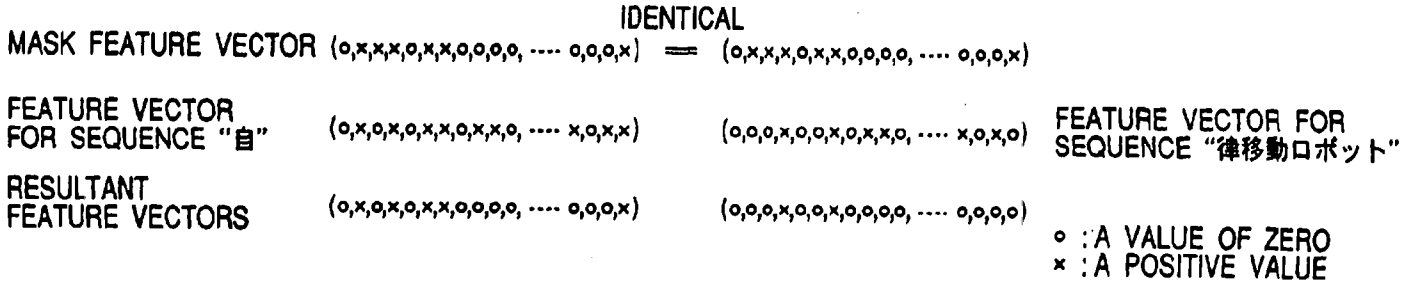


FIG. 11

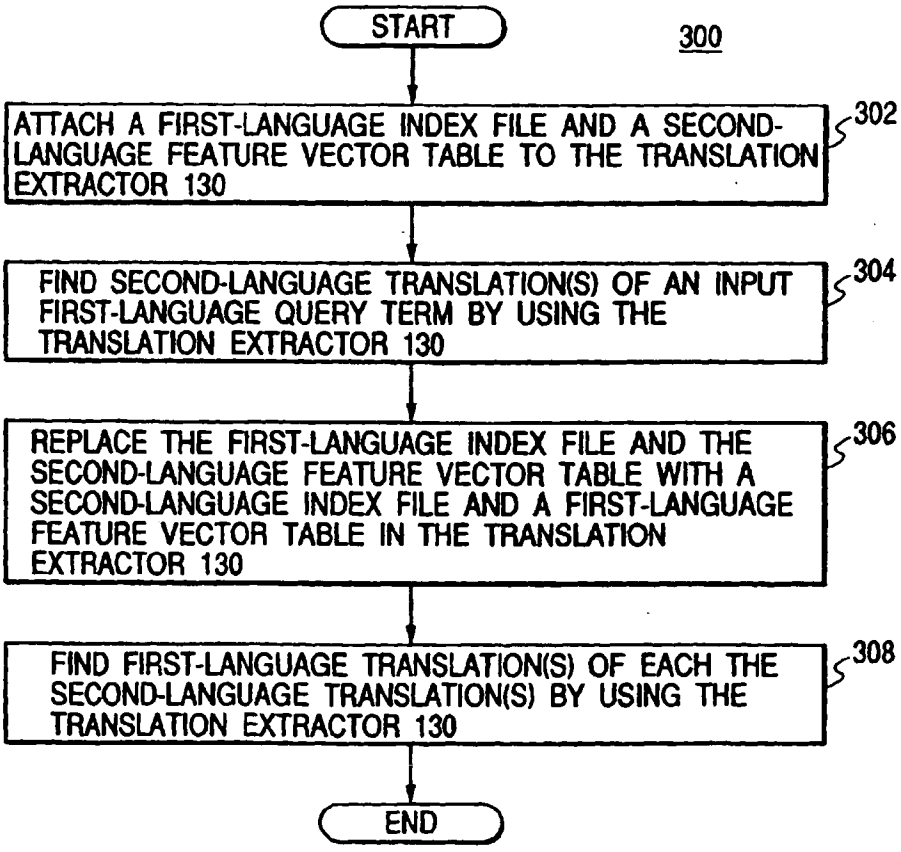


EP 1 072 982 A2

FIG. 12

	DEGREE OF SIMILARITY	DIVISION PROPRIETY	
(自、律移動ロボット)	0.01	NO	
(自律、移動ロボット)	0.40	YES	
(自律移、動ロボット)	0.02	NO	
(自律移動、ロボット)	0.50	YES	
(自律移動ロ、ボット)	0.02	NO	THRESHOLD=0.35
(自律移動ロボ、ット)	0.01	NO	
(自律移動ロボッ、ト)	0.02	NO	

FIG. 13



(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 095 326 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
30.01.2002 Bulletin 2002/05

(21) Application number: **99933296.8**

(22) Date of filing: **09.07.1999**

(51) Int Cl.[7]: **G06F 1/00**

(86) International application number:
PCT/NO99/00233

(87) International publication number:
WO 00/03315 (20.01.2000 Gazette 2000/03)

(54) **A SEARCH SYSTEM AND METHOD FOR RETRIEVAL OF DATA, AND THE USE THEREOF IN A SEARCH ENGINE**

EIN SUCHSYSTEM UND VERFAHREN ZUM ZURÜCKHOLEN VON DATEN UND DIE ANWENDUNG IN EINEM SUCHGERÄT

SYSTEME ET PROCEDE EN VUE DE LA RECUPERATION DE DONNEES ET SON UTILISATION DANS UN AUTOMATE DE RECHERCHE

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **10.07.1998 NO 317598**

(43) Date of publication of application:
02.05.2001 Bulletin 2001/18

(73) Proprietor: **Fast Search & Transfer ASA**
0120 Oslo (NO)

(72) Inventor: **RISVIK, Knut, Magne**
N-7014 Trondheim (NO)

(74) Representative: **Liesegang, Eva**
Forrester & Boehmert, Pettenkoferstrasse 20-22
80336 München (DE)

(56) References cited:
WO-A1-92/15954
WO-A1-96/00945
US-A- 5 511 159
US-A- 5 627 748

Printed by Jouve, 75001 PARIS (FR)

EP 1 095 326 B1

Description

[0001] The present invention concerns a search system for information retrieval, particularly information stored in form of text, wherein a text *T* comprises words and/or symbols *s* and sequences thereof *S*, wherein the information retrieval takes place with a given or varying degree of matching between a query *Q*, wherein the query *Q* comprises words and/or symbols *q* and sequences *P* thereof, and retrieved information *R* comprising words and/or symbols and sequences thereof from the text *T*, wherein the search system comprises a data structure for storing at least a part of the text *T*, and a metric *M* which measures the degree of matching between the query *Q* and retrieved information *R*, and wherein the search system implements search algorithms for executing a search, particularly a full text search on the basis of keywords *kw*, and a method in a search system for information retrieval, particularly information stored in form of text, wherein a text *T* comprises words and symbols *s* and sequences *S* thereof, wherein the information retrieval takes place with a given or varying degree of matching between a query *Q*, wherein the query *Q* comprises words and/or symbols *q* and sequences *P* thereof, and retrieved information *R* comprising words and/or symbols and sequences thereof from the text *T*, wherein the search system comprises a data structure for storing at least a part of the text *T*, and a metric *M* which measures the degree of matching between the query *Q* and retrieved information *R*, and wherein the search system implements search algorithms for executing a search, particularly a full text search on the basis of keywords *kw*, wherein the information in the text *T* is divided into words *s* and word sequences *S*, the words being substrings of the entire text separated by word boundary terms and forming a sequence of symbols, and wherein each word is structured as a sequence of symbols.

[0002] The invention also concerns the use of the search system.

[0003] A tremendous amount of information in various fields of human knowledge is collected and stored in computer memory systems. As the computer memory systems increasingly are linked in public available data communication networks, there has been an increasing effort to develop systems and methods for searching and retrieving information for public or personal use. Present search methods for data have, however, limitations that seriously reduce the possibility of efficiently retrieving and using information stored in this manner.

[0004] Information may be stored in the form of different data types, and in the context of information search and retrieval it will be useful to discern between dynamic data and static data. Dynamic data is data that change often and continuously, so that the set of valid data varies all the time, while static data only changes very seldom or never at all. For instance will economic data, such as stock values, or meteorological data be subject to very quick changes and hence dynamic. On the other hand archival storage of books and documents are usually permanent and static data. The concept the volatility of the data relates to how long the information is valid. The volatility of data has some bearing upon how the information should be searched and retrieved. Large volumes of data require some structure in order to facilitate searching, but the time cost of building such structures must not be higher than the time the data are valid. The cost of building a structure is dependent on the data volume and hence the building of data structures for searching the information should take both the data volume and the volatility into consideration. The information collected are stored in databases and these may be structured or unstructured. Moreover, the databases may contain several types of documents, including compound documents which contain images, video, sound and formatted or annotated text. Particularly structured databases are usually furnished with indexes in order to facilitate searching and retrieving the data. The growth of the World Wide Web (WWW) offers a steadily growing collection of compound and hyperlinked documents. A great many of these are not collected in structured databases and no indexes facilitating rapid searching are available. However, the need for searching documents in the World Wide Web is obvious and as a result a number of so-called search engines has been developed, enabling searching at least parts of the information in the World Wide Web.

[0005] With a search engine it is commonly understood one or more tools for searching and retrieving information. In addition to the search system proper, a search engine also contains an index, for instance comprising text from a large number of uniform resource locators (URLs). Examples of such search engines are Alta Vista, HotBot with Inktomi technology, Infoseek, Excite and Yahoo. All these offer facilities for performing search and retrieval of information in the World Wide Web. However, their speed and efficiency do by no means match the huge amount of information available on the World Wide Web and hence the search and retrieval efficiency of these search engines leaves much to be desired.

[0006] Searching a large collection of text documents can usually be done with several query types. The most common query type is matching and variants of this. By specifying a keyword or set of keywords that has to be present in the queried information the search system retrieves all documents that fulfils this requirement. The basic search method is based on so-called single keyword matching. The keyword *p* is searched for and all documents containing this word shall be retrieved. It is also possible to search for a keyword prefix p_j and all documents where this prefix is present in any keyword in the documents, will be retrieved. Instead of searching with keywords, the search is sometimes based on so-called exact phrase matching, where the search uses several single keywords in particular sequence. As well-known by persons skilled in the art, the exact matching of keyword phrases in many search systems may be done with

the use of Boolean operators, for instance based on operators such as AND, OR, and NOT which allow a filtering of the information; e.g. using an AND phrase results in that all documents containing the two keywords linked by the AND operator will be returned. Also a NEAR operator has been used for returning just the documents with the keywords matching and located "near" to each other in the document text. In many structured database the documents contained in the database have been annotated, e.g. provided with fields which denote certain parts or types of information in the document. This allows the search for matches in only parts of the documents and is useful when the type of queried information is known in advance.

[0007] When searching in text documents the data are structured and most likely present in some natural language, likc English, Norwegian etc. When searching for documents with a certain context it is possible to apply proximity metrics for matching keywords or phrases that match the query approximately. Allowing errors in keywords and phrases are common method for proximity, using a thesaurus is another common method. A proximity search requires only that there shall be a partial match between the information retrieved and the query. International published application WO96/00945 titled "Variable length data sequence matching method and apparatus" (Döringer & al.) which has been assigned to International Business Machines, Corp., discloses the building, maintenance and use of a database with a trie-like structure for storing entries and retrieving at least a partial match, preferably the longest partial match or all partial matches of a search argument (input key) from the entries.

[0008] In order to further illuminate the general prior art mention can be made of international published patent application WO92/15954 (Kimball & al., assigned to Red Brick System, U.S.A.) and US patent no. 5 627 748 (Baker & al., assigned to Lucent Technologies, Inc., U.S.A.), both disclosing data structures in the form of suffix trees for searching/matching in a square matrix. Neither of these two publications disclose anything beyond a regular suffix tree, except for the use of a linked list during matching and do not teach or suggest approaches to limit the search space when searching for approximate matches. However, such approaches would be most desirable when applying data structures based on suffix trees to searching, particularly for approximate matches in extremely large document collections, such as may be found on the World Wide Web.

[0009] The main object of the present invention is thus to provide a search system and a method for fast and efficient search and retrieval of information in large volumes of data. Particularly it is an object of the present invention to provide a search system suited for implementing search engines for searching of information systems with distributed large volume data storage, for instance Internet. It is to be understood that the search system according to the invention by no means shall be limited to searching and retrieving information stored in the form of alphanumeric symbols, but equally well may be applied to searching and retrieving information stored in the form of digitalized images and graphic symbols, as the word text used herein also may interpreted as images when these are represented wholly or partly as sets of symbols. It is also to be understood that the search system according to the invention can be implemented as software written in a suitable high-level language on commercially available computer systems, but it may also be implemented in the form of a dedicated processor device for searching and retrieving information of the aforementioned kind.

[0010] The above-mentioned objects and advantages are realized according to the invention with a search system which is characterized in that the data structure comprises a tree structure in the form of a non-evenly spaced sparse suffix tree ST(*T*) for storing suffixes of words and/or symbols *s* and sequences *S* thereof in the text *T*, that the metric *M* comprises a combination of an edit distance metric D(*s,q*) for an approximate degree of matching between words and/or symbols *s;q* in respectively the text *T* and a query *Q* and an edit distance metric $D_{ws}(S,P)$ for an approximate degree of matching between sequences *S* of words and/or symbols *s* in the text *T* and a query sequence *P* of words and/or symbols *q* in the query *Q*, the latter edit distance metric including weighting cost functions for edit operations which transform a sequence *S* of words and/or symbols *s* in the text *T* into the sequence *P* of words and/or symbols *q* in the query *Q*, the weighting taking place with a value proportional to a change in the length of the sequence *S* upon a transformation or dependent on the size of the words and/or symbols s;q in sequences *S;P* to be matched, that the implemented search algorithms comprise a first algorithm for determining the degree of matching between words and/or symbols *s;q* in the suffix tree representation of respectively the text *T* and a query *Q*, and a second algorithm for determining the degree of matching between sequences *S;P* of words and/or symbols *s;q* in the suffix tree representation of respectively the text *T* and the query *Q*, said first and/or second algorithms searching the data structure with queries *Q* in the form of either words, symbols, sequences of words or sequences of symbols or combinations thereof, such that information *R* is retrieved on the basis of query *Q* with a specified degree of matching between the former and the latter, and that the search algorithms optionally also comprise a third algorithm for determining exact matching between words and/or symbols *s;q* in the suffix tree representation of respectively the text *T and* the query *Q* and/or a fourth algorithm for determining exact matching between sequences *S;P* of words and/or symbols *s;q* in the suffix tree representation of respectively the text *T* and the query *Q*, said third and/or fourth algorithms searching the data structure with queries *Q* in the form of either words, symbols, sequences of words, or sequences of symbols or combinations thereof, such that information *R* is retrieved on the basis of the query *Q* with an exact matching between the former and the latter.

[0011] In an advantageous embodiment of the search system according to the invention the suffix tree ST(T) is a word-spaced sparse suffix tree $SST_{ws}(T)$, comprising only a subset of the suffixes in the text T.

[0012] Preferably is then the word-spaced sparse suffix tree $SST_{ws}(T)$ a keyword-spaced sparse suffix tree $SST_{kws}(T)$.

[0013] In further advantageous embodiments of the search system according to the invention the first algorithm for detecting the degree of keyword matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ is implemented as disclosed by dependent claim 4, the second algorithm for determining the degree of sequence matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ implemented as disclosed by dependent claim 5, whereby a subroutine of the second algorithm preferably is implemented as disclosed by dependent claim 6, the third algorithm for determining an exact keyword matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ implemented as disclosed by dependent claim 7, and finally the fourth algorithm for determining an exact keyword sequence matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ implemented as disclosed by dependent claim 8.

[0014] The above-mentioned objects and advantages are also realized according to the invention with a method which is characterized by generating the data structure as a word-spaced sparse suffix tree $SST_{ws}(T)$ of a text T for representing all the suffixes starting at a word separator symbol in the text T, storing sequence information of the words s in the text T in the word-spaced sparse suffix tree $SST_{ws}(T)$, generating a combined edit distance metric M comprising an edit distance metric $D(s,q)$ for words s in the text T and a query word q in a query Q and a word-size dependent edit distance metric $D_{ws}(S,P)$ for sequences S of words s in the text T and a sequence P of words q in the query Q, the edit distance metric $D_{ws}(S,P)$ being the minimum sum of costs for edit operations transforming a sequence S into the sequence P, the minimum sum of costs being the minimum sum of cost functions for each edit operation weighted by a value proportional to the change in the total length of the sequence S or by the ratio of the current word length and average word length in the sequences $S;P$, and determining the degree of matching between words $s;q$ by calculating the edit distance $D(s,q)$ between the words s of the retrieved information R and the word q of a query Q, or in case the words s,q are more than k errors from each other, determining the degree of matching between the word sequences S_R; P_Q of retrieved information R and a query Q respectively by calculating the edit distance $D_{ws}(S_R,P_Q)$ for all matches.

[0015] Advantageously the method according to the invention additionally comprises weighting an edit operation which changes a word s into word q with a parameter for the proximity between the characters of the words $s;q$, thus taking the similarity of the words $s;q$ in regard when determining the cost of the edit operation in question.

[0016] In an advantageous embodiment of the method according to the invention the number of matches is limited by calculating the edit distance $D_{WS}(S_R,P_Q)$ for restricted number of words in the query word sequence P_Q.

[0017] In another advantageous embodiment of the method according to the invention the edit distance $D(s, q)$ between word s and a word q is defined recursively and calculated by means of a dynamic programming procedure; and the edit distance $D_{ws}(S,P)$ between sequences S and a sequence P is correspondingly recursively defined and calculated by means of a dynamic programming procedure.

[0018] According to the invention the above-mentioned objects and advantages are also realized with the use of the search system according to the invention in an approximate search engine.

[0019] The search system and the method according to the invention shall now be discussed in greater detail in the following with reference to the accompanying drawing figures, of which

fig. 1 shows an example of a suffix tree,

fig. 2 examples of word-spaced sparse suffix trees as used with the present invention,

fig. 3 an example of a so-called PATRICIA trie as known in prior art,

fig. 4 a further example of a word-spaced sparse suffix tree as used with the present invention,

fig. 5 an example of explicitly stored word sequence information as used with the present invention,

fig. 6 a leaf node structure as used with the present invention, and

fig. 7 schematically the structure of a search engine with the search system according to the present invention.

[0020] The search system according to the invention consists essentially of three parts, namely the data structure, the metrics for approximate matching and the search algorithm. When full text retrieval is the target, as essentially will be the case with the search system according to the present invention, then the entire data set which shall be retrievable, will be stored in a data structure which supports a high query performance.

[0021] The basic concepts underlying the present invention shall first be discussed in some detail. Stored information in the form of text T is divided into words *s* and word sequences *S*. Words are substrings of the entire text separated by word boundary terms. The set of word boundary terms is denoted BT_{word}. A common set of word boundary terms could be the set {'', ':', '\t', '\n', '\0', '.', ',', '?'} where \t denotes a tab character, \n denotes a linefeed character and \0 denotes an end-of-document indicator. In connection with the following description of the present invention it will be useful with some definitions concerning strings and sequences.

Definition 1: String

[0022] A string is a sequence of symbols taken from an alphabet, such as the ASCII characters. Then the length of a string is the number of instances of symbols or characters comprising the string, and is denoted |x|. If *x* has the length *m* the string may also be written as $x_1x_2...x_i...x_m$, where x_i represents the *i*th symbol in the string.

[0023] A substring of *x* is a string given by a contiguous group of symbols within *x*. Thus, a substring may be obtained from *x* by deleting one or more characters from the beginning or the end of the string.

Definition 2: Substring, suffix and prefix

[0024] A substring of *x* is a string $x_i^j = x_ix_{i+1}...x_j$ for some $1 \le i \le j \le n$. The string $x_i = x_i^n = x_i..x_n$ is a suffix of string *x* and the string $x_j = x_1^j = x_1\, x_2...x_j$ is a prefix of string *x*.

[0025] Also the notion of a word sequence will be used.

Definition 3: Word sequence

[0026] A word sequence is a sequence of separated, consecutive words. A word sequence $S = s_1, s_2, ..., s_n$ consists of *n* single words (or strings) s_1, s_2, up to s_n.

[0027] Word sequences are delimited by sequence boundary terms. The set sequence boundary terms are denoted BT_{seq}. A common set of sequence boundary terms could be the set {'0\'}, where \0 indicates an end-of-document marker.

[0028] The concept approximate word matching can be described as follows.

[0029] Given a string $s = s_1s_2...s_n$ and a query term $q = q_1q_2...q_m$. Then the task is to find all occurrences of *q* in *s* that is at most *k* errors away from the original query term *q*. A proximity metric determines how to calculate the errors between *q* and a potential match $s_i...s_j$.

[0030] A common metric for approximate word matching is the Levenstein distance or edit distance (V.I. Levenstein, "Binary codes capable of correcting deletions, insertions, and reversals", (Russian) Doklady Akademii nauk SSSR, Vol. 163, No. 4, pp. 845-8 (1965); also Cybernetics and Control Theory, Vol. 10, No. 8, pp. 707-10, (1966)). This metric is defined as the minimum number of edit operations needed to transform one string into another. An edit operation is given by any rewrite rule, for instance:

- $(a \rightarrow \varepsilon)$, deletion
- $(\varepsilon \rightarrow a)$, insertion
- $(a \rightarrow b)$, change

[0031] Let *p* and *m* be two words of size *i* and *j*, respectively. Then $D(i,j)$ denotes the edit distance between the *i*th prefix of *p* and the *j*th prefix of *m*. The edit distance can then recursively be defined as:

$$D(i,0) = D(0,i) = i$$

$$D(i,j) = \min\begin{cases} D(i-1,j)+1 \\ D(i,j-1)+1 \\ D(i-1,j-1)+\partial(i,j) \end{cases} \qquad (1)$$

where

$$\partial(i,j) = 0 \text{ if } p_1 = m_j \text{ else } 1$$

[0032] It is also possible to define an approximate matching on the level of words in a word sequence and this can be described as follows.

[0033] Given a text T consisting of the n words $w_1, w_2 .. w_n$ where each of the words is a string of characters. A sequence pattern P consists of the m words $p_1, p_2, ..., p_m$. The sequence pattern P is said to have an approximate occurrence in T *if* the sequence $p_1, p_2, ..., p_m$ differs with at most k errors from a sequence $w_i, w_{i+1}, ..., w_j$ for some i, j, such that $1 \leq i \leq j \leq n$. Again, a proximity metric determines how to calculate the number of errors between the two sequences.

[0034] A text that shall be retrieved in a search system must be indexed in a manner which facilitates searching the data. Consequently the data structure is a kernel data structure of the search system according to the present invention and is based on so-called suffix trees and particularly a sparse suffix tree. These two kinds of structures shall be defined in the following. A suffix tree $S(T)$ is a tree representation of all possible suffixes in the text T. All unary nodes in a suffix tree $S(T)$ are concatenated with its child to create a compact variant.

[0035] Fig. 1 shows the suffix tree for the text T = "structure".

[0036] Even more particularly the present invention is based on sparse suffix trees. These were introduced by J. Kärkkäinen & E. Ukkonen, in "Sparse Suffix Trees", Proceedings of the Second Annual International Computing and Combinatorics Conference (COCOON '96), Springer Verlag, pp.219-230, which again was based on ideas published by D.R. Morrison, "PATRICIA-Practical Algorithm To Retrieve Information Coded in Alphanumeric", Journal of the ACM, 15, pp. 514-534 (1968). A sparse suffix tree is defined as follows.

Definition 4: Sparse suffix tree

[0037] A sparse suffix tree $SST(T)$ of the text T is a suffix tree, containing only a subset of the suffixes present in the suffix tree $ST(T)$ of the text.

[0038] When using the search system according to the present invention searching for entire words, advantageously a non-evenly spaced sparse suffix tree may be created by storing suffixes starting at word boundaries only. The concept words-spaced sparse suffix tree is defined as follows.

Definition 5: Word-spaced sparse suffix tree

[0039] A word-spaced sparse suffix tree $SST_{ws}(T)$ of a text T is a sparse suffix tree $SST(T)$ containing only the suffixes starting at a word separator character in the text.

[0040] Fig. 2 shows two examples of word-spaced sparse suffix trees. Parts of the suffixes have been omitted to enhance the readability. The word-spaced sparse suffix tree for T = "to be the best" is the left structure, and T = "to make the only major modification" is the right structure in fig. 2.

[0041] In the search system of the present invention the text is naturally divided into words which are stored independently in the word-spaced sparse suffix tree.

[0042] As the atomic search term for searching is the word itself, advantageously each suffix will be terminated at the end of the word. This reduces the sparse suffix tree to a so-called PATRICIA trie (Morrison, op.cit.). A trie as defined in the literature is a rooted tree with the properties that each node, except the root, contains a symbol of the alphabet and that no two children of the same node contain the same symbol. It should be noted that the word trie derives from the word "retrieval" and hence indicates that the trie is a tree structure suitable for retrieval of data. A PATRICIA trie is defined as a keyword-spaced sparse suffix tree (KWS tree) where the suffixes stored in the leaf nodes are limited by keyword delimiters. An example of a PATRICIA trie for the set of keywords {"avoid", "abuse", "be", "become", "breathe", "say"} is shown in fig. 3. The structure used in the search system of the present invention differs from the PATRICIA trie because the search system explicitly stores sequence information of the words. Reducing the suffix length requires that the representation of the leaf node is changed. Pointers to the original text are replaced by the suffix string itself. A suffix length reduction of this kind is shown in fig. 4 for one of the strings shown in fig. 2. In other words fig. 4 shows the word-spaced sparse suffix tree for T = "to make the only major modification" and with suffixes cut off at word boundaries. A leaf node will contain a list of all positions where the word represented by the leaf node occurs.

[0043] Instead of using the implicit sequence of information found in the original text, the present invention explicitly stores sequence information in the word-spaced sparse suffix tree. This is done by using pointers between the leaf nodes that represent consecutive words in the original text. As at least all the occurrences of the word represented by a particular leaf node are available, a pointer must be added to the next consecutive leaf.

[0044] A leaf node contains only the suffix of the word it represents, so when traversing the sequence pointers in the occurrence list only the suffixes of each of the consecutive words are revealed. This is handled by storing the entire word in the leaf node instead of just the suffix and thus also data structure of the invention differs from the PATRICIA trie in this respect. The data structure for explicitly stored word sequence information with an occurrence list with pointers to the next consecutive word and to its occurrence is shown in fig. 5.

[0045] The search system according to the present invention uses a PATRICIA trie for organizing the occurrence list (Morrison, op.cit.). The PATRICIA trie enables the search system to access the list of all consecutive words matching the string p_2 in a time $O(|p_2|)$, where $|p_2|$ of course is the length of p_2. By using a PATRICIA trie to organize the list of occurrences, a completely defined tree structure is obtained for storing words from a text and maintaining the sequence information. A typical leaf node, with both a PATRICIA trie for the organized occurrence list and the extra unsorted list of occurrences, is shown in fig. 6. As an example the memory requirement for an occurrence list as used in the search system of the present invention, a database with about 742358 documents has a total of 333 856 744 words and a lexicon of 538 244 distinct words. The total size of the database is 2054.52 MB. The average word length is thus 6.45 bytes. A sparse suffix tree will use 8 bytes for each internal node, using 32 bit pointers. It is assumed that an average of 3 internal nodes is used for each word. The leaf node would then require 6.45 bytes for storing the entire word plus 32 bits for a pointer to an occurrence list. A total of 34.45 bytes/word gives a total size of 18.108 MB. In addition the occurrence list has the size of 4 bytes per entry and 12 bytes if the full version is to be used. Hence the total memory requirement of the occurrence list varies from 1273 MB to 3820 MB. The data structure using a sparse suffix tree will have a size between 60% to 200% of the original text. This is comparable with the requirements of an inverted file, but the sparse suffix tree as used in the search system according to the invention provides much faster searching, enables approximate matching and makes sequence matching easy to perform.

[0046] In approximate searching, a metric is used to give an error measure of a possible match. The search system according to the present invention employs several metrics, and particularly a unique combination of metrics. These metrics along with the combined metric shall be discussed in the following.

[0047] An edit distance metric as defined above allows the operations deletion, insertion and change which intuitively apply to words as well as characters. Common errors in matching phrases are missing, extra or changed words. Hence the edit distance metric as previously defined shall be adapted and extended in order to apply to the approximate word sequence matching problem. Edit operations for sequences are defined below.

Definition 6: Edit operations for sequences

[0048] For transforming one sequence S of words into another sequence P of words, the edit operations allowed on the word in the sequences may be written according to the following rewrite rules:

- $(a \rightarrow \varepsilon)$, deletion of word a from the sequence
- $(\varepsilon \rightarrow a)$, insertion of word a into the sequence
- $(a \rightarrow b)$, change of word a into word b
- $(ab \rightarrow ba)$, transposition of adjacent words a and b.

[0049] Instead of characters as atoms, the search system according to the invention applies the edit operations to words which then should be regarded as the operational atoms.

[0050] A cost function $c_{edit}(x \rightarrow y)$ is a constant which is defined as

$$c_{edit}(x \rightarrow y) = \begin{cases} 1 & \text{delete} \\ 1 & \text{insert} \\ 1 & \text{transpose} \\ \partial(x,y) & \text{change} \end{cases} \qquad (2)$$

where $\partial(x,y)$ is defined as

$$\partial(x,y) = \begin{cases} 0 & x = y \\ 1 & \text{else} \end{cases} \qquad (3)$$

[0051] By using the edit operations as defined above the edit distance for sequences can now be defined.

Definition 7: Edit distance for sequences

[0052] The edit distance metric for sequences defines the distance $D_{seq}(S,P)$ between the sequence $S = s_1,s_2,...,s_n$ and the sequence $P = p_1,p_2, ...,p_m$ as the minimum sum of cost $c(x \rightarrow y)$ for the sequence of edit operations transforming the sequence S into the sequence P.

[0053] The search system according to the present invention enhances the edit distances metric for sequences to weight the cost of the edit operations by the size of the words operated upon.

Definition 8: Word size-dependent edit distance for sequences

[0054] The word size dependent edit distance for sequences is defined as the minimum sum of costs for the editing operations needed to transform one sequence into the other. The cost functions are dependent on the word size of their operands.

[0055] In the search system according to the invention a definition of cost functions is given by the equations

$$\begin{aligned} c_{insert}(\varepsilon \rightarrow a) &= \frac{|a|}{l} \\ c_{delete}(a \rightarrow \varepsilon) &= \frac{|a|}{l} \\ c_{transpose}(ab \rightarrow ba) &= 1 \\ c_{change}(a \rightarrow b) &= \frac{\max(||a|-|b||, l)}{l} \end{aligned} \qquad (4)$$

where l denotes the average length of a word in the two sequences being compared. The cost of each edit operation is weighted by a value proportional to the change in the total length of the sequence or by the ratio of the current word length and the average word length in the sequences considered.

[0056] Now the distance metric reflects the assumption of some relation between the word length and how important the word is to the semantic context of the word sequence. Furthermore the search system according to the invention employs proximity at the character level when the change edit operation $(a \rightarrow b)$ is used. Replacing a word a by another word b should be related to the similarity between these two words. The new cost function for the change edit operation hence is given as:

$$c_{change}(a \rightarrow b) = \partial_{approx}(a,b)\frac{\max(||a|-|b||, \overline{l}}{\overline{l}} \qquad (5)$$

where

$$\partial_{approx}(a,b) = D(a,b) \qquad (6)$$

[0057] Where $D(a,b)$ is the normalized edit distance measuring function for words, 0 means full similarity, 1 means no similarity.

[0058] The search system according to the invention combines the edit distance metric for sequences with the cost functions as given by formulas (4), (5) and (6), with an edit distance metric for words as given by formula (1). This means that sequence edit operations are only used when the words being matched are more than k errors away from each other.

[0059] The algorithms used in the search system according to the invention perform efficient searching of the described structures. Matches are found according to the metrics as given above.

[0060] Approximate word matching in a word-spaced sparse suffix tree is done by combining the calculation of the edit distance matrix and a traversal of the suffix tree. An algorithm for this is written in pseudo-code and given in table I.

[0061] This algorithm is adapted from a trie-matching algorithm as proposed by H. Shang & T.H. Merrettal, "Tries for

Approximate String Matching", IEEE Transactions on Knowledge and Data Engineering, Vol.5, No. 4, pp. 540-547 (1996). The expected worst case running time of the algorithm is $O(k|\Sigma|^k)$ according to Shang & Merrettal (op.cit.).

[0062] Approximate word sequence matching requires the calculation of the word sequence edit distance for all possible matches. However, the number of possible matches can be limited by starting the calculation of the edit distance only on the possible words. The cost of deleting a word from the sequences determines the number of possible start words. If the accumulated cost of deleting the i first words in a query sequence P_Q rises above a given error threshold, the candidate sequence starting with the ith word of the query cannot possibly be a match. Therefore for a query sequence P_Q of i words, at most i possible start words will be tried. Since there are no backpointers in the sequence structure of the tree, it will not be ensured that all possible matches are obtained. Adding backpointers would solve this problem. The algorithm for approximate word sequence matching as used in the search system according to the present invention, is given in pseudo-code in table II below. This algorithm tries to match the first keyword with $p_1, p_2 ..$ sequentially, testing all possible start positions.

[0063] In the ApproxSequenceMatch algorithm in table II the *ApproxMatchRest* function is defined by the algorithm in table III below. This function matches the remaining sequence, using an initial error value.

Table I

```
FindApproximate (root,p,k)
  node ← root;
  i ← 1;

  nodes ← Children (node);   // A stack of    nodes
  for all v ∈ nodes do

   if IsLeaf (v) then
    for j ← i to length (Suffix(node)) do
     w_j ← Suffix(node)_{j-i};
     if w_j = '$' then      // '$' is a stopchar
      output w_{1..j};
      return;
     if EditDist (i) = ∞ then
      break;
   else              //Internal node
     i ← i+1;
     w_i ← label (v)
     if EditDist (i) = ∞ then
      break;
     nodes ← Children (v) □ nodes;
  // end for

EditDistance(j)     // Calculates jth row
   for i← 1 to length (P) do
     if p_i = w_j then ∂ ← 0 else ∂ ← 1;
    c_1 = D[i-1,j] + c_ins(m_j);
    c_2 = D[i,j-1] + c_del(p_i);
    c_3 = D[i-1,j-1] + c_change(p_i,m_j);
    c_4 = D[i-2,j-2] + c_transpose(p_i,m_{j-1});
    D[i,j] ← c_fraction(j/l)· min(c_1, c_2, c_3, c_4);
   if D[i,j] > k
    return ∞; // No distance below k

   return D[i,j]
```

Table II

```
ApproxSequenceMatch_ED (root, P(=p1,p2,...,Pm),k)
  m← |p|
  matches ← ∅
  startError ← 0
  startIndex ← 1

  while startError ≤ k OR startIndex ≤ m do
   startNode ← FindExact (p_startIndex);
   list ← UnorderedOccurrenceList (startNode);
   for all v ∈ lists do
    if ApproxMatchRest  (v,P,k,startError) then
     matches ← □ v;

  startError ← startError + c_del (p_startIndex);
  startIndex ← startIndex + 1;
```

Table III

```
ApproxMatchRest (u,P,K,startError)
  error ← startError;
  lastError ← startError;
  column ← 0 ;
  node ← u;
 for v ← p2 to p|p| do
  node ← NextOccurrence(node);
  word ← Keyword(node);
  lastError ← error;
  error ← startError + EditDistance (column);
  if error > k AND lastError > k then
   return false;

 return true;

EditDistance(j)     // Calculates jth row
   for i← 1 to length (P) do
     if pi = wj then ∂ ← 0 else ∂ ← 1;
    c1 = D[i-1,j] + cins(mj);
    c2 = D[i,j-1] + cdel(pi);
    c3 = D[i-1,j-1] + cchange(pi,mj);
    c4 = D[i-2,j-2] + ctranspose(pi,mj-1);
    D[i,j] ← cfraction(j/l)· min(c1,c2,c3,c4);
   if D[i,j] > k
    return ∞; // No distance below k

   return D[i,j]
```

[0064] The algorithms in tables II and III are written in the same pseudo-code as the algorithm in table I.

[0065] The *FindExact* function used to find the leaf node matching the first word in the sequence performs a simple traversal of the tree and its running time is $O|p_1|$ where p_1 denotes the first word in a query sequence P_Q. Calculating the edit distance can be done in $|P|^2$ time using straightforward dynamic programming or in $O(k)$ time (where k denotes the error threshold) using improved versions of the calculation algorithm, see E. Ukkonen, "Finding Approximate Patterns in Strings", Journal of Algorithms, vol. 6, pp. 132-137 (1985).

[0066] If $\Sigma n_{occ}(p_i)$ denotes the total sum of the number of occurrences of each word p_i in the word sequence, then the worst case running time is $O(k\Sigma n_{occ}(p_i))$.

[0067] Finally the implementation of a search engine based on the search system according to the invention shall

briefly be discussed. Particularly a search engine based on the search system according to the invention is implemented as an approximate search engine (ASE) and is intended as a search engine for indexing large document collections and providing algorithms for exact and approximate searching of these document collections. ASE shall provide a data structure for storing large texts or collection of documents. It is to be understood that the data structure may be generated from documents which contain additional information, such as images, video, sound, and the text may be formatted and/or annotated. The data structure is identical to the word-spaced sparse suffix tree as discussed above and it is, of course, to be understood that the words is the keywords of the search system, hence the word-spaced sparse suffix tree may instead be termed a keyword-spaced sparse suffix tree (KWS tree). The ASE shall contain algorithms for indexing documents in the KWS tree. These algorithms, of course, do not form a part of the search system according to the present invention, but they are well-known to persons skilled in the art and described in the literature, see for instance J. Kärkkäinen & E. Ukkonen (op.cit.) and D.R. Morrison (op.cit).

[0068] The search system according to the invention and as used in the ASE employs algorithms both for exact and approximate matching of a pattern in a KWS tree. The algorithms given above in table I and table II are used for approximate word and word sequence matching with the non-uniform edit distance as a metric. Finding an exact match of keyword p with length m in a KWS tree is known in the art and easily implemented as a simple traversal of the tree structure. An appropriate algorithm for exact keyword matching written in pseudo-code is given in table IV. The search system according to the invention also shall be able to support algorithms for exact keyword sequence matching. Algorithms for exact keyword sequence matching are known in the art and easily implemented as e.g. shown in pseudo-code in table V below. The algorithm given here will find the exact match of the first keyword, if any. Then it will for all occurrences of the first keyword check if the second keyword matches the second keyword of the query. If so, the *MatchRest* procedure in table V is used to determine if the occurrence of the two first keywords are matching in the entire sequence. For approximate keyword matching in a KWS tree the search system implements the algorithm in table I above. For approximate keyword sequence matching the search system implements the algorithm in table II above, matching a first keyword sequentially with $p_1, p_2, \ldots$ and testing all possible start positions, applying the *ApproxMatchRest* function as given in table III to match a sequence starting at a particular position and handle the initial error value.

[0069] Finally, the ASE shall need a simple front end which gives the user control of indexing and querying the document collection. The front end should also be able to furnish statistics of the document collection and provide both a network interface for remote access, e.g. via WWW, and a local server user interface.

[0070] The ASE with the search system according to the invention should be general in a manner that allows for the adding new indexing and searching algorithms easily. Also, storing extra information about each document or keyword shall be possible to implement in an easy manner. Particularly the front end should be independent of the data structure and the search algorithms, such that internal changes in these has no effect on the design of the former.

[0071] The use of the search system according to the invention the ASE should be designed to have as low memory overhead as possible in the data structure. Also, searching should be designed to be as fast as possible. However, there will usually be a trade-off between these two factors.

Table IV

```
FindExact (root,p)
 i← 1;
  node ← Find Child (root,p_i);
  while node AND i ≤ length (p) do
   if IsLeaf(node) AND Suffix(node) = P_1...P_m then
    return node;
   i ← i+1
   node ← FindChild (node, p_i);

  return NIL;
```

Table V

```
MatchSequenceExact (P, root)
 matches ← ∅;
 v ← FindExact (p_1,root);
 if |P| > 1 then
  if v ≠ NIL then
   list ← UnorderedOccurrenceList (v);
   for all u ∈ list do
    if NextKeyword(u) = p_2 then
     if MatchRest (p_3...p_m,u) then
      matches ← matches □ Occurrence(u);

 return matches;

MatchRest (P,u)
 node ← u;
 for v ← p_2 to p_|P| do
  node ← NextOccurrence(node);
  word ← Keyword(node);
  if v ≠ word then
   return false;
```

[0072] To sum up, an ASE with a search system according to the invention shall comprise four major modules.

1. Document indexing module DIM for indexing documents in the KWS tree structure. This module should also

contain all extensions to support several document types.

2. Data storage module DSM based on a keyword-spaced sparse suffix tree (KWS tree).

3. Search algorithm module SAM for searching the KWS tree, comprising algorithms for exact and/or approximate matching of respectively words and word sequences.

4. User interface front-end module FEM comprising both a local server user interface and a network interface for remote queries.

[0073] The four modules of the ASE works together to offer a complete search engine functionality. The data flow between the different modules is shown in fig. 7. Indexing a collection of documents in done in the document indexing module DIM comprising indexing algorithms. This module is, of course, not a part of the search system according to the invention, but indexing algorithms that can be used are well-known in the art. The text found in the documents is passed on to the data storage DSM module for storage. The data storage module is, of course, a part of the search system according to the invention and is as stated based on the KWS tree structure. The search algorithm module SAM contains algorithms for searching the data located in the data storage module. This module implements the search system according to the present invention and allows for a search process querying the data structure for tree and node information, while maintaining state variables. The front-end module may for instance be implemented on a work station or a personal computer and the like, providing the functionality as stated above.

[0074] As already stated in the introduction, it is to be understood that the search system according to the invention can be implemented as software written in a suitable high-level language on commercially available computer systems, including workstations. It may also as stated be implemented in the form of a dedicated processor device which advantageously may comprise a large number of parallel processors being able to process large word sequences in parallel for approximate matching with a large number of query word sequences. The fixed operational parameters of the processor may then be entered in a low-level code, while keyword sequences input from the KWS tree structure allows for an extremely fast processing of queries on a huge amount of data, and the search system according to present invention shall hence in high degree be suited for performing searches on e.g. the World Wide Web, even in a KWS tree structure large enough to index all documents presently offered on the World Wide Web and moreover capable of handling the expected data volume growth on the World Wide Web in the future.

Claims

1. A search system for information retrieval, particularly information stored in form of text, wherein a text *T* comprises words and/or symbols *s* and sequences *S* thereof, wherein the information retrieval takes place with a given or varying degree of matching between a query *Q*, wherein the query *Q* comprises words and/or symbols *q* and sequences *P* thereof, and retrieved information *R* comprising words and/or symbols and sequences thereof from the text *T*, wherein the search system comprises a data structure for storing at least a part of the *text T*, and a metric *M* which measures the degree of matching between the query *Q* and retrieved information *R*, and wherein the search system implements search algorithms for executing a search, particularly a full text search on the basis of keywords *kw*, **characterized in that** the data structure comprises a tree structure in the form of a non-evenly spaced sparse suffix tree ST(*T*) for storing suffixes of words and/or symbols *s* and sequences *S* thereof in the text *T*, that the metric *M* comprises a combination of an edit distance metric D(*s,q*) for an approximate degree of matching between words and/or symbols *s;q* in respectively the text *T* and a query *Q* and an edit distance metric $D_{ws}(S, P)$ for an approximate degree of matching between sequences *S* of words and/or symbols *s* in the text *T* and a query sequence *P* of words and/or symbols *q* in the query *Q*, the latter edit distance metric including weighting cost functions for edit operations which transform sequences of words and/or symbols *s* in the text *T* into the sequence *P* of words and/or symbols *q* in the query *Q*, the weighting taking place with a value proportional to a change in the length of the sequence *S* upon a transformation or dependent on the size of the words and/or symbols *s;q* in sequences *S;P* to be matched, that the implemented search algorithms comprise a first algorithm for determining the degree of matching between words and/or symbols *s;q* in the suffix tree representation of respectively the text *T* and a query *Q*, and a second algorithm for determining the degree of matching between sequences *S;P* of words and/or symbols *s;q* in the suffix tree representation of respectively the text *T* and the query *Q*, said first and/or second algorithms searching the data structure with queries *Q* in the form of either words, symbols, sequences of words or sequences of symbols or combinations thereof, such that information *R* is retrieved on the basis of query *Q* with a specified degree of matching between the former and the latter, and that the search algorithms optionally also comprise a third algorithm for determining exact matching between words and/or symbols

s;q in the suffix tree representation of respectively the text *T* and the query *Q* and/or a fourth algorithm for determining exact matching between sequences *S;P* of words and/or symbols *s;q* in the suffix tree representation of respectively the text *T* and the query *Q*, said third and/or fourth algorithms searching the data structure with queries *Q* in the form of either words, symbols, sequences of words, or sequences of symbols or combinations thereof, such that information *R* is retrieved on the basis of the query *Q* with an exact matching between the former and the latter.

2. A search system according to claim 1, **characterized in that** the non-evenly spaced sparse suffix tree ST(*T*) is a word-spaced sparse suffix tree $SST_{ws}(T)$ comprising only a subset of the suffixes in the text *T*.

3. A search system according to claim 2, **characterized in that** the word-spaced sparse suffix tree $SST_{ws}(T)$ is a keyword-spaced sparse suffix tree $SST_{kws}(T)$.

4. A search system according to claim 3, **characterized in that** the first algorithm for detecting the degree of keyword matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ is implemented in pseudo-code as follows:

```
FindApproximate (root,p,k)
  node ← root;
  i ← 1;

  nodes ← Children (node);   // A stack of    nodes
  for all v ∈ nodes do

   if IsLeaf (v) then
    for j ← i to length (Suffix(node)) do
     wj ← Suffix(node)j-i;
     if wj = '$' then      // '$' is a stopchar
      output w1..j;
      return;
     if EditDist (i) = ∞ then
      break;
   else              //Internal node
     i ← i+1;
     wi ← label (v)
     if EditDist (i) = ∞ then
```

```
            break;
          nodes ← Children (v) □ nodes;
        // end for

    EditDistance(j)     // Calculates jth row
      for i← 1 to length (P) do
        if p_i = w_j then ∂ ← 0 else ∂ ← 1;
       c_1 = D[i-1,j] + c_ins(m_j);
       c_2 = D[i,j-1] + c_del(p_i);
       c_3 = D[i-1,j-1] + c_change(p_i,m_j);
       c_4 = D[i-2,j-2] + c_transpose(p_i,m_j-1);
       D[i,j] ← c_fraction(j/l)· min(c_1,c_2,c_3,c_4);
      if D[i,j] > k
       return ∞; // No distance below k

      return D[i,j]
```

5. A search system according to claim 3, **characterized in that** the second algorithm for determining the degree of keyword sequence matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ is implemented in pseudo-code as follows:

```
ApproxSequenceMatch_ED (root, P(=p_1,p_2,...,P_m),k)
  m← |p|
  matches ← Ø
  startError ← 0
  startIndex ← 1

  while startError ≤ k OR startIndex ≤ m do
   startNode ← FindExact (p_startIndex);
   list ← UnorderedOccurrenceList (startNode);
   for all v ∈ lists do
    if ApproxMatchRest  (v,P,k,startError) then
     matches ← □ v;

  startError ← startError + c_del (p_startIndex);
  startIndex ← startIndex + 1;
```

6. A search system according to claim 5, **characterized in that** the *ApproxMatchRest* subroutine of the second algorithm is implemented in pseudo-code as follows:

```
ApproxMatchRest (u,P,K,startError)
  error ← startError;
  lastError ← startError;
  column ← 0 ;
  node ← u;
 for v ← p2 to p|p| do
  node ← NextOccurrence(node);
  word ← Keyword(node);
  lastError ← error;
  error ← startError + EditDistance (column);
  if error > k AND lastError > k then
   return false;

 return true;

EditDistance(j)     // Calculates jth row
   for i← 1 to length (P) do
     if pi = wj then ∂ ← 0 else ∂ ← 1;
    c1 = D[i-1,j] + cins(mj);
    c2 = D[i,j-1] + cdel(pi);
    c3 = D[i-1,j-1] + cchange(pi,mj);
    c4 = D[i-2,j-2] + ctranspose(pi,mj-1);
    D[i,j] ← cfraction(j/l)· min(c1,c2,c3,c4);
   if D[i,j] > k
    return ∞; // No distance below k

   return D[i,j]
```

7. A search system according to claim 3, **characterized in that** the third algorithm for determining an exact keyword matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ is implemented in pseudo-code as follows:

```
FindExact (root,p)
 i← 1;
  node ← Find Child (root,p_i);
  while node AND i ≤ length (p) do
   if IsLeaf(node) AND Suffix(node) = P_i…P_m then
    return node;
  i ← i+1
  node ← FindChild (node, p_i);

 RETURN NIL;
```

8. A search system according to one of claims 3 to 7, **characterized in that** the fourth algorithm for determining an exact keyword sequence matching in a keyword-spaced sparse suffix tree $SST_{kws}(T)$ is implemented in pseudo-code as follows:

```
MatchSequenceExact (P, root)
 matches ← Ø;
 v ← FindExact (p_1,root);
 if |P| > 1 then
  if v ≠ NIL then
   list ← UnorderedOccurrenceList (v);
   for all u ∈ list do
    if NextKeyword(u) = p_2 then
     if MatchRest (p_3…p_m,u) then
      matches ← matches □ Occurrence(u);

 return matches;

MatchRest (P,u)
 node ← u;
 for v ← p_1 to p_|P| do
  node ← NextOccurrence(node);
  word ← Keyword(node);
  if v ≠ word then
   return false;
```

9. A method in a search system for information retrieval, particularly information stored in form of text, wherein a text *T* comprises words and/or symbols *s* and sequences *S* thereof, wherein the information retrieval takes place with a given or varying degree of matching between a query *Q*, wherein the query *Q* comprises words and/or symbols *q* and sequences *P* thereof, and retrieved information *R* comprising words and/or symbols and sequences thereof from the text *T*, wherein the search system comprises a data structure for storing at least a part of the text *T*, and

a metric *M* which measures the degree of matching between the query *Q* and retrieved information *R*, and wherein the search system implements search algorithms for executing a search, particularly a full text search on the basis of keywords *kw*, wherein the information in the text *T* is divided into words *s* and word sequences *S*, the words being substrings of the entire text separated by word boundary terms and forming a sequence of symbols, and wherein each word is structured as a sequence of symbols, **characterized by** generating the data structure as a word-spaced sparse suffix tree $SST_{ws}(T)$ of a text *T* for representing all the suffixes starting at a word separator symbol in the text *T*, storing sequence information of the words *s* in the text *T in* the word-spaced sparse suffix tree $SST_{ws}(T)$, generating a combined edit distance metric *M* comprising an edit distance metric *D(s,q)* for words *s* in the text *T* and a query word *q* in a query *Q* and a word-size dependent edit distance metric $D_{ws}(S,P)$ for sequences *S* of words *s* in the text *T and* a sequence *P* of words *q* in the query *Q*, the edit distance metric $D_{ws}(S, P)$ being the minimum sum of costs for edit operations transforming a sequence *S* into the sequence *P*, the minimum sum of costs being the minimum sum of cost functions for each edit operation weighted by a value proportional to the change in the total length of the sequence *S* or by the ratio of the current word length and average word length in the sequences *S;P*, and determining the degree of matching between words *s,q* by calculating the edit distance $D(s,q)$ between the words *s* of the retrieved information R and the word *q* of a query *Q*, or in case the words *s,q* are more than *k* errors from each other, determining the degree of matching between the word sequences S_R; P_Q of retrieved information *R* and a query *Q* respectively by calculating the edit distance $D_{ws}(S_R,P_Q)$ for all matches.

10. The method according to claim 9, **characterized by** additionally weighting an edit operation which changes a word *s* into word *q* with a parameter for the proximity between the characters of the words *s;q*, thus taking the similarity of the words *s;q* in regard when determining the cost of the edit operation in question.

11. The method according to claim 9, **characterized by** limiting the number of matches by calculating the edit distance $D_{ws}(S_R,P_Q)$ for restricted number of words in the query word sequence P_Q.

12. A method according to claim 9, **characterized by** defining the edit distance $D(s,q)$ between words *s* and a word *q* recursively and calculating the edit distance $D(s,q)$ by means of a dynamic programming procedure.

13. A method according to claim 9, **characterized by** defining the edit distance $D_{ws}(S,P)$ between sequences *S* and a sequence *P* recursively and calculating the edit distance $D_{ws}(S,P)$ by means of a dynamic programming procedure.

14. The use of a search system according to claim 1 in an approximate search engine.

Patentansprüche

1. Suchsystem für die Informationswiedergewinnung, insbesondere für Information, die in Form von Text gespeichert ist, wobei ein Text *T* Wörter und/oder Symbole *s* und Sequenzen S hieraus umfaßt, wobei die Informationswiedergewinnung mit einem gegebenen oder variierenden Grad von Übereinstimmung zwischen einer Anfrage *Q* und wiedergewonnener Information *R* abläuft, wobei die Anfrage *Q* Wörter und/oder Symbole *q* und Sequenzen *P* hieraus umfaßt und wobei die Information *R* Wörter und/oder Symbole und Sequenzen hieraus aus dem Text *T* umfaßt, wobei das Suchsystem eine Datenstruktur zum Speichern wenigstens eines Teils des Textes *T* und ein Maß *M* umfaßt, welches den Grad der Übereinstimmung zwischen der Anfrage *Q* und der wiedergewonnenen Information *R* mißt, und wobei das Suchsystem einen Suchalgorithmus zur Ausführung einer Suche realisiert, insbesondere einer Volltextsuche auf der Basis von Schlüsselwörtern *kw*, **dadurch gekennzeichnet, daß** die Datenstruktur eine Baumstruktur in der Form eines nicht gleichmäßigen verteilten dünnbesiedelten Suffix-Baums $ST(T)$ zum Speichern von Suffixen von Wörtern und/oder Symbolen *s* und Sequenzen S daraus in dem Text *T* umfaßt, daß das Maß *M* eine Kombination aus einem Edit-Abstandsmaß $D(s,q)$ für einen ungefähren Grad an Übereinstimmung zwischen Wörtern und/oder Symbolen *s;q* in dem Text *T* bzw. einer Anfrage *Q* und ein Edit-Abstandsmaß $D_{ws}(S,P)$ für einen ungefähren Grad an Übereinstimmung zwischen Sequenzen *S* von Wörtern und/oder Symbolen *s* in dem Text *T* und einer Anfragesequenz *P* von Wörtern und/oder Symbolen *q* in der Anfrage *Q* umfaßt, wobei das zuletzt genannte Edit-Abstandsmaß eine Gewicht-Kostenfunktion für Editieroperationen umfaßt, welche Sequenzen von Wörtern und/oder Symbolen *s* in dem Text *T* in die Sequenz *P* von Wörtern und/oder Symbolen *q* in der Anfrage *Q* transformiert, wobei die Gewichtung mit einem Wert erfolgt, der proportional zu einer Änderung der Länge der Sequenz *S* bei einer Transformation oder abhängig von der Größe der Wörter und/oder Symbole *s;q* in den abzugleichenden Sequenzen *S;P* ist, daß der implementierte Suchalgorithmus einen ersten Algorithmus zum Ermitteln des Grades der Übereinstimmung zwischen Wörtern und/oder Symbolen *s;q* in der

Suffix-Baumdarstellung des Textes *T* bzw. einer Anfrage *Q* und einen zweiten Algorithmus zum Ermitteln des Grades der Übereinstimmung zwischen Sequenzen *S;P* von Wörtern und/oder Symbolen *s;q* in der Suffix-Baumdarstellung des Textes *T* bzw. der Anfrage *Q* umfaßt, wobei der erste und/oder zweite Algorithmus die Datenstrukturen mit Anfragen *Q* in der Form von entweder Wörtern, Symbolen, Wortsequenzen oder Symbolsequenzen oder Kombinationen daraus absucht, so daß Information *R* auf der Basis der Anfrage *Q* mit einem bestimmten Grad an Übereinstimmung zwischen ersterer und letzterer wiedergewonnen wird, und daß der Suchalgorithmus optional auch einen dritten Algorithmus zum Ermitteln der exakten Übereinstimmung zwischen Wörtern und/oder Symbolen *s;q* in der Suffix-Baumdarstellung des Textes *T* bzw. der Anfrage *Q* und/oder einen vierten Algorithmus zum Ermitteln der exakten Übereinstimmung zwischen Sequenzen *S;P* von Wörtern und/oder Symbolen *s;q* in der Suffix-Baumdarstellung des Textes *T* bzw. der Anfrage *Q* umfaßt, wobei der dritte und/oder vierte Algorithmus die Datenstruktur mit Anfragen *Q* in der Form von entweder Wörtern, Symbolen, Wortsequenzen oder Symbolsequenzen oder Kombinationen daraus absuchen, so daß Information R auf der Basis der Anfrage Q mit einer exakten Übereinstimmung zwischen ersterer und letzterer wiedergewonnen wird.

2. Suchsystem nach Anspruch 1, **dadurch gekennzeichnet, daß** der nicht gleichmäßig verteilte dünnbesiedelte Suffix-Baum ST (*T*) ein dünnbesiedelter Wortabstand-Suffix-Baum $SST_{ws}(T)$ ist, der nur eine Untermenge der Suffixe in dem Text *T* umfaßt.

3. Suchsystem nach Anspruch 2, **dadurch gekennzeichnet, daß** der dünnbesiedelte Wortabstand-Suffix-Baum $SST_{ws}(T)$ ein dünnbesidelter Stichwort-Abstand-Suffix-Baum $SST_{kws}(T)$ ist.

4. Suchsystem nach Anspruch 3, **dadurch gekennzeichnet, daß** der erste Algorithmus zum Erfassen des Grads der Stichwortübereinstimmung in einem dünnbesiedelten Stichwort-Abstand-Suffix-Baum $SST_{kws}(T)$ in Pseudo-Code wie folgt implementiert wird:

```
FindApproximate (root,p,k)
  node ← root;
  i ← 1;

  nodes ← Children (node);   // A stack of    nodes
  for all v ∈ nodes do

   if IsLeaf (v) then
    for j ← i to length (Suffix(node)) do
       wj ← Suffix(node)j-i;
```

```
            if w_j = '$' then      // '$' is a stopchar
             output w_1..j;
             return;
            if EditDist (i) = ∞ then
             break;
          else              //Internal node
            i ← i+1;
            w_i ← label (v)
            if EditDist (i) = ∞ then
            break;
            nodes ← Children (v) □ nodes;
         // end for

       EditDistance(j)     // Calculates jth row
          for i← 1 to length (P) do
            if p_i = w_j then ∂ ← 0 else ∂ ← 1;
           c_1 = D[i-1, j] + c_ins(m_j);
           c_2 = D[i, j-1] + c_del(p_i);
           c_3 = D[i-1, j-1] + c_change(p_i, m_j);
           c_4 = D[i-2, j-2] + c_transpose(p_i, m_j-1);
           D[i, j] ← c_fraction(j/l) · min(c_1, c_2, c_3, c_4);
          if D[i, j] > k
           return ∞;  // No distance below k

        return D[i, j]
```

5. Suchsystem nach Anspruch 3, **dadurch gekennzeichnet, daß** der zweite Algorithmus zum Bestimmen des Grads der Übereinstimmung einer Stichwortsequenz in einem dünnbesiedelten Stichwort-Abstand-Suffx-Baum SST_{kws} (*T*) in Pseudo-Code wie folgt implementiert wird:

```
ApproxSequenceMatch_ED (root, P(=p_1,p_2,...,P_m),k)
  m← |p|
  matches ← Ø
  startError ← 0
  startIndex ← 1

  while startError ≤ k OR startIndex ≤ m do
   startNode ← FindExact (p_startIndex);
   list ← UnorderedOccurrenceList (startNode);
   for all v ∈ lists do
    if ApproxMatchRest  (v,P,k,startError) then
      matches ← □ v;

  startError ← startError + c_del (p_startIndex);
```

```
startIndex ← startIndex + 1;
```

6. Suchsystem nach Anspruch 7, **dadurch gekennzeichnet, daß** die Unterroutine *ApproxMatchRest* des zweiten Algorithmus in Pseudo-Code wie folgt implementiert wird:

```
ApproxMatchRest (u,P,K,startError)
  error ← startError;
  lastError ← startError;
  column ← 0 ;
  node ← u;
for v ← p2 to p|p| do
 node ← NextOccurrence(node);
 word ← Keyword(node);
 lastError ← error;
 error ← startError + EditDistance (column);
 if error > k AND lastError > k then
  return false;

return true;

EditDistance(j)    // Calculates jth row
  for i← 1 to length (P) do
    if pi = wj then ∂ ← 0 else ∂ ← 1;
   c1 = D[i-1,j] + cins(mj);
   c2 = D[i,j-1] + cdel(pi);
   c3 = D[i-1,j-1] + cchange(pi,mj);
   c4 = D[i-2,j-2] + ctranspose(pi,mj-1);
   D[i,j] ← cfraction(j/l)· min(c1,c2,c3,c4);
  if D[i,j] > k
   return ∞;  // No distance below k

    return D[i,j]
```

7. Suchsystem nach Anspruch 3, **dadurch gekennzeichnet, daß** der dritte Algorithmus zum Ermitteln der exakten Stichwortübereinstimmung in einem dünnbesiedelten Stichwort-Abstand-Suffix-Baum $SST_{kws}(7)$ in Pseudo-Code wie folgt implementiert wird:

```
FindExact (root,p)
  i← 1;
   node ← Find Child (root,p_i);
   while node AND i ≤ length (p) do
     if IsLeaf(node) AND Suffix(node) = P_i...P_m then
      return node;
   i ← i+1
  node ← FindChild (node, p_i);

                    return NIL;
```

8. Suchsystem nach einem der Ansprüche 3 bis 7, **dadurch gekennzeichnet, daß** der vierte Algorithmus zum Ermitteln der exakten Übereinstimmung einer Stichwortsequenz in einem dünnbesiedelten Stichwort-Abstand-Suffix-Baum $SST_{kws}(T)$ in Pseudo-Code wie folgt implementiert wird:

```
MatchSequenceExact (P, root)
 matches ← Ø;
 v ← FindExact (p_1,root);
 if |P| > 1 then
  if v ≠ NIL then
   list ← UnorderedOccurrenceList (v);
   for all u ∈ list do
    if NextKeyword(u) = p_2 then
     if MatchRest (p_3...p_m,u) then
      matches ← matches □ Occurrence(u);

 return matches;

MatchRest (P,u)
 node ← u;
 for v ← p_1 to p_|P| do
  node ← NextOccurrence(node);
  word ← Keyword(node);
  if v ≠ word then
     return false;
```

9. Verfahren in einem Suchsystem für die Informationswiedergewinnung, insbesondere von Information, die in Form von Text gespeichert ist, wobei ein Text *T* Wörter und/oder Symbole *s* und Sequenzen *S* hieraus umfaßt, wobei die Informationswiedergewinnung mit einem gegebenen oder variierenden Grad an Übereinstimmung zwischen einer Anfrage *Q* und wiedergewonnene Information *R* abläuft, wobei die Anfrage *Q* Wörter und/oder Symbole *q* und Sequenzen *P* hieraus umfaßt und wobei die Information *R* Wörter und/oder Symbole und Sequenzen hieraus aus dem Text *T* umfaßt, wobei das Suchsystem eine Datenstruktur zum Speichern wenigstens eines Teils des Textes *T* und ein Maß *M* umfaßt, welches den Grad der Übereinstimmung zwischen der Anfrage *Q* und der wiedergewonnenen Information *R* mißt, und wobei das Suchsystem einen Suchalgorithmus zur Ausführung einer Suche realisiert, insbesondere einer Volltextsuche auf der Basis von Schlüsselwörtern *kw*, wobei die Information in dem Text *T* in Wörter *s* und Wortsequenzen S aufgeteilt wird, wobei die Wörter Unterstrings des gesamten

Textes sind, welche durch Wortgrenzterme getrennt sind und eine Symbolsequenz bilden, wobei jedes Wort als eine Symbolsequenz strukturiert ist, **gekennzeichnet durch**:

Erzeugen der Datenstruktur als einen dünnbesiedelten Wortabstand-Suffix-Baum $SST_{ws}(T)$ eines Textes T zum Darstellen aller Suffixe, die bei einem Worttrennsymbol in dem Text T beginnen;
Speichern von Sequenzinformation der Wörter s in dem Text T in dem dünnbesiedelten Wortabstand-Suffix-Baum $SST_{ws}(T)$,
Erzeugen eines kombinierten Editabstandsmaßes M, daß ein Editabstandsmaß $D(s,q)$ für Wörter s in dem Text T *und* ein Anfragewort q in einer Anfrage Q und ein wortgrößenabhängiges Editabstandsmaß $D_{ws}(S,P)$ für Wortsequenzen S in dem Text T und eine Wortsequenz P in der Anfrage Q umfaßt, wobei das Editabstandsmaß $D_{ws}(S,P)$ die minimale Summe der Kosten für Editieroperationen ist, welche eine Sequenz S in die Sequenz P umwandeln, wobei die minimale Summe der Kosten die minimale Summe der Kostenfunktionen für jede Editieroperation ist, gewichtet mit einem Wert, der proportional zur Änderung in der Gesamtlänge der Sequenz S ist, oder mit dem Verhältnis der momentanen Wortlänge und der mittleren Wortlänge in der Sequenz $S;P;$ und
Bestimmen des Grads der Übereinstimmung zwischen Wörtern s,q **durch** Berechnen des Editabstandsmaßes $D(s,q)$ zwischen den Wörtern s der wiedergewonnenen Information R und den Wörtern q einer Anfrage Q, oder in dem Fall, daß die Wörter s,q mehr als k Fehler voneinander abweichen, Bestimmen des Grads der Übereinstimmung zwischen den Wortfolgen $S_R;P_Q$ der wiedergewonnenen Information R bzw. einer Anfrage Q **durch** Berechnen des Editabstandsmaßes $D_{ws}(S_R,P_Q)$ für alle Übereinstimmungen.

10. Verfahren nach Anspruch 9, **gekennzeichnet durch** das zusätzliche Gewichten einer Editieroperation, die ein Wort s in ein Wort q ändert, mit einem Parameter für die Nähe zwischen den Zeichen der Wörter $s;q$, und somit Berücksichtigen der Ähnlichkeit der Wörter $s;q$ bei der Ermittlung der Kosten der in Rede stehenden Editieroperation.

11. Verfahren nach Anspruch 9, **gekennzeichnet durch** :

Begrenzen der Anzahl der Übereinstimmungen **durch** Berechnen des Editabstandsmaßes $D_{ws}(S_R,P_Q)$ für eine beschränke Anzahl von Wörtern in der Anfragewortsequenz P_Q.

12. Verfahren nach Anspruch 9, **gekennzeichnet durch**:

Definieren des Editabstandsmaßes $D(s,q)$ zwischen Wörtern s und einem Wort q auf rekursive Weise und Berechnen des Editabstandsmaßes $D(s,q)$ mit Hilfe einer dynamischen Programmierprozedur.

13. Verfahren nach Anspruch 9, **gekennzeichnet durch**:

Definieren des Editabstandsmaßes $D_{ws}(S,P)$ zwischen Sequenzen S und einer Sequenz P auf rekursive Weise und Berechnen des Editabstandsmaßes $D_{ws}(S,P)$ mit Hilfe einer dynamischen Programmierprozedur.

14. Verwendung des Suchsystems nach Anspruch 1 in einer Näherungs-Suchmaschine.

Revendications

1. Système de recherche pour la récupération d'informations, en particulier d'informations stockées sous forme de texte, dans lequel un texte T comprend des mots et/ou des symboles s et des séquences S de ces éléments, dans lequel la récupération d'informations a lieu avec un degré prédéterminé ou variable de correspondance entre une demande Q, où la demande Q comprend des mots et/ou des symboles q et des séquences P de ces éléments, et des informations récupérées R comprenant des mots et/ou des symboles et des séquences de ces éléments du texte T, dans lequel le système de recherche comprend une structure de données pour stocker au moins une partie du texte T, et un moyen de mesure M qui donne une mesure du degré de correspondance entre la demande Q et les informations récupérées R, et dans lequel le système de recherche met en oeuvre des algorithmes de recherche pour exécuter une recherche, en particulier une recherche de texte seulement sur la base de mots clefs kw, **caractérisé en ce que** :

la structure de données comprend une structure en arbre ayant la forme d'un arbre de suffixes peu dense et

irrégulier ST(T) pour stocker des suffixes de mots et/ou de symboles s et de séquences S de ces éléments du texte T ;

le moyen de mesure M comprend une combinaison d'un moyen de mesure de distance d'édition D(s,q) donnant un degré approximatif de correspondance entre des mots et/ou des symboles s,q respectivement du texte T et d'une demande Q et un moyen de mesure de distance d'édition $D_{WS}(S,P)$ donnant un degré approximatif de correspondance entre des séquences S de mots et/ou de symboles s du texte T et une séquence de demande P de mots et/ou de symboles q de la demande Q, le dernier moyen de mesure de distance d'édition comprenant des fonctions de pondération de coût pour des opérations d'édition qui transforment des séquences de mots et/ou de symboles s du texte T en des séquences P de mots et/ou de symboles q de la demande Q, la pondération ayant lieu avec une valeur qui est proportionnelle à un changement de la longueur de la séquence S lors d'une transformation, ou qui dépend de la taille des mots et/ou des symboles s,q des séquences S,P à comparer ;

les algorithmes de recherche mis en oeuvre comprennent un premier algorithme pour déterminer le degré de correspondance entre des mots et/ou des symboles s,q dans la représentation sous forme d'arbre de suffixes respectivement du texte T et d'une demande Q, et un deuxième algorithme pour déterminer le degré de correspondance entre des séquences S,P de mots et/ou de symboles s,q dans la représentation sous forme d'arbre de suffixes respectivement du texte T et de la demande Q, lesdits premier et/ou deuxième algorithmes recherchant dans la structure de données des demandes Q ayant la forme de mots, de symboles, de séquences de mots ou de séquences de symboles ou de combinaisons de ces éléments, de telle manière que les informations R sont récupérées sur la base d'une demande Q avec un degré prédéterminé de correspondance entre les informations et la demande ; et

les algorithmes de recherche comprennent également de manière optionnelle un troisième algorithme pour déterminer une correspondance exacte entre des mots et/ou des symboles s,q dans la représentation sous forme d'arbre de suffixes respectivement du texte T et de la demande Q et/ou un quatrième algorithme pour déterminer une correspondance exacte entre des séquences S, P de mots et/ou de symboles s,q dans la représentation sous forme d'arbre de suffixes du texte T et de la demande Q, les troisième et/ou quatrième algorithmes recherchant dans la structure de données les demandes Q sous la forme de mots, de symboles, de séquences de mots, ou de séquences de symboles ou de combinaison de ces éléments, de telle manière que les informations R sont récupérées sur la base de la demande Q avec une correspondance exacte entre les informations et la demande.

2. Système de recherche selon la revendication 1, **caractérisé en ce que** l'arbre de suffixes peu dense et irrégulier ST(T) est un arbre de suffixes peu dense dont les intervalles sont des mots $SST_{ws}(T)$ comprenant seulement un sous ensemble des suffixes du texte T.

3. Système de recherche selon la revendication 2, **caractérisé en ce que** l'arbre de suffixes peu dense dont les intervalles sont des mots $SST_{ws}(T)$ est un arbre de suffixes peu dense dont les intervalles sont des mots clefs $SST_{kws}(T)$.

4. Système de recherche selon la revendication 3, **caractérisé en ce que** le premier algorithme pour détecter le degré de correspondance de mots clefs dans un arbre de suffixes peu dense dont les intervalles sont des mots clefs $SST_{kws}(T)$ est mis en oeuvre au moyen du pseudo-code suivant :

```
FindApproximate (root, p, k)
   node ← root ;
   i ← 1 ;

nodes ← Children (node) ;   // A stack of nodes
for all v ∈ nodes do

if IsLeaf (v) then
   for j ← i to length (Suffix(node)) do
      Wj ← Suffix(node)j-i ;
      if Wj = '$' then   // '$' is a stopchar
       output W1...j ;
      return ;
if EditDist (i) = ∞ then
 break ;
else        //Internal node
   i ← i+1 ;
   Wi ← label (v)
if EditDist (i) = ∞ then
      break ;
   nodes ← Children (v) □ nodes ;
// end for

EditDistance(j)        //Calculates jth row
      for i ← 1 to length (P) do


       if pi = Wj then ∂ ← 0 else ∂ ← 1 ;
C1 = D[i-1, j] + Cins(mj) ;
C2 = D[i, j-1] + Cdel (pi) ;
C3 = D[i-1, j-1] + Cchange (pi, mj) ;
C4 = D[i-2, j-2] + Ctranspose (pi, mj-1) ;
D[i, j] ← Cfraction (j/l)..min (C1, C2, C3, C4) ;
if D[i, j] > k
      return ∞ ;   // No distance below k

return D[i,j]
```

5. Système de recherche selon la revendication 3, **caractérisé en ce que** le deuxième algorithme pour déterminer le degré de correspondance de séquences de mots clefs dans un arbre de suffixes peu dense dont les intervalles sont des mots clefs $SST_{kws}(T)$ est mis en oeuvre par le pseudo-code suivant :

```
ApproxSequenceMatch_ED (root, P(=p1, p2, ..., pm),k)
    m ← |p|
    matches ← ∅
startError ← 0
startIndex ← 1

while startError ≤ k OR startIndex ≤ m do
    startNode ← FindExact (PstartIndex) ;
    list ← UnorderedOccurrenceList (startNode) ;
    for all v ∈ lists do
        if ApproxMatchRest (v, P, k, startError) then
        matches ← □ v ;

startError ← startError + Cdel (PstartIndex) ;
startIndex ← startIndex + 1 ;
```

6. Système de recherche selon la revendication 5, **caractérisé en ce que** le sous programme ApproxMatchRest du deuxième algorithme est mis en oeuvre par le pseudo-code suivant :

```
     ApproxMatchRest (u, P, K, startError)
error ← startError ;
lastError ← startError ;
column ← 0 ;
node ← u ;
     for v ← p2 to P|p| do
node ← NextOccurrence (node) ;
word ← Keyword (node) ;
lastError ← error ;
error ← startError + EditDistance (column) ;
if error > k AND lastError > k then
     return false ;

     return true ;

     EditDistance (j)        // Calculates jth row
          for i ← 1 to length (P) do
          if pi = wj then ∂ ← 0 else ∂ ← 1 ;
     C1 = D[i-1, j] + Cins (mj) ;
     C2 = D[i, j-1] + Cdel(pi) ;
     C3 = D[i-1, j-1] + Cchange (pi, mj) ;
     C4 = D[i-2, j-2] + Ctranspose (pi, mj-1) ;
     D[i,j] ← Cfraction (j/l).min (C1, C2, C3, C4) ;
     if D[i, j] > k
     return ∞ ;   // No distance below k
     return D[i, j]
```

7. Système de recherche selon la revendication 3, **caractérisé** en que le troisième algorithme pour déterminer une correspondance exacte de mots clefs dans un arbre de suffixes peu dense dont les intervalles sont des mots clefs $SST_{KWS}(T)$ est mis en oeuvre par le pseudo-code suivant :

```
FindExact (root, p)
i ← 1 ;
   node ← Find Child (root, pi) ;
while node AND i ≤ length (p) do
```

```
if IsLeaf (node) AND Suffix(node) = Pi...Pm then
 return node ;
i ← i+1
node ← FindChild (node, pi) ;

RETURN NIL ;
```

8. Système de recherche selon l'une quelconque des revendications 3 à 7, **caractérisé en ce que** le quatrième algorithme pour déterminer une correspondance exacte de séquences de mots clefs dans un arbre de suffixes peu dense dont les intervalles sont des mots clefs $SST_{kws}(T)$ est mis en oeuvre par le pseudo-code suivant :

```
MatchSequenceExact (P, root)
   matches ← ∅ ;
   v ← FindExact (p1, root) ;
   if |P| > 1 then
   if v ≠ NIL then
      list ← UnorderedOccurrenceList (v) ;
   for all u ∈ list do
      if NextKeyword (u) = p2 then
       if MatchRest (p3...pm, u) then
        matches ← matches □ Occurrence (u) ;
   return matches ;

MatchRest (P, u)
   node ← u ;
   for v ← p1 to p|p| do
    node ← NextOccurrence (node) ;
    word ← Keyword (node) ;
    if v ≠ word then
     return false ;
```

9. Procédé de récupération d'informations dans un système de recherche, en particulier d'informations stockées sous forme de texte, dans lequel un texte T comprend des mots et/ou des symboles s et des séquences S de ces éléments dans lequel la récupération d'informations a lieu avec un degré prédéterminé ou variable de correspondance entre une demande Q, dans lequel la demande Q comprend des mots et/ou des symboles q et/ou des séquences P de ces éléments, et les informations récupérées R comprennent des mots et/ou des symboles et des séquences de ces éléments du texte T, dans lequel le système de recherche comprend une structure de données pour stocker au moins une partie du texte T, et un moyen de mesure M qui donne une mesure du degré de correspondance entre la demande Q et les informations récupérées R, et dans lequel le système de recherche met en oeuvre des algorithmes de recherche pour exécuter une recherche, en particulier une recherche de texte seulement sur la base de mots clefs kw, dans lequel les informations du texte T sont divisées en mot s et en séquences de mot S, les mots étant des sous-chaînes du texte entier séparées par des termes de limite de mots et formant une séquence de symboles, et dans lequel chaque mot a la structure d'une séquence de symboles,

caractérisé en ce qu'il comprend les étapes consistant à :

produire la structure de données sous la forme d'un arbre de suffixes peu denses dont les intervalles sont des mots SST_{WS} (T) d'un texte T pour représenter tous les suffixes commençant par un symbole de séparation de mots dans le texte T ;
stocker des informations de séquence des mots s du texte T dans l'arbre de suffixes peu dense dont les intervalles sont des mots $SST_{WS}(T)$;
produire un moyen de mesure de distance d'édition combiné M comprenant un moyen de mesure de distance d'édition D(s,q) entre des mots s du texte T et un mot de demande q d'une demande Q et un moyen dépendant de la taille des mots $D_{WS}(S,P)$ de mesure de distance d'édition entre des séquences S de mots s du texte T et une séquence P de mots q de la demande Q, le moyen de mesure de distance d'édition $D_{ws}(S,P)$ étant la somme minimum du coût des opérations d'édition transformant une séquence S en la séquence P, la somme minimum des coûts étant la somme minimum des fonctions de coût pour chaque opération d'édition pondérée par une valeur proportionnelle au changement de la longueur totale de la séquence S ou par le rapport de la longueur de mot courante et de la longueur de mot moyenne dans les séquences S,P ; et
déterminer le degré de correspondance entre des mots s,q en calculant la distance d'édition D(s,q) entre les mots s des informations récupérées R et du mot q d'une demande Q, ou dans le cas où les mots s,q sont distants de plus de k erreurs l'un de l'autre, déterminer le degré de correspondance entre des séquences de mots S_R, P_Q des informations récupérées R et d'une demande Q respectivement en calculant la distance d'édition D_{ws} (S_R, P_Q) pour toutes les comparaisons.

10. Procédé selon la revendication 9, **caractérisé en ce qu'**il comprend en outre l'étape consistant à pondérer une opération d'édition qui change un mot s en un mot q avec un paramètre de proximité des caractères des mots s, q, prenant ainsi en compte la ressemblance des mots s,q lors de la détermination du coût de l'opération d'édition en question.

11. Procédé selon la revendication 9, **caractérisé en ce qu'**il comprend l'étape consistant à limiter le nombre de comparaisons en calculant la distance d'édition $D_{WS}(S_R,P_Q)$ pour un nombre restreint de mots de la séquence de mots de demande P_Q.

12. Procédé selon la revendication 9, **caractérisé en ce qu'**il comprend l'étape consistant à définir la distance d'édition D(s,q) entre des mots s et un mot q de manière récursive et à calculer la distance d'édition D(s,q) au moyen d'une procédure de programmation dynamique.

13. Procédé selon la revendication 9, **caractérisé en ce qu'**il comprend l'étape consistant à définir une distance d'édition $D_{ws}(S,P)$ entre des séquences S et une séquence P de manière récursive et à calculer la distance d'édition $D_{ws}(S, P)$ au moyen d'une procédure de programmation dynamique.

14. Procédé consistant à utiliser un système de recherche selon la revendication 1 dans un moteur de recherche à approximation.

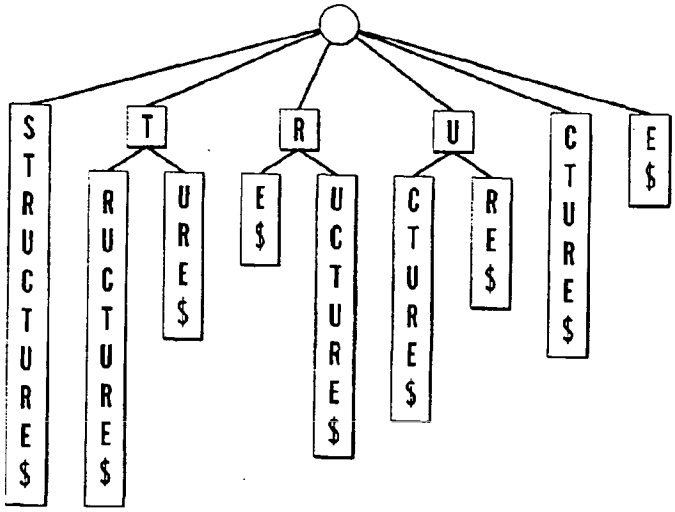


Fig. 1

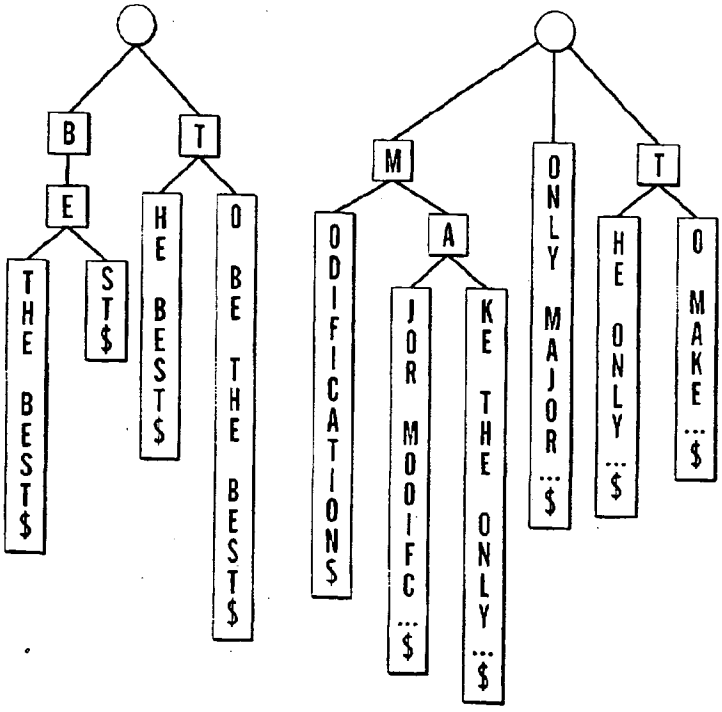


Fig. 2

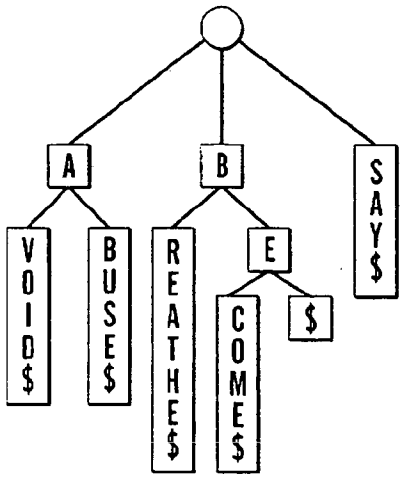

A
B
SAY$
VOID$
BUSE$
REATHE$
E
COME$
$

Fig.3

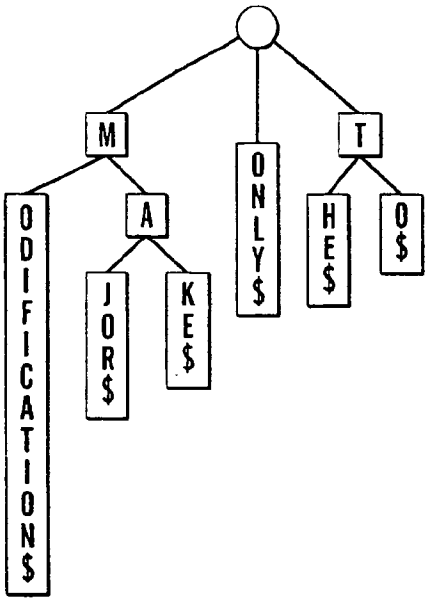
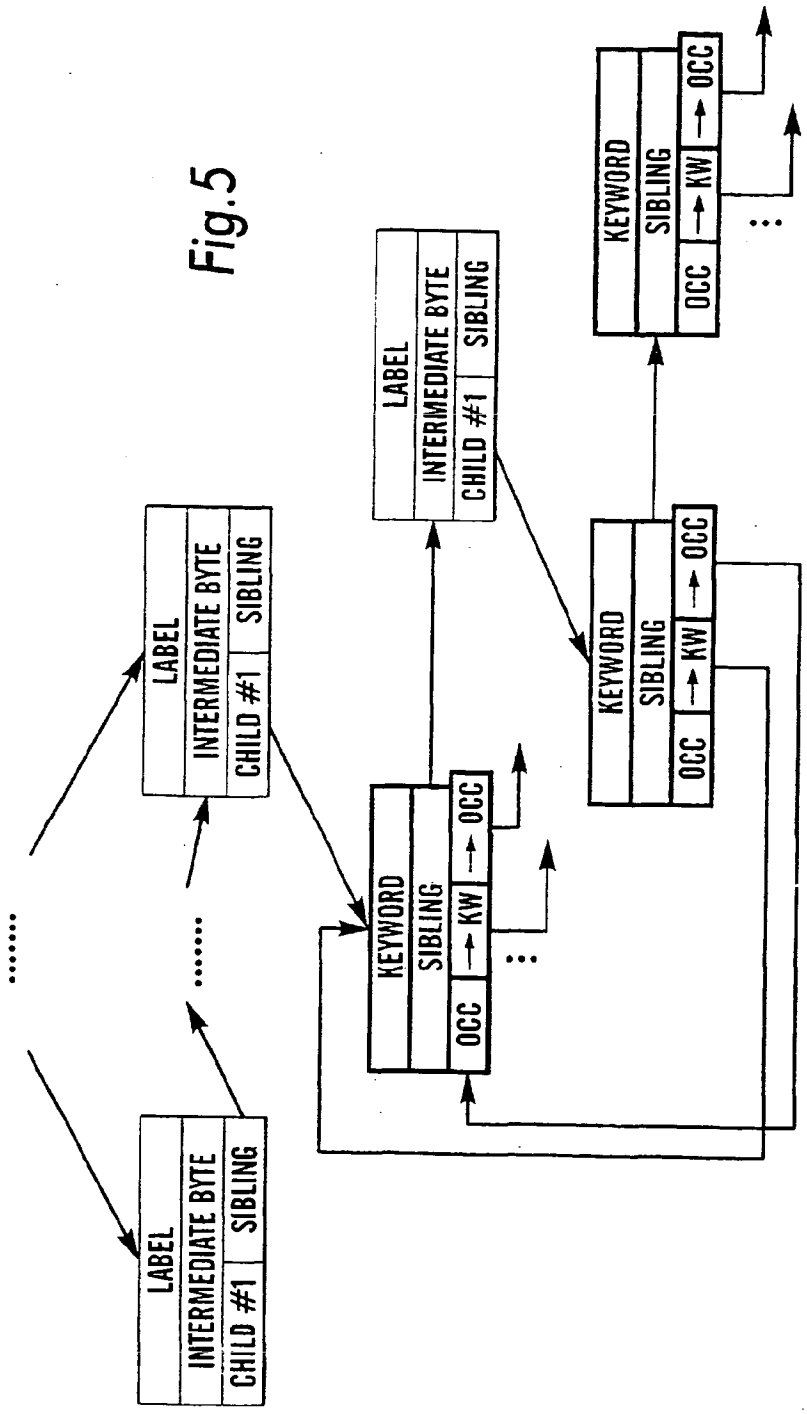

M
ONLY$
T
ODIFICATIONS$
A
HE$
O$
JOR$
KE$

Fig.4

Fig.5



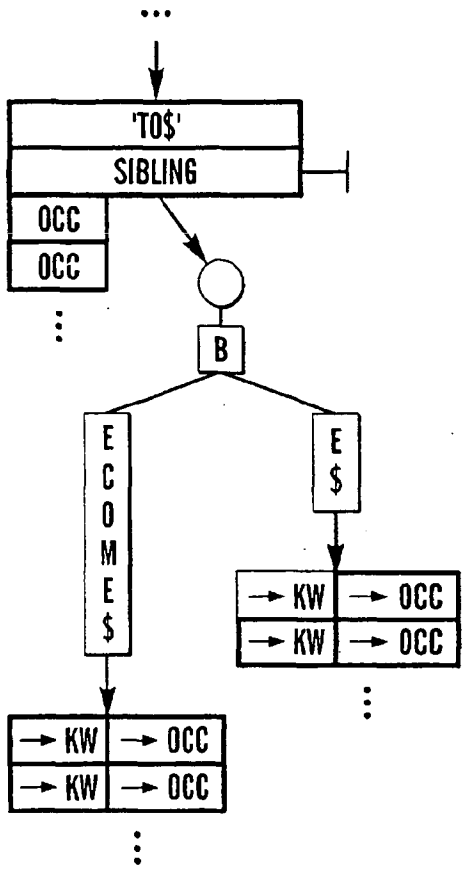


Fig.6

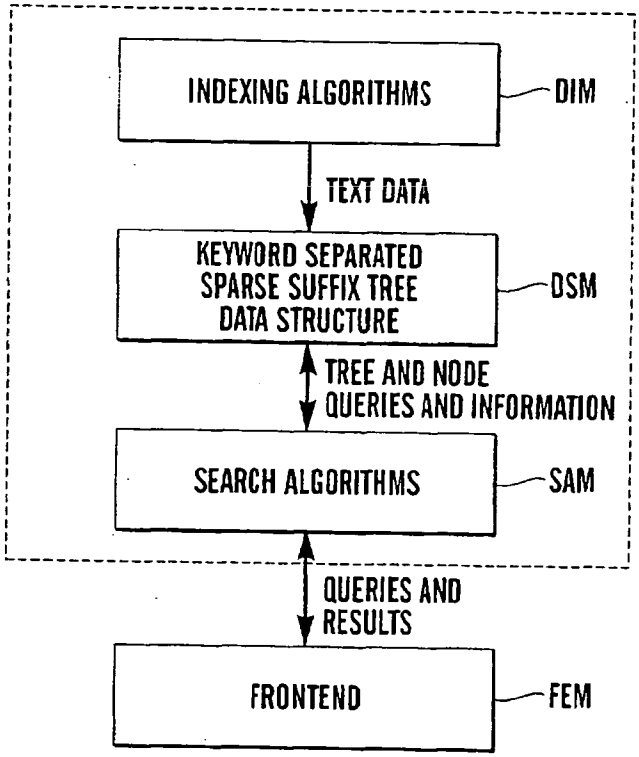


Fig.7

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 284 461 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
19.02.2003 Bulletin 2003/08

(51) Int Cl.[7]: **G06F 17/30**

(21) Application number: **02018111.1**

(22) Date of filing: **13.08.2002**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **13.08.2001 US 311857 P**
05.12.2001 US 683236

(71) Applicant: **Xerox Corporation**
Rochester, New York 14644 (US)

(72) Inventors:
- **Shanahan, James G.**
  Pittsburg, PA 15232 (US)
- **Grefenstette, Gregory T.**
  38610 Gieres (FR)
- **Fernstrom, Christer**
  38330 St-Ismier (FR)
- **Hubert, Laurence**
  38660 St Bernard du Touvet (FR)
- **Guerin, Nicolas**
  38000 Grenoble (FR)

(74) Representative: **Grünecker, Kinkeldey,**
Stockmair & Schwanhäusser Anwaltssozietät
Maximilianstrasse 58
80538 München (DE)

(54) Meta-document management system with user definable personalities

(57) A system operates using meta-documents which include document content associated with one or more personalities. Each personality is associated with a set of document service requests. Users are provided different techniques for creating personalities and modifying existing personalities. These techniques include: the use of an algebra to tailor existing personalities, the use of a list of links or documents to create a personality, the use of predefined personalities and knowledge levels in a field to create new personalities, the use of question answering techniques, and the use of learning personalities. Specified personalities are then used to enrich document content by integrating into corresponding meta-documents the results received from their document service requests.

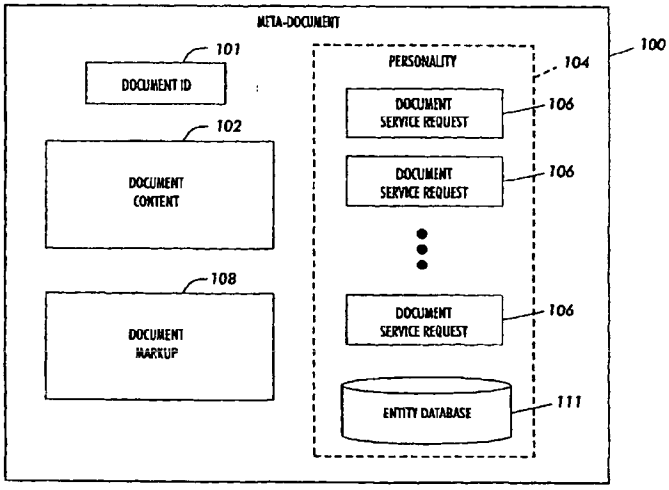


FIG. 1

EP 1 284 461 A1

Printed by Jouve, 75001 PARIS (FR)

Description

[0001] The invention relates generally to the management and use of documents, and in particular, to improved management and use of documents which may act as agents, generating requests for information, then seeking, retrieving and packaging responses to enrich the documents while facilitating reading comprehension, understanding relationships with other documents, and content creation. In particular this invention relates to a meta-document server with user definable personalities.

[0002] Knowledge management through document management forms an important part of the knowledge creation and sharing lifecycle. A typical model of knowledge creation and sharing is cyclical, consisting of three main steps: synthesizing (search, gather, acquire and assimilate), sharing (present, publish/distribute), and servicing (facilitate document use for decision making, innovative creativity).

[0003] Most systems consider documents as static objects that only acquire new content when acted upon by an authorized user. A user's decision to read and modify a document, or to run a program on it which may change its contents (for example, by adding hyperlinks), is needed for the document to acquire new information. This view of the document as a passive repository leads to the current situation in which documents remain static unless a user is in front of the screen piloting the system. OpenCola Folders™ offers one solution to the view of the document as a passive repository by creating folders on a user's computer that look for a limited set of document types, according to criteria set by the user (i.e., a single purpose information retrieval system).

[0004] Both agent-based systems and content-based retrieval systems provide some management of information without user intervention. An agent is a software program that performs a service, such as alerting the user of something that needs to be done on a particular day, or monitoring incoming data and giving an alert when a message has arrived, or searching for information on electronic networks. An intelligent agent is enabled to make decisions about information it finds. Both such systems, however, consider documents to be fixed and static entities.

[0005] Many products provide various solutions for individual aspects of the overall problem of knowledge management: anticipatory services, unstructured information management, and visualization of information and knowledge. Watson, for example, from the InfoLab at the University of Northwestern, is a program which operates while a user is creating a document. Watson retrieves information as the user works, from which the user can select for further investigation. Information retrieved by Watson comes from a service provider, and Watson stores the retrieved information in memory associated with Watson.

[0006] Also, Autonomy.com's ActiveKflowledge™ analyzes documents that are being prepared on the user's computer desktop and provides links to relevant information. In addition, online services such as Alexa.com, Zapper.com, and Flyswat.com suggest links that are relevant to the content currently viewed highlighted in a browser window. The suggested links appear in an additional window inside or separate from the current browser window. These services treat documents as static objects. Specifically, using Zapper.com's engine, when a user right clicks on selected text, words surrounding the selected text are analyzed to understand the context of the search request, and to reject pages that use those words in a different context.

[0007] Various products, such as commercial information retrieval systems, provide unstructured information, such as web pages, documents, emails etc. (which content may consist of text, graphics, video, or audio). Typical management services for unstructured information include: search and retrieval; navigation and browsing; content extraction, topic identification, categorization, summarization, and indexing; organizing information by automatic hyperlinking and creation of taxonomies; user profiling by tracking what a user reads, accesses, or creates create communities; etc. For example, Inxight's parabolic tree is an example of a system that organizes unstructured information and presents it in an intuitive tree-like format.

[0008] Furthermore, it is known how to embed executable code in documents to perform certain functions at specified times. For example, European Patent Applications EP 0986010 A2 and EP 1087306 A2 set forth different techniques in which to define active documents (i.e., documents with embedded executable code). More specifically, these publications set forth that executable code within the document can be used to control, supplement, or manipulate their content. Such active documents are said to have active properties.

[0009] Notwithstanding these existing methods for statically and actively enriching document content, there continues to exist a need to provide an improved document enrichment architecture that allows ubiquitous use of document enrichment services. Such an improved document enrichment architecture would advantageously provide methods for facilitating the use of such services by automatically attaching, monitoring, and suggesting such services for users.

[0010] In accordance with the invention, there is provided a system, and method therefor, for enriching document content. The system includes a user interface for specifying a personality defining a set of document service requests that identify enrichment themes. A user-manager receives document content and the personality specified at the user interface. The user-manager forms a meta-document with the document content and the personality. A scheduler selects a document service request from the set, for initiating and managing communication with a service provider to satisfy the selected document service. A content manager integrates results from the selected document service into

the meta-document as document markup. The user interface, the user-manager, the scheduler, and the content manager operate together to generate or recommend new personalities or modify existing personalities using one or more in combination of a set of algebraic computations, a set of document references, a set of predefined personalities, a set of learning personalities, a knowledge level, and a classification of personalities.

[0011] These and other aspects of the invention will become apparent from the following description read in conjunction with the accompanying drawings wherein the same reference numerals have been applied to like parts and in which:

[0012] Figure 1 is a schematic of a meta-document according to one embodiment of the invention;

[0013] Figure 2 illustrates a block diagram of a system incorporating a meta-document server;

[0014] Figure 3 is a schematic of meta-document enrichment according to one embodiment of the invention;

[0015] Figure 4 illustrates an example of meta-document enrichment as illustrated in Figure 3;

[0016] Figure 5 illustrates an electronic identification tag having a specified personality that is affixed or positioned proximate to a physical object;

[0017] Figure 6 illustrates an embodiment in which a hardcopy document has encoded thereon a personality identifier in embedded data;

[0018] Figure 7 illustrates a tag reader for receiving document identifiers from a mobile computing device or tag associated with a particular object;

[0019] Figure 8 illustrates a client interface for invoking a print command at a computer with enrichment selections;

[0020] Figure 9 illustrates a properties interface for the client interface shown in Figure 8;

[0021] Figure 10 illustrates a client interface for accessing the meta-document server shown in Figure 2;

[0022] Figure 11 illustrates a blow up of the window 1014 shown in Figure 10 for an architecture personality in which hay bale homes and tire homes personalities are selected;

[0023] Figure 12 illustrates an example of a properties window 1210 that is displayed when the properties configuration button 1022 is selected in Figure 10;

[0024] Figure 13 illustrates one embodiment of a client interface for creating and/or modifying personalities;

[0025] Figure 14 illustrates a client window for specifying properties of searches performed at the search engine defined in Figure 13;

[0026] Figure 15 illustrates another embodiment of a client interface for creating and/or modifying personalities;

[0027] Figure 16 illustrates a client interface for creating and/or modifying personalities by performing operations to groups of personalities;

[0028] Figure 17 is a flow diagram illustrating steps for generating a personality;

[0029] Figure 18 illustrates an example of an expanded document 1800, developed by descending two levels;

[0030] Figure 19 illustrates a form that can be used to create services;

[0031] Figure 20 illustrates four services that can be generated using the form shown in Figure 19;

[0032] Figure 21 is a flow diagram that depicts one method for filtering services at act 1716 in Figure 17;

[0033] Figure 22 illustrates a graphical representation of a selection process for selecting services with the highest similarity measure;

[0034] Figure 23 is a flow diagram that depicts another method for filtering services at act 1716 in Figure 17;

[0035] Figure 24 is a flow diagram that depicts one embodiment for identifying an answer of an instantiated question;

[0036] Figure 25 illustrates an example list of services available when an e-learning personality is selected to enrich document content;

[0037] Figure 26 illustrates an example list of services available when a language learning personality is selected to enrich document content;

[0038] Figure 27 illustrates a client interface for selectively specifying personality and/or service behaviors to entities recognized in specified content or documents;

[0039] Figure 28 illustrates a client interface for specifying different modes for determining when to annotate an identified entity;

[0040] Figure 29 is a flow diagram that sets forth the steps for propagating enrichment between electronic documents;

[0041] Figure 30 is a flow diagram for creating and updating an interaction history that are performed at act 2912 in Figure 29;

[0042] Figure 31 is a flow diagram for identifying what entities to markup at act 3008 in Figure 30;

[0043] Figure 32 illustrates the propagation of enrichment between accessed documents;

[0044] Figure 33 illustrates an interaction history;

[0045] Figure 34 illustrates the manner in which to apply pairs of entities and in addition identify third party entities;

[0046] Figure 35 illustrates entity types organized hierarchically;

[0047] Figure 36 illustrates a text categorizer;

[0048] Figure 37 illustrates a personality recommender;

[0049] Figure 38 illustrates the elements and flow of information for generating a query;

[0050] Figure 39 illustrates an example of a query contextualized using classification labels of document categorization hierarchy;

[0051] Figure 40 is a flow diagram which depicts one embodiment in which both categories and aspect vectors can be used to improve the accuracy of an information retrieval system;

[0052] Figure 41 illustrates a client interface similar to the client interface that illustrates an augmented query that can be performed using a recognized entity;

[0053] Figure 42 illustrates an information space that surrounds meta-document (i.e., a meta-document information space);

[0054] Figure 43 illustrates an auto-completion module that operates with a text editor and the meta-document information space;

[0055] Figure 44 illustrates an alternate embodiment in which an auto-completion module operates integrally with elements of the meta-document server shown in Figure 2;

[0056] Figure 45 is a flow diagram for creating and updating an entity database dynamically from the document information space;

[0057] Figure 46 illustrates a flow diagram for selecting words using the auto-completion system shown in Figure 44;

[0058] Figure 47 illustrates an example of the auto-completion process performed using the auto-completion entity database presented in Figure 48;

[0059] Figure 48 illustrates an example of an auto-completion entity database;

[0060] Figure 49 illustrates a document-centric auto-correction system that iteratively corrects errors in meta-document using a meta-document information space;

[0061] Figure 50 is a flow diagram for performing error correction using the system shown in Figure 49;

[0062] Figure 51 is a flow diagram depicting a process for identifying and correcting errors in document content for act 5026 shown in Figure 50;

[0063] Figure 52 illustrates a block diagram of the elements for forming a directed search;

[0064] Figure 53 illustrates an example of a user interface for invoking a directed search;

[0065] Figure 54 illustrates an example of the output of the directed search specified in Figure 53;

[0066] Figure 55 illustrates one embodiment of an interface for specifying a meta-document exchange;

[0067] Figures 56, 57, 58A, and 58B illustrate a detailed example of an export format;

[0068] Figure 59 illustrates another embodiment of a meta-document;

[0069] Figure 60 illustrates an embodiment of the contents of a personality;

[0070] Figure 61 illustrates an embodiment of the contents of a service request;

[0071] Figure 62 illustrates an alternate embodiment of the client interface shown in Figure 10;

[0072] Figure 63 illustrates a status window that displayed when enrichment is invoked for a specified document;

[0073] Figures 64 and 65 illustrate two examples of popup windows that appear when identified entities are selected;

[0074] Figure 66 illustrates an example of a document storing management view of a user's files;

[0075] Figure 67 illustrates an example interface for selecting document marking options; and

[0076] Figure 68 illustrates an example of an interface for configuring services.

A. Definition Of Terms

[0077] The terms defined below have the indicated meanings throughout this application, including the claims:

[0078] "Annotate" is used herein to mean to create a reference between an entity in a document, or region of a document, and some set of links, text segment, images, or embedded data (e.g., glyphs).

[0079] "Content retrieval" is used herein to mean an annotation that consists of content obtained by following a series of one or more links and retrieving their content, which content may be filtered or reformatted after retrieval.

[0080] A "document" is used herein to mean an electronic (e.g., digital) or physical (e.g., paper) recording of information. In its electronic form, a document may include image data, audio data, or video data. Image data may include text, graphics, or bitmaps.

[0081] Document "mark-up" is used herein to mean the annotation applied to a document.

[0082] A "document soul" is used herein to mean a personality that remains attached to a document for an extended period of time that may be indefinite or pre-specified of finite duration.

[0083] "Enrich" is used herein to mean to annotate a document in accordance with a predefined personality.

[0084] "Entity" is used herein to mean something recognized in a document (e.g., a person's name, a location, a medical term, a graphics entity that may include image data, graphics data, audio data or video data) that can be in the form of an image, text, embedded data, HTML, etc.

[0085] "Information space" is used herein to mean the entire set of annotations associated with an entity, a document segment, a document, or a set of documents.

[0086] A "lexicon" is used herein to mean a data structure, program, object, or device that indicates a set of words

that may occur in a natural language set. A lexicon may be said to "accept" a word it indicates, and those words may thus be called "acceptable" or may be referred to as "in" or "occurring in" the lexicon.

[0087] A "link" is used herein to mean, by way of example, a URL (Uniform Resource Locator) associated with a text segment or an image segment.

[0088] A "morphological variant" is used herein to mean the conjugated form of a word or expression (e.g., plural form), or a derivational form of a word (e.g., presidential is a variant of president). Morphological variants can be reduced to stems or lemmas using known techniques such as stemming algorithms such as Porter's algorithm or a lemmatization scheme in Inxight's LinguistX Platform.

[0089] A "personality" is used herein to mean a thematic set of services that can be applied to enrich a document.

[0090] A "service" is used herein to mean a program that provides new markup based on content and meta-data in a document in its current state. For example, the program may identify entities in a document, and annotate each entity with data associated to that entity (e.g., in a database). For example, a service may enrich a document with external information and/or add new services.

[0091] A "text segment" is used herein to mean a continuous sequence of bytes in a document, or a group of such segments.

B. General Features

[0092] A block diagram of a meta-document or "document soul" 100 is shown in Figure 1. The meta-document 100 includes an identifier 101, a content portion 102, which is a document created by a user or obtained by a user, and a personality 104. The personality 104 is a set of one or more document service requests 106 and an entity database 111. The entity database may include one or more separate entity databases, where each entity database identifies a class of entities (e.g., people names, city names, business names, etc.). In one embodiment, the personality 104 does not include the entity database 111 but instead includes document service requests that identify entities. In another embodiment, the entity database 111 records document-centric entities (i.e., entities that are related exclusively to the document content 102) that are specified by a user or by the system. It will be appreciated by those skilled in the art that the document service requests 106 and the entity database(s) 11 forming part of the meta-document 100 may include the content of a document service request and an entity database and/or may include references to a document service request and an entity database (in, for example, services database 210). The identifier 101 may include other administrative data such as creator, owner, size, access permissions, etc.

B.1 The Knowledge Management Cycle

[0093] Figure 2 illustrates a meta-document management system 201, within which the meta-document 100 is produced as the result of a knowledge crystallization process, where the process may last the lifetime of the document. Typically a meta-document's life begins with a focus and purpose which helps direct and refine the synthesis phase. During the synthesis phase, the meta-document 100 anticipates the information needs of the writer or reader, either independently through a pre-defined set of document service requests or by following specific or customized instructions, and performs the sometimes tedious tasks of searching, gathering, assimilating, and organizing information relevant to the document content.

[0094] The actions of the synthesis phase occur through the activation of one or more document service requests 106. Document service requests 106 may be activated while the user is creating or working on the meta-document 100 or when user has set aside the meta-document 100 so that the service requests can benefit from idle computer time, unused network bandwidth, etc. Activating a document service request 106 while the user works on the document has the additional advantage of allowing the meta-document to learn about the user's preferences. Document service requests 106 may be activated automatically by a scheduler 204 or manually by a user.

[0095] The next phase in the knowledge management cycle is concerned with sharing the information produced during the synthesizing phase. Typically the sharing phase consists of integrating the information gathered during the synthesizing phase into the contents of the meta-document 100 in a format useful for the user, person, or community that will use the document. The document content can be further enhanced for the user by assigning a personality to the document which marks up the document with information that eases the understanding of the content or that regularly provides more recent updates related to the content. The final servicing step in the cycle deals with periodic updates whereby the meta-document performs predefined service requests on behalf of the user. For example, the meta-document can keep up-to-date information of the temperature of an identified city.

B.2 Services

[0096] Referring again to Figure 2, one or more meta-documents 100 are stored in a meta-document server 200 at

meta-document database 202. In an alternate embodiment, document references (e.g., URLs) are stored in meta-document database 202 and their content referenced on network file server 220. Each meta-document 100 in the meta-document server 200 is endowed with a set of document service requests which each meta-document 100 exercises under control of a scheduler or scheduling demon 204, which wakes up each meta-document in database 202 in accordance with some predetermined time schedule. The scheduler 204 may be implemented in a software mechanism which accesses the document service requests 106, entity database 111, and content in a meta-document 100.

[0097] As illustrated in Figure 3, after the scheduler 204 wakes up the meta-document 100, the meta-document 100 informs the scheduler 204 of its current set of document service requests 301. Depending on the resources (e.g., service providers which can fulfill or satisfy a particular document service request) available to the meta-document server 200, the scheduler 204 chooses a document service request 106 to fulfill (indicated by arrow 300). Subsequently, the scheduler 204 invokes service providers 206 identified using services database 210 to satisfy those requests.

[0098] The services database 210 includes "service provider methods" for lookup and selecting service providers (including authentication data associated with each service), "entity methods" for identifying entities in document content using entity database 111 or entity databases in services database 210 or available as a network service 206, "notification methods" for notifying a user of new enrichment, regular expressions, lexicons, and a categorizer. In other embodiments, the services database 210 also includes content rights management methods.

[0099] Fulfilling a document service request means accessing a service provider from the services database 210 (e.g., selecting a service provider from a list of possible service providers) which includes some processes (or programs) that are invoked by the scheduler to access to the document content 102 (indicated by arrow 302) and document markup 108 (indicated by arrow 304). The results received from service providers 206 are integrated back into the original meta-document 100 by content manager 208. That is, these processes terminate by producing document-specific markup 108 (indicated by arrow 306) and/or new document service requests 106 (indicated by arrow 308), both of which are added to the meta-document 100 by content manager 208.

[0100] Various standards for attaching metadata exist, for example, DOM (Document Object Model) and XML (extended markup language) may be used. In one embodiment, both meta-document document service requests and resulting knowledge can be represented as XML metadata and added to the document at the end of each waking cycle. For example, a meta-document's document service requests are expressed as XML fields: <DSR> ... </DSR> (where DSR is short for "DOCUMENT-SERVICE-REQUEST"). For example, one document service can be expressed as: <DSR> who-am-i </DSR>.

[0101] In order to fulfill this document service request, the scheduler 204 invokes a "who-am-i" process stored in services database 210. This process uses the document content 102 and its document markup 108 as input. For example, the "who-am-i" process may return the filename of a document with identical content and additional document service requests to know who created this document, and where the document resides. The manager 208 marks the initial document service as fulfilled, or deletes it from the meta-document, and adds the additional knowledge returned by the process fulfilling the document service. In this example the following three lines are added to the initial document:

```
<MYNAME> xerox.txt </MYNAME>
<DSR> who-made-me </DSR>
<DSR> where-am-i </DSR>
```

[0102] During each operating cycle of the meta-document server 200, a meta-document 100 may acquire new markup 108 and new document service requests 106 as a function of document service requests that have been fulfilled. Some document service requests may add markup 108, and replicate the same document service request or other document service requests. Some document service requests may indicate to the content manager 208 markup 108 that should be eliminated when these requests are fulfilled.

[0103] In general, document service requests 106 correspond to services which add markup 108 to the document, based on the document's existence as a file in a file system; based on the content of the document as it was originally authored; and based on the content of the markup added to the document by some other process. When the document or the document's location is altered, the knowledge in the document may have to be regenerated or changed. Some knowledge, such as the relation of the contents of the document to other files in an intranet, or to documents found on the Internet may have to be update of periodically, possibly by some <DSR> continuous-education </DSR> document service request. For example, a textbook document may reference the population of the world and may need to be updated periodically to remain current.

[0104] Some document service requests may take a long time (for example, finding all the company names mentioned on a page and accessing all WWW pages mentioning two of those companies together). Other document services may be satisfied periodically (for example, finding the closing price of a stock share price). Besides document service requests, other functions not shown can be included in the meta-document server: a coordination system to orchestrate the concurrent execution of the functions described for the scheduler, a visualization and interaction system that allows various levels of display and interaction of metadata-enhanced documents, and a learning system that learns by observing the user interactions with the document. Likewise the meta-document 100 may be physically stored as a number

of destination files (e.g., a file corresponding to the original content 102, a file corresponding to markup 108, and a file corresponding to document service requests 104, which files may all be related by known naming schemes).

B.3 Personalities

[0105] The meta-document server 200 provides end-to-end solution for document-based knowledge creation and sharing in a customizable fashion. Customization is provided by the mechanism of personalities within a meta-document server. Personalities are assigned to a document thereby assisting a user in the acquisition, sharing and utilization of knowledge; this creates a document view of the world as opposed to a global view as in current web portals. One or more personalities can be attached to a document. Each personality thematically and/or contextually encodes a collection of documents service requests 106 which will allow the document to act autonomously on behalf of the creator or reader, anticipating the information needs of both the writer and reader of documents, keeping the document connected and up-to-date with the rest of information world.

[0106] A meta-document 100, for example, may be given a personality 104 that is: (a) inquisitive: a set of document service requests to find out more information about concepts present in the document content, find biographies of people mentioned in the content, (b) polyglot: search out translations of the words, terms and phrases contained in the document, (c) private: marked to keep the document's metadata invisible to other documents, (d) scientific: search for online versions of the papers cited in the document content, or (e) genealogical: looking for documents containing similar contents as itself.

B.4 Methods For Identifying And Using Entities

[0107] As shown in Figure 3, a personality 104 identifies one or more service requests 106. Each service request includes methods for: (a) recognizing entities in the document content 102; and (b) accessing a service using the recognized entities.

[0108] Entities include proper names (e.g., people, places, organizations, etc.), times, locations, amounts, citations (e.g., book titles), addresses, etc. Entities can be recognized using a variety of known techniques that may include any one or a combination of regular expressions, lexicons, keywords, and rules. A lexicon is typically a database of tuples of the form <entity-string, part-of-speech-tag, entity-type> where: an entity-string is the string characters that make up the entity (e.g., a person's name "John Smith"); a part-of-speech-tag, which is optional, denotes the grammatical usage of the entity (e.g., as a noun, noun phrase, verb, etc.); and entity-type denotes whether the entity belongs to one or more predefined classes (i.e., categories) of entities (e.g., person, organization, company name, etc.). A contiguous text string is recognized as an entity if the string is accepted as belonging to the lexicon.

[0109] Entities can be recognized by string matching or by using regular expressions. For example, a person's name could be recognized as two capitalized words. Regular expressions can be expressed in terms of the actual textual document content (i.e., words) or in terms of the linguistic markup associated with the textual content. This linguistic markup could include part of speech tags (such as noun phrases, nouns, etc.) or shallow parsing tags.

[0110] As an alternative means of recognizing entities some rules can be used. For example the following rule could be used to recognize proper names: if "word" is capitalized and is not in the lexicon (or dictionary, or thesaurus) then the word is a proper name.

[0111] Figure 4 illustrates an example in which a meta-document 100 is enriched using a personality 104 specified therein. At some predefined time or at pre-specified time intervals the scheduler 204 wakes up and identifies document service request 410. The scheduler then invokes the methods of the document service request referenced at 412 in the service database 210. The methods of a document service request referenced in the service database 210 may include regular expressions, lexicons, service provider selection, authentication data associated with each service, and content rights management. In executing the method identified by reference 412, a service is identified from network services 206 that recognizes entities from entity type "Company Name" also stored in services database 210.

[0112] Once the identified service is executed by the scheduler 204, it provides its results to content manager 208 which subsequently performs one or more of the following tasks: (a) marks the document service request 410 as completed at 414; (b) marks the document service request 416 as no longer waiting for input but waiting to be executed; and (c) inserts entities from the Company Name entity type and that appear in the document content 102 as well as their location(s) in the document content 102 at 418.

[0113] When the scheduler subsequently identifies the document service request 416, the scheduler will similarly identify a method 420 from service database 210 that will use the stock quote service to check the Company Name entities identified at 418. These stock quote results will similarly be inserted into document markup 108 and linked directly to the entities 418 which in turn link to locations in the document content. Alternatively, instead of inserting the stock quote results in 418, the service identified by request 416 will be inserted as document markup 108 to be initiated when the user accesses the identified entities in the document content 102.

C. Ubiquitous Personalities

[0114] This section pertains to methods for attaching personalities to documents, whether physical or in electronic form, and to objects, whether animate or inanimate. That is, depending on the particular form of the document, there exist different manners in which to preferably attach personalities thereto. Once a personality is attached, an enriched view of the document can be produced using the meta-document management system. Users of the meta-document management system can then be notified using the notification service when further enrichments of interest take place to the document. The general steps for attaching a personality to a document include: (a) uploading the document to a meta-document server; (b) attaching one or more personality to the document; (c) periodically enriching the document in accordance with the personality.

[0115] More specifically, in this section personality tokens and readers enable document enrichment to occur more ubiquitously and systematically throughout the life of a meta-document. In addition, in-transit enrichment services enable document enrichment to be invoked when a document service is initiated.

C.1 Personality And Service Tokens

[0116] A personality token records an identifier to a personality in personality database 212 shown in Figure 2. In one form, a personality token is an electronic tag that includes a digitally readable identifier. In operation, the digitally readable identifier of the electronic tag has prerecorded thereon with suitable graphic, symbolic, or textual indicia a personality identifier that is pre-associated with a predefined personality in the personality database 212 of the meta-document server 200.

[0117] Methods for transferring instruction sequences and information from one or more electronic tags to an electronic tag reader connected computer is disclosed in the following U.S. Patent Application Serial Nos., which are hereby incorporated by reference: 09/404,734; 09/391,898; 09/391,462. For example, electronic tag and tag reader systems can be based on temporary direct connection between a tag and a computing system (e.g., a magnetic card strip and card reader, or a small integrated circuit in a "smart card" with associated reader). Alternatively, the electronic tag is read by the electronic tag reader through a wireless infrared or radio frequency connection.

[0118] In operation, at least one electronic identification tag is located on or proximate to each physical object or location (hereinafter referred to together as physical item) that is to be assigned one or more personality identifiers. Figure 5 illustrates an example in which an electronic identification tag 502 is affixed or positioned proximate to a physical object 504 (e.g., the Eiffel tower). The tag 502 can be a small radio frequency transponder comprised of an integrated circuit, containing a unique user accessible identification number. A small coil inductively powers the tag, and an antenna is used to broadcast the personality identifier to an electronic identification tag reader 506.

[0119] The electronic identification tag reader 506 includes transmitter and receiver components that are integrated with in its computer system. The tag reader momentarily energizes the tag through its coil until it has sufficient power for transient transmission of its personality identifier. The communication between the tag 502 and the tag reader 506 only occurs when both are proximate, with an actual distance varying based on size of the antenna attached to the tag and to the transmitter, from a distance of a few inches to that of several feet.

[0120] Once the personality identifier is received, the tag reader 506 passes this on to its computer system as for example an ASCII string via some suitable connection, while simultaneously providing user feedback to confirm reading of the electronic tag. User feedback can be visual (e.g., blinking or turning on an LED status light, text based or iconic display presentations), auditory (e.g., an audible buzz or beep), tactile (e.g., a button being raised or a perceptible structure rotation), or combinations of the foregoing.

[0121] Upon receipt of the personality identifier, the computing system of the tag reader 506 determines the context in which the information is received, and identifies the appropriate meta-document(s) to attach the personality identifier thereto. Context information can include location and/or time information. For example, the context related to location may be determined using a GPS (Global Positioning System) in the tag reader 506 that identifies where the personality identifier is being received. This context information is used to assign a personality identifier to a meta-document based on information related to location. Also, context information related to time can also be used to assign a personality identifier to a meta-document based on information related to time. For example, depending on the time of year a personality identifier is read, the tag reader 506 assigns different personality identifiers to document content (e.g., a seasonally dependent document).

[0122] The tag reader 506 in one embodiment is programmed to use context information (i.e., location and time information) to assign a personality identifier to documents and/or document tokens on the tag reader 506 by using document metadata (e.g., document title, creation date, author, etc.) and/or document content. In one instance of this embodiment, a personality identifier read by tag reader 506 is assigned to each document or document token recorded on the tag reader 506 based on the time and/or region its content was last accessed and/or modified relative to the time and/or region the personality identifier is located. In this manner, personality identifiers are automatically associ-

ated with existing meta-documents by the tag reader 506. Alternatively, the user of the tag reader 506 may manually associate a personality identifier with a meta-document.

[0123] In either instance, the tag reader 506 may be embodied in a token-enabled mobile computing device. Token-enabled mobile computing devices are described in for example the following patent and patent applications, which are hereby incorporated by reference: U.S. Patent No. 5,862,321; U.S. Patent Application Serial No. 09/118,322 (entitled: "Token-Based Document Transactions"); and U.S. Patent Application Serial No. 09/270,320 (entitled "Secure Token-Based Document Server"). A transaction involves a series of steps that include: (a) a request for information from a source that will satisfy the request; (b) an estimate to satisfy the request from the provider; and (c) formalizing and submitting the request.

[0124] As shown in Figure 5, the token-enabled mobile computing device with tag reader 506 is bridged to the wire-based networks 516 and 522 through either gateway 508 or gateway 510, all of which forms part of network 221 shown in Figure 2. The mobile computing device 506 communicates with other wire-based or wireless devices using either an IR (infrared) transceiver or a radio (RF) transceiver integrated therein. The radio transceiver operates over any suitable wireless network using, for example, Bluetooth™ wireless personal area network (PAN) technology, PCS (Personal Communications Services), GSM (Global System for Mobile Communications), or pager messaging. The infrared transceiver uses, for example, communication standards set by the infrared data association (IrDA).

[0125] The wire-based network is further populated with a token-enabled server 526 to provide users of the mobile computing device 506 with access to document services available on wire-based networks 516 and 522. A user of token-enabled mobile computing device 506 is capable of browsing through directories of document tokens. These document tokens represent the user's documents stored on wired-based networks 516 or 522. In addition using a token-enabled mobile computing device, the user is able to apply document services available on networks 516 or 522 to selected document tokens, including but not limited to services offered by the meta-document server 200.

[0126] In one embodiment, one or more personality identifiers is embedded in a general document token along with one or more document references (e.g., URL's) that each identify a meta-document. In one embodiment, each personality is specified as a service parameter of the requested service, which is encoded as a service identifier in the general document token. (See specifically elements 36 and 38 in Figures 3A and 3B of U.S. Patent Application Serial No. 09/118,322, and their associated description.) Once the contents of a general document token embodying a personality identifier is communicated to the meta-document server and received by the user manager 214 (as shown in Figure 2), the user manager 214 then adds the one or more personalities to the referenced meta-document(s).

[0127] Document enrichment is initiated by the scheduler 204 as set forth above in accordance with the personality associated therewith. The enriched document is then made available to the owner of the document, either by placing the enriched document in the owner's list of enriched documents, or by storing the enriched document in a computer storage space available to the object referenced by the personality, or by emailing the location of the enriched document to the owner of the mobile computing device 506.

[0128] Further variants of this embodiment include adding a timestamp along with the specific GPS/GSM location and personality so that time-dependent information is also transmitted to the meta-document server 200. This information would allow for time-specific services to be activated (e.g., public displays related to the recorded time and place). Yet another variant adds a video camera at the location 504, and the image captured is matched against possible images associated with that GPS/GSM location, and a document specific to that identified image (e.g., a building) is enriched with the personality. Yet a further variant adds a laser distance-measuring device at the location 504 to pinpoint an exact location being viewed by the user possessing the GPS/GSM device, thereby providing more position specific enrichment to the user.

[0129] Figure 6 illustrates an alternate embodiment in which a hardcopy document 614 has associated therewith a personality identifier. In one embodiment, the personality identifier may be encoded thereon in embedded data 612. Embedded data is digital data carried by a document that is machine readable. In one representation of embedded data, a halftone pattern such as a serpentine halftone pattern is used to encode the personality identifier and document content (or reference thereto) as digital data in the halftone pattern. In this representation, a halftone cell is rotated depending on the particular value of the digital encoding required for the halftone cell.

[0130] Further details for forming serpentine halftone images are disclosed in U.S. Patent No. 5,706,099 to Curry, which is incorporated herein by reference. In an alternate representation of embedded data, hyperbolic serpentine halftone cells are used to encode the embedded data instead of circular serpentine halftone cells. Further details of hyperbolic serpentine halftone cells are set forth in U.S. Patent No. 6,081,345, which is incorporated herein by reference.

[0131] Another form of embedded data is data glyphs, which encode digital information in the form of binary ones and zeros that are then rendered in the form of very small linear marks. Each small mark represents a digit of binary data. Whether the particular digit is a binary one or zero depends on the linear orientation of the particular mark. U.S. Patent Nos. 5,091,966, 5,128,525, 5,168,147, 5,221,833, 5,245,165, 5,315,098, 5,449,895, and 5,486,686, which are all hereby incorporated by reference, provide additional information about the uses, encoding and decoding techniques

of data glyphs.

[0132] Referring again to Figure 6, the hardcopy document 614 includes document content 616 and embedded data 612. The embedded data 612 includes at least a personality identifier. In addition, the embedded data 612 may include a digital representation of the document content 616. In one operational embodiment, multifunctional (i.e., fax, scan, print, store, email) device 512 scans in a document 614 with embedded data 612. Once the personality identifier and document content is detected and converted to a digital form, the multifunctional device 512 transmits them to the meta-document server 200.

[0133] In an alternate embodiment, the personality identifier associated with a hardcopy document is encoded as part of the human readable content of the hardcopy document. In another embodiment, the personality identifier is encoded as part of a smart coversheet that gives a user the ability to select from one or more personality identifiers. Smart cover sheet are disclosed in U.S. Patent Application No. 09/746,913, which is incorporated herein by reference.

[0134] At the meta-document server 200, the document content and personality identified by the personality identifier is used to create a meta-document. As set forth above, the meta-document is enriched with content in accordance with the specified personality. Once enriched with content, the user that made the content enrichment request is notified by email as illustrated by reference number 618. In an alternate embodiment, the marked up content can be sent to multifunctional device 512 to be rendered on hardcopy output 620. In yet another embodiment, the user requesting the service may request both electronic notification and hardcopy output of the enriched document. Electronic notification can be performed using, for example, SMS (Short Message Service) text messaging, a paging service, etc.

[0135] Personalities may be alternatively developed using predefined service tags. In this alternate embodiment, the personality tag 502 represents a service and not a personality. In this embodiment, users capture a collection of one or more document service requests 106 that are stored in a service tag 502 using for example tag reader 506. Using the captured collection, the user defines a personality with it. This personality can then be attached to a document for enrichment in accordance therewith. Unlike personality tags, document service tags can be much more specific. In one embodiment shown in Figure 2, a hardcopy newspaper 230 includes a service tag 232 that identifies a service that will enrich a document in accordance the content from the newspaper of the service tag attached thereto.

[0136] In addition, services attached by personalities to document content may be content and/or media sensitive. For example, a personality may annotate a document depending on the format of the content (e.g., textual, graphical, and image) or form of the content (e.g., audio, video, static). Also, a personality may annotate document content differently depending on the media used to represent document content and/or enrichment. For example, document content may be annotated differently depending on the capabilities of device to which media is to be directed (e.g., whether the display screen is large or small, whether audio capabilities exist, etc.).

[0137] It will be appreciated by those skilled in the art that the personality (or personality identifier) may be textually and/or digitally recorded on a tag. In addition, it will be appreciated by those skilled in the art that the contents of the tag may be manually or semi-automatically recorded by a user and input into a mobile computing device. In one embodiment, the user of the mobile computing device may be permitted to compare the tag and select a personality from a list of personalities available on the mobile computing device that represents the tag. In another embodiment, the personality identifier on the tag is recorded in an image with a camera of a mobile computing device. Subsequently processing is performed on the recorded image to identify the personality identifier that may be digitally and/or textually encoded on the tag. In yet another embodiment, the content of the tag may be manually input by the operator of the mobile computing device.

C.2 Personalities Identified By Location

[0138] In this section, personalities are attached to document content from the location at which the request is made to enrich document content. In one embodiment, a personality is selected or suggested using a global positioning system. In another embodiment, a personality is assigned to a document token reading system at the location and attached to document references recorded by the document token reading system.

[0139] In a further embodiment, personalities are suggested using context such as the physical location of the user of a token-enabled mobile computing device 506 or the time and/or date at which the suggestion is made. In this embodiment, after selecting a document or document reference on a token-enabled mobile computing device 506, the mobile computing device converts positioning coordinates given by a GPS or GSM device or the like or a combination thereof (e.g., snaptrack.com), into a personality identifier. Personality suggestion is performed either at the mobile computing device or at the meta-document server using a lookup table that relates worldwide positioning information with personality identifiers. For example, a location in a city may be associated with personalities concerning particular monuments, streets, restaurants, buildings, or tour guides. The lookup table may be user specific so that time of day or week and/or position may reference either personalities that are directed at either work or personal interests. As set forth above in section C.1, once a personality is identified, the document content or reference thereto is transmitted along with the personality information to the meta-document server for content enrichment.

[0140] In yet another embodiment, physical locations are assigned a specific personality identifier that is related to a physical object at the location or something which is associated with that location. Figure 7 illustrates a device such as a tag reader 702 for receiving document identifiers from a mobile computing device 704 or a tag 706 associated with a particular object 708. Once the tag reader 702 or similar device (e.g., scanner) receives a document identifier or content, the system 705 coupled to the tag reader 702 directs a pre-assigned personality identifier and the document identifier or content to the meta-document server 200 for processing. As set forth above, the meta-document server after enriching the document content with the pre-assigned personality, either delivers a notice 710 by email to an identified user or places the enriched content in a folder associated with the particular location at which the tag reader 702 is positioned.

[0141] In yet a further embodiment, the tag reader 702 resembles a poker chip that includes a user identifier, a personality identifier, and a communications system for communicating with the meta-document server 200. A user to which the identifier is associated may have a plurality of these chips while working with document objects 708. When the user reads a particular document, the user can select and place any one of the plurality of poker chips upon the document to read tag 706. Upon receipt of the document identifier, the selected poker chip communicates the document identifier (or document content), personality identifier (or personality), and user identification (or user ID) to the meta-document server 200. Subsequently, the meta-document server 200 uses this information to enrich the identified document content with the identified personality and makes it available to the user upon completion. In this way, personalities may be readily applied to objects such as documents.

[0142] Context (e.g., physical location) in this section is used to identify or suggest a personality, unlike section C.1 where context is used to identify or suggest document content or a reference thereto that is to be enriched. In the example shown in Figure 7, the personality assigned to the tag reader 702 is one that is commonly used by a person working at the location. The personality identified tag reader 702 is used by the person to quickly input documents identified using tag 706 or mobile computing device 704 to meta-document server 200. It will be appreciated that in another embodiment context may be used to perform both the actions of identifying or suggesting a personality as well as identifying or suggesting document content to be enriched.

C.3 Transit Triggered Enrichment

[0143] Personalities may alternatively be automatically or manually specified at capture or in-transit using personality buttons. Referring back to Figure 2, a personality button is a button that is associated with a document capture, processing, and/or output device 218 that is programmable with one or more personalities from the personality database 212 of the meta-document sever 200 shown in Figure 2. The processing of a document involves any action performed on a document (e.g., move, copy, print, email, etc.). The device 218 may be coupled to network 221 permanently or temporarily. In addition, the device may alternatively be a mobile device 219 that communicates with the network 221 through gateway or tag reader 222. Examples of document capture, processing, and/or output device include a scanner, a camera, a printer, a display, a facsimile, an email client/server, SMS text messaging, etc.

[0144] In operation, after programming a personality button on the devices 218 or 219, the user has the option of selecting the button during document capture, processing, and/or output. When selected, a program is activated which associates the programmed personality with the document being captured, processed, and/or output. At a specified interval (e.g., after capture, after or before processing, or before output), the document content generated or input to the device is sent to the meta-document server 200 for enrichment. Once enriched, the document content is either made available to the user at the meta-document server 200 or delivered to a specified device for output.

[0145] In one embodiment, a user at a computer 226 is provided personality buttons when printing a document at a network device 218 coupled to the network 221. Figure 8 illustrates a client interface 800 for invoking a print command at the computer 226. In addition to well known print property settings, the client interface offers enrichment property buttons 802. The enrichment property buttons 802 enable a user to manually select a personality to apply to a given print request at 804 or have the meta-document server select a personality automatically for the user at 806. In addition, the enrichment property buttons 802 allow a user to apply the enrichment to selected pages or content at 808. Also, the enrichment property buttons 802 allow a user to specify whether the enrichment is inserted in the print request in the form of links or as additional content at 810.

[0146] Figure 9 illustrates a properties interface 900 for the client interface 800, which is invoked by selecting properties button 814 in the interface 800. The properties that may be set in the properties interface 900 are default enrichment properties that may be applied to any user print request. For example, the user is given the ability to specify whether enrichment should be provided as links or content at 901, and whether provide enrichment by printing it or storing it (on the meta-document server) at 902. If stored on the server the user is given the ability to specify at 904 whether to be notified by email when such enrichment is completed, or when significant changes occur in the document markup.

[0147] The properties interface 900 also provides the user the ability to specify an insertion point at 906. Advanta-

geously, enrichment may be automatically or manually set to occur at any insertion point of a document service request. A document insertion point is a point in the processing of a document at which a stage of processing has begun or ended. For example, in printing a document it may be converted into postscript before being directed to a specific printer. Thus, in this example insertion points exist before or after conversion to postscript.

[0148] If the insertion point is selected to occur after the document is rendered to postscript, the postscript driver is alerted to create a copy of the document before sending it to the selected printer. Subsequently, a personality to attached to the rendered postscript document. Both the postscript document and its attached personality are stored on the meta-document server 200 for enrichment to take place. If either the name of the document, and/or the identifier of the user who submitted the document to the device are available to the device driver, then the location of the personality-enriched document is communicated to the user via some notification service (e.g., email, smart cover sheet, etc.). Smart coversheets are disclosed in U.S. Patent Application No. 09/746,913, which is incorporated herein by reference.

[0149] In summary, the combination of the interfaces 800 and 900 provide a user with the ability to specify what content to enrich, when to enrich it, and in which form the enrichment should be provided to the user. It will be appreciated by those skilled in the art that additional enrichment property buttons 802 can be specified as part of the client interface 800 or properties interface 900, and that these properties can be interchanged between the two interfaces.

[0150] Personality buttons may appear in many alternate forms, besides those shown in the drawings. For example, a personality button can be added to a document capture devices such as scanners, printers, email clients, digital cameras, mobile phones, and community walls (such as described in U.S. Patent Application No. 09/746,914, entitled "Electronic Board System", which is incorporated herein by reference). The personality button may be a physical button or ones formed using software on a display screen of a device.

[0151] In the event an image is scanned at a scanner, the personality applied to the image using a personality button at the scanner's interface can be one that includes a service for identifying objects, such as people or buildings, therein. Identification can be performed using a variety of pattern recognition techniques. Once objects are identified, additional services automatically or manually selected can then be used to enrich the identified object (e.g., photo).

[0152] In another embodiment, a personality button is added to a Dictaphone or other voice-input capturing device. When the user records audio, the user has the option of selecting one or more programmable personality buttons to apply a personality to all or a portion of the recorded audio. Note that when there are multiple personality buttons, multiple personalities can be indexed to different portions of a continuous audio recording using audio indexing. Audio indexing is further described in U.S. Patent No. 5,321,396, which is incorporated herein by reference.

[0153] Once the recording of voice terminates, one or more personalities are attached to the recorded voice as specified using one or more personality buttons by the user. When invoked, a first service in one of the personalities converts the audio into text. The recorded audio is converted to text by the first service using well-known voice recognition software such as ViaVoice sold by IBM®. Subsequently, one or more additional services are applied to enrich the text in accordance with the attached personalities. If the audio is indexed with multiple personalities, then each indexed portion of the audio after being converted to text is associated with the particular personality that it is assigned, and delivered to the meta-document server 200 for enrichment in accordance with the indexed personality.

[0154] In variation of voice-input capturing using personality buttons, the user adds the personality to the voice capture by pressing an index button that marks the recording in some symbol (e.g., a predefined tune) to indicate that the subsequent name will identify a known personality. In yet another embodiment, a sequence of one or more index marks on the audio recording could be used to identify known personalities to specified sections of the audio recording. These index marks, which are identifiable by the voice recognition software could be recorded either audibly (e.g., "Add Personality") or with a special button on the audio capture device.

[0155] In yet another embodiment, a personality button is added to a video capture device (e.g., video camera) or display device (e.g., television). Similar to the audio capture device, one or more personality buttons can be used to apply one or more personalities to the video capture device or to the display using personality buttons on the device or a remote of the device while content is being captured and/or displayed.

[0156] In yet a further embodiment, video and/or audio is annotated directly or used to annotate textual content. For example, audio can be compared to other audio tracks. In identifying a similar audio track, the similar audio track can be used to annotate the audio to which it is being compared. Similarly, textual content, whether or not derived from audio data, can be annotated with audio and/or video tracks to further enrich the textual content.

[0157] In the case of video capture, video image data is matched against stored images, or decoded to identify the video from which the image was drawn. For example, screen credits can be captured and decoded by an OCR (Optical Character Recognition) program, and then the names matched against a database of movies and the video identified.

[0158] In the case of display, if the video is broadcast on a publicly available channel, the image can be matched against videos shown on those stations at the time of data capture. Alternatively, time and channel information of public broadcasts can be used to identify a video at the time a user selects a personality to apply to the video. Personality buttons in one embodiment can be part of a television remote control. Such personality buttons can be turned on and off while viewing to create index points associated with video content that associate one or more personalities to

different intervals of the program. In one embodiment, the video is not captured but instead is identified using the title of the video, and the text of the video thereafter enriched. Alternatively, the audio of the video production can be processed as set forth above using a voice input capturing device with personality buttons.

[0159] Once the video is identified, the text of the audio from a video recording is accessed in one embodiment from a script or subtitle database. The selected personality along with any index points (specifying particular sections to which one or more personalities are to be attached to the video) are associated with the retrieved text (at indexed points if specified) and stored in the meta-document server 200 for enrichment. For example, if the video is a film of a Shakespeare play, then the personality might be a play critic personality that would link up references in the play text to footnotes, glossaries, analyses, or liner notes. It will be appreciated by those skilled in the art that if the video is recorded with multiple personalities selected using index points, then the index points and text must be synchronized. In addition, it will be appreciated by those skilled in the art that in addition to audio from a video recording, images can be reduced to textual content using OCR programs.

D. Creating And Modifying Personalities

[0160] This section pertains to the formation and/or customization of individual or groups of personalities. It will be appreciated by those skilled in the art that the different methods described herein for forming and/or customizing personalities may be used on their own or in combination.

[0161] Figure 10 illustrates a client interface 1010 for directly accessing the meta-document server 200 shown in Figure 2. Such a client interface can operate at a user computer 226 or mobile computing device 219. In one form, the client interface 1010 is invoked by specifying an address (e.g., URL) of the meta-document server 200 in any conventional Internet or web browser. Other forms of the interface may be for example accessed using an application specific program.

[0162] After logging in on a login screen (not shown) through user manager 214, a user is given the ability to specify a location of a document to be uploaded and stored in meta-document database 202 at 1012. After the specified document reference at 1012 is uploaded and stored in the document database 202, a personality 1016 is selected from personality window 1014.

[0163] Once the personality 1016 is selected, the meta-document server may immediately and/or at a later point in time, depending on the document services specified in the personality, enrich the uploaded document content as described herein. In the event document content is immediately annotated with document services set forth in the selected personality, results are displayed in window 1018 and global service results in window 1020.

[0164] The personalities in window 1014 can be arranged in a variety of views that can specify private, shared, or public personalities. These personalities are recorded in the personality database 212 shown in Figure 2. Shared personalities may be given different access permissions (e.g., some users may be able to read or modify a personality while other may only be able to read a personality).

[0165] In addition, the window 1014 allows one or more personalities to be selected and simultaneously applied to enrich an uploaded document. In one embodiment, this is accomplished by selecting a single personality or a folder of personalities, as shown in Figure 10 at the analyst personality 1016. Alternatively, specific personalities can be selected to be applied to uploaded document content. Figure 11 illustrates a blow up of window 1014 shown in Figure 10 for the architecture personality in which hay bale homes and tire homes personalities are selected, at 1102 and 1104 respectively.

[0166] Invoking button 1022 on interface 1010 brings up a properties window for a user. Figure 12 illustrates an example of a properties window 1210 that is displayed when the properties configuration button 1022 is selected in Figure 10. In window 1210 a user is able to specify a default personality at 1212 or to have a personality recommended when a document is uploaded to the meta-document server 200 at 1214. A default personality at 1212 available to a user is "none", which if selected requires a user to specify a personality manually from the window 1014 after uploading a document. In addition, properties window 1210 allows a user to create and/or modify specific personalities by selecting button 1216, the details of which are discussed below in section D.1.

D.1 Generally

[0167] In one embodiment to create and/or modify personalities, a window 1310 shown in Figure 13 is revealed with two sub-windows 1312 and 1314 after selecting button 1216 shown in Figure 12. The first sub-window 1312 presents a list of all available personalities, while the second sub-window 1314 presents a list of categories of available services for a personality selected from sub-window 1312. In the example shown in Figure 13, the "watch business" personality is selected at 1316. Each category of services shown in sub-window 1314 is selectable to permit a user to specify one or more specific document services (e.g., information retrieval service 1318). [0214] Figure 14 illustrates a window 1400 with the information retrieval service 1318 in which searches can be selected for specific categories. In the ex-

ample shown in Figure 14, the categories of computing and people are selected at 1404 and 1408, respectively. In operation during enrichment, only those services that are selected are invoked. In addition, the searches performed by services will be limited to specified categories. That is, searches performed by a selected service can be limited to a specified category in the information provider's directory (e.g., Google™) of information content. For example, the service 1408 is limited to the "people" category of content of the information provider "Google".

[0168] In addition, Figure 14 illustrates that selected services can be updated or refreshed at 1450 on a periodic bases such as either a daily, weekly, monthly, or automatic bases at 1452-1455, respectively. The period of automatic updating at 1455 is determined using for example: (a) the history or access log of a browser (e.g., how recently the address of the services has been accessed by a user); and/or (b) monitoring results received from the service over a period of time and if they change at a frequent rate then set the refresh period to be frequent. Alternatively, a selected service can be specified to be updated never (i.e., to perform a single act) or until the end of a predefined period (e. g., until 2003), as illustrated at 1451 and 1456, respectively. In another embodiment not shown, the selected service can be specified to be updated. until a specified purpose expires (e.g., as long as a person is a minor). In yet a further embodiment not shown, the selected service can be specified to be updated until it no longer returns relevant results. Also, Figure 14 illustrates that cost can be defined for each service at 1440 as free at 1442 or for payment at 1444 for which a maximum amount may be defined.

[0169] In another embodiment to create and/or modify personalities, a window 1502 is revealed as shown in Figure 15 when the button 1216 is selected in Figure 12. In this embodiment, a user is given the ability to specify a name of a personality at 1504 and create it by either (a) modifying existing personalities at 1506, (b) using a selected set of files and/or folder with files at 1508, or (c) using content from a selected file or website at 1510.

D.2 Using An Algebra

[0170] In one embodiment, personalities can be specified through modification at 1506 in Figure 1 5 by tailoring existing personalities using an algebra. A specific personality can be tailored using an algebra that merges, adds, subtracts, composes (i.e., personalities that are composed together using a composition operator allow the results of one personality to be used as input of another personality), or intersects sets of two or more personalities. For example, it may be desirable to eliminate any references to computer science in a very general "tech watch" personality. Figure 16 illustrates an example of a user interface 1600 in which personalities are either added together or subtracted from one another, at 1602 and 1604 respectively, to form a new or modified personality.

[0171] In one embodiment, personalities are defined using a collection of tuples (i.e., set of ordered elements) of services S and lexicons L $[S_I, L_J]$. A first personality A and a second personality B are merged by forming the union of their tuples $[S_I^A, L_J^A]$ and $[S_K^B, L_M^B]$, respectively. If any of the services S_I^A or S_K^B in either personality are the same then the new service consists of $[S_I^A, L_N]$ where L_N is the union of L_J^A and L_M^B. In addition, a first personality A can be restricted by removing an existing personality B from it by creating a new personality by: (a) removing any services that are the same in both personalities A and B, and/or (b) subtracting the lexicon L^B from the lexicon L^A corresponding to that service. Alternatively, the techniques outlined in section D.3 can be applied to the services in selected personalities to select and/or organize the services of the new personality.

D.3 Using A List Of Links

[0172] In another embodiment, the meta-document server automatically generates on demand a personality using a specified set of documents or references thereto. For example, the set of documents could be defined using all of the files in a folder of a personal computer, where the set of files could contain textual content that is linked to or references other content (e.g., using hyperlinks). Alternatively, the set of documents could be identified using a predefined query such as an SQL query. In yet another embodiment, the set of document can comprise all of the documents in the meta-document information space 4200 (shown in Figure 44), which is described in detail below in section G.2.

[0173] Advantageously, the personalities generated are user-centric since they build on information that is explicitly selected by a user. In one embodiment, this service for automatically generating personalities is invoked in the window 1502 at 1508. Once a user specifies a set of documents at 1509 and initiates a request for the service at 1512, a process set forth in the flow diagram depicted in Figure 17 is performed by the meta-document server for generating a personality. It will be appreciated that in an alternate embodiment, the meta-document server uses this process on its own given a collection of documents identified by, for example, a search.

[0174] Initially at 1702, the personality creation process receives a specified set of documents and/or folders containing a set of documents. This set of documents is defined as a level N=0 document set. At 1704, all links are extracted from the level N document set. At 1706, content pointed to by the extracted links is fetched and used to define a level N+1 document set. At 1708, if additional levels are to be descended then the action at 1704 is repeated; otherwise, an expanded document is defined using the N document sets defined at 1702 and 1706.

[0175] More generally, the collection of N documents sets are referred to as an expanded document. The expanded document, which can be viewed as a list of documents, consists of documents selected by the user and the documents linked to those selected documents. Figure 18 illustrates an example of an expanded document 1800, developed by descending two levels from a level N=0 document set 1801. That is, the expanded document 1800 consists of the level N=0 document set 1801, a level N=1 document set 1803, and a level N=2 document set 1805. In this example, the level N=0 document consists of a single document with three links, that reference the documents in the level N=1 document set 1803.

[0176] Referring to Figures 17 and 18, an entity extractor 1802 constructs an entity database 1804 using the expanded document 1800 (which in one embodiment the entity extractor created), at 1710. The entity extractor 1802 includes generic rules for extracting entity types such as names of cities, people, products, dates, noun phrases, etc. These generic rules do not specify entities per se. Instead they specify generic entities that are capable of detecting that a capitalized noun is likely to be a person's name rather than a name of a city, which can be performed by using the context surrounding the identified noun.

[0177] Further details of entity extraction are described below in section G.2. In addition, an entity that is extracted is indexed to point back to the location at which it referenced. In addition, the entity database includes contextual information related to the use of the entity. An example of an entity database is shown in Figure 33 and described in more detail below.

[0178] Subsequently, the entity database 1804 is used by a service generator 1806 to generate document services or document service requests 1808. The combination of the entity database 1804 and document service requests 1808 are then used to define a new personality. This new personality can thereafter be applied to a document uploaded to the meta-document server and enriched as described above.

[0179] Initially at 1714, the service generator 1806 identifies and extracts all queryable forms in the expanded document 1800. Queryable forms can be identified by, for example, one or more tags. In one embodiment, each page of the expanded document 1800 is scanned for XML (EXtensible Markup Language) and HTML (HyperText Markup Language) forms. Typically, a form consists of input fields, choice fields such toggle buttons, menus, etc. HTML forms are described for example in "XForms 1.0" by World Wide Web Consortium (W3C) published on the Internet at http//www.w3.org/Markup/Forms.

[0180] At step 1716, the service generator 1806 creates at least one service for each form identified therein. In the event a page contains multiple forms, the service generator will generate multiple services to account for the different possible combinations of queries that could result. Further details of the act of creating services is set forth in section D.3.1 below.

[0181] At 1716, the service generator 1806 filters the services created at 1714 that likely provide little added utility. Services that add little or no utility return no results or irrelevant results. Methods for measuring the utility of incorporating a service (that was induced from a form) into a personality is set forth in section D.3.2 below. Both approaches rely on a Boolean or vector space retrieval model, a brief description of which is set forth in section F.1.4 below.

[0182] Finally, at 1718 a personality is defined using the filtered services and the entity database. The entities in the entity database are limited to the types of entities that provide utility as measured through the service to which they are associated (e.g., by measuring the utility of a word). As a further refinement, services are limited in scope to entity types for which they provide added value (i.e., return relevant results). Determining relevant results or added utility can be accomplished by filtering and ranking results after running a service that for example queries an information provider.

[0183] In one embodiment, filtering and ranking of results of a query returned by an information provider related to document content to which a personality is attached with the service is accomplished by: acquiring a list of hyperlinks and summaries ordered by relevance from the information provider; performing a similarity measure between the summaries and the context surrounding entities in the document content to which the query is directed; ranking the results based on the computed similarity measure; and filtering out only the highest ranked results.

[0184] In an alternate embodiment, the similarity measure is performed using document content referenced by the hyperlinks in addition to the summaries. Also, a Cosine distance metric or a correlation measure can be used to measure the similarity between the content acquired from the information provider and the related document content.

[0185] It will be appreciated by those skilled in the art that before using a Cosine distance metric the document content are converted to features (e.g., word, word phrases, etc.) and stemmed. In addition it will be appreciated by those skilled in the art that distance measurements can be performed on originally extracted features that are remapped to define a reduced feature space using latent semantic indexing (LSI).

[0186] In one embodiment, a personality created at 1718 by packaging the services filtered at step 1716 is ranked and organized hierarchically into groups using the hyperlinks of the specified services. Such ranking and organizing can be performed using known ranking, (agglomerative) clustering or hyperlink techniques. An example of hyperlinked techniques is disclosed by Kleinberg, in "Authoritative Sources In A Hyperlinked Environment," IBM Technical Report RJ 10076, May 1997.

[0187] In an alternate embodiment, the organizational structure of the documents with hyperlinks specified at step

1702 is used to create one or more personalities at step 1718 (e.g., a personality for each branch in a hierarchical collection of documents is created). These personalities could be organized as shown in Figure 11. It will be appreciated by those skilled in the art that the steps set forth in Figure 17 may be partially or entirely automated.

[0188] In another embodiment, the expanded document 1800 is further developed by attaching a generic personality thereto. The generic personality could be applied to one or more levels of the expanded document and only depending on whether there exists a need for further expansion of the information space surrounding the original document content at level N=0. For example, in one instance the generic personality is applied only if the expanded document references less than a predetermined threshold number of documents.

D.3.1 Creating Services

[0189] In one embodiment, each service created at step 1714 is created with the following properties: (a) the service is specified such that it takes as input a new text segment identified for example by a document reference (e.g., URL); (b) the service includes methods for recognizing entities and their offsets in the new text segment or accepting recognized entities and their locations from another service; (c) the service includes methods for associating the recognized entities from the new text segment with (i) the concepts in the retrieved content of the N level document sets, (ii) an instantiated query (i.e., a concrete instance defined therefor) involving the recognized entity and a form, and/or (iii) a result (possibly reformatted or filtered) of the instantiated query (ii) with the recognized entities at (b); and (d) the service includes methods for returning a list with the recognized entities and their original offsets at (b) and the newly associated information at (c). In an alternate embodiment if a service for a specified hyperlink already exist then no new service is created and the existing service is used.

[0190] In cases c(ii), a form may be instantiated as follows. If the form contains one input field and one or no submit buttons, then the input field is filled with the recognized concept or entity, and the form submission protocol (e.g., GET or POST as disclosed in XForms 1.0) is followed with the filled-in input field. In case the form contains more than one field, then either all possible combinations of fields and recognized entities or concepts are created for submission, with only those producing non-null results being returned. Alternatively, the form may be filled in using automated techniques such as those disclosed on the Internet at www.roboform.com.

[0191] By way of example, consider form 1900 with input field 1901 and toggle buttons 1902-1905 depicted in Figure 19. In this example, the form 1900 is a front end for a content provider of scientific material. The form is composed of a text field 1901, where the user is expected to input one's query and toggle fields 1902-1905, where the user can indicate to the system, in which folder the query should be executed (e.g., by selecting the all-folders toggle button 1902, the system will search the entire content of the content provider). In this example, the following four different services shown in Figure 20 would be generated. Each service includes the input field 1901 and one of the four toggle buttons 1902-1905 in an activated state. Each service is associated with a particular type of entity, which is determined using method disclosed below in the following section D.3.2.

D.3.2 Filtering Services

[0192] The purpose of filtering as set forth above is to remove services that have been created but that have little or no utility. More specifically, given an entity database and a list of extracted services, three different utility measures are set forth below to determine the utility of a list of services. It will be appreciated by those skilled in the art that one or a combination of the three utility measures can be used. Generally, each utility measure ranks the services according to their potential usefulness for entities in the entity database.

[0193] A first utility measure is set forth in Figure 21, which depicts a flow diagram of the act at 1716 of filtering services created at 1714 using the entity database created at 1710. Initially at 2140, a list of services and an entity database are received. It is assumed that the entity database includes the frequency of occurrence of each entity in the expanded document.

[0194] Each service provides means for accessing an information service provided by a content provider. The frequency of entities in the entity database of the databases provided by content providers accessed by each service in the list of services is computed at 2142-2146. More specifically at 1242-2146, the following statistics relating to the frequency of each entity in a service are computed: (a) f_{ij} the number of documents in the database of content provider CP_DB_i that contain entity (or feature) f_j; and (b) w_{ij} the sum of the weights of each feature f_j over all documents in the database of content provider CP_DB_i. At 2144, each content provider is represented as a list of tuples of the form <entity, frequency, weight>, where frequency and weight are as "f" and "w" defined above. The expanded document is represented using a similar list but in this case f denotes the number of documents in which the entity occurs and w denotes the sum of the weights of each entity over all documents.

[0195] In one embodiment, information relating to the frequency of entities in services can be acquired for each service by running periodically a system that constructs a query for each feature f_j that are executed at content provider

CP_DB_i and subsequently extracts the values f_{ij} and w_{ij} from the returned results of the query. Extractors for values f_{ij} and w_{ij} can be constructed automatically using wrapper approaches or Hidden Markov Models (HMMs).

[0196] Approaches for generating wrappers is disclosed in U.S. Patent Application No. 09/361,496, which is incorporated herein by reference. Additional information regarding wrapper generation is disclosed by Chidlovskii et al. in: "Automatic Wrapper Generation for Web Search Engines", Proc. 1st Intern. Conf. on Web-Age Information Management, WAIM'2000, LNCS Series, Shanghai, China, June 2000; and "Wrapper Generation via Grammar Induction", 11th European Conference on Machine Learning, ECML'00, Lect.Notes Comp. Science, Vol. 1810, Barcelona, Spain, May 2000.

[0197] In another embodiment, the STARTS protocol is used to export summaries from the content provider to provide information relating to the statistics of entities in services. STARTS is a protocol proposal for Internet searching coordinated by Stanford University, that involves private and public organizations. STARTS specifies that content providers should export summaries that include entity statistics f_{ij} and w_{ij}. Details of the STARTS protocol are described by Gravano et al., in "STARTS: Stanford proposal for Internet meta-searching", Proceedings of the 1997 ACM SIGMOD Conference, 1997.

[0198] At 2152, any of a number of well-known similarity measures can subsequently be used to measure the similarity between each service and the expanded document. For example the Cosine distance metric can be used. Alternatively, a correlation measure could be used at 2152 to measure similarity. For more background relating to the computation of distance metrics see "Foundations of Statistical Natural Language Processing" by Manning and Schutze, MIT Press, 1999. In addition, see section F.1.4 herein that describes a correlation measure in terms of entities and associated frequencies and weights.

[0199] In yet another embodiment, the entities and associated frequencies (i.e., similarity for entities and weights) could be remapped to define a reduced feature space using latent semantic indexing (LSI) (for background relating to LSI see articles authored by Dumais available on the Internet at http://www.cs.utk.edu/~lsi/), thereby overcoming problems associated with synonyms and polynyms (i.e., same word has different interpretations depending on the context). Subsequently in this alternate embodiment, similarity measures can be carried out in this reduced feature space.

[0200] At 2154, the top N services (i.e., with the highest similarity measures) could then be selected as the services and incorporated into the new personality. Figure 22 illustrates a graphical representation of this selection process with an expanded document and two services A and B. The horizontal axis of the graph sets forth each entity in the entity database (i.e., $e_1 \ldots e_n$), and the vertical axis sets forth the weighted frequency of occurrence of each entity. In the example shown, service A has a greater degree of similarity than service B to the expanded document.

[0201] Another utility measure ranks the list of services after acquiring entity statistics using a utility measure as set forth in the flow diagram show in Figure 23. Figure 23 sets forth a method for filtering services at 1716. Initially at 2355, a list of services and entity database are received. At 2356 a next service in the list of services is selected, and at 2357 a next entity is picked from the database of entities. At 2358, a query is formulated for the selected service using the selected entity as set forth above. At 2359, the query is submitted to the service. Using the top N results of the service at 2359, a similarity measure between the entity and contextual information related to the selected entity and each of the top N results is computed at 2360, as follows:

$$EntityUtility(Entity, Service) = \sum_{Doc \in TopMatchesForService} Similarity\ (Entity, Doc),$$

[0202] where "entity" is one of the entities in the entity database; "service" is a service; and "doc" is one of the N top results.

[0203] More specifically, "entity" in the equation denotes both an entity string and a surrounding context. For simplicity it may be assumed that an entity only occurs in one location in the expanded document. The surrounding context for an entity can be determined in a number of ways using known parsing techniques that delimit sentences, paragraphs, etc. For example, techniques for determining the context surrounding an entity include: (a) letting the context be the textual content of the whole document, which forms part of an expanded document, be the context; (b) letting the context be the sentence in which the entity string occurs; (c) letting the context be the paragraph in which the entity string occurs; or (d) letting the context be the topic text in which then entity string occurs as detected by known topic detection techniques.

[0204] Also in the equation, "doc" refers to either the document summary that appears (as an element in a result list) in the results page of the service or alternatively to the entire document, from which the summary was derived. The similarity measure can be performed using either resulting form. In this equation a similarity measure is generated for each entity (represented as the entity plus a context) and result document "doc" (represented as a summary or the entire document content).

[0205] In order to compute such a similarity measure both the entity and the result document are first processed as

follows: (a) stop words are eliminated; and (b) each word is stemmed using known stemming techniques such as Porter's stemmer. Subsequently, a similarity measure such as the Cosine measure could be used to calculate the degree of similarity between the entity and the result document based upon text features (for details of text features see U.S. Patent Application Serial No. 09/928,619, entitled "Fuzzy Text Categorizer" which is incorporated herein by reference).

[0206] In an alternate embodiment, the text features are transformed using LSI into a reduced features space. This LSI transformation is calculated using entity and entity frequency database that is extracted as described above. Having transformed the features using LSI, a similarity measure such as a Cosine distance measure can be used to calculate the similarity between the entity (and its context) and the resulting document "doc".

[0207] In the instance in which an entity occurs in multiple contexts exist for an entity (i.e., the entity exists in multiple locations in a document or expanded document), each location of the entity and its associated context are treated separately (i.e., as different entities).

[0208] At 2361, if it is determined that the last entity in the entity database has been examined, then the measured similarities are summed for all the entities related to the selected service at 2362 as follows:

$$ServiceUtility(Service) = \sum_{E \in EntityDB} EntityUtility(E, Service),$$

[0209] where E is an entity in the entity database, and service is a service. At 2363, if this is performed for all services, then the top N services are selected with the highest service utility measure to specify the filtered services; otherwise, the process continues at 2356 with the next service in the list.

[0210] Services can be organized in a number of ways such as flat or hierarchically. The services as represented in these ways could be clustered and a representative service could be selected from each cluster. In this embodiment, a multi-dimensional graph is defined with one dimension for each entity in the entity database. The frequency of each entity occurring in the expanded document and the services are plotted against each other. Clusters are formed and associated with a service. These clusters can then be used to hierarchically organize the services.

[0211] In an alternative embodiment, a generic service is applied to the expanded document subsequent to act 2363. The generic service uses the contents of the expanded document to query a general purpose information provider instead of an information provider that specializes in a specific subject. In yet another embodiment, a service utility is computed for an entity type instead of for all entity types as described above. In this alternative embodiment, the utility of services can be evaluated for particular types of entities. For example, a service utility is computed for the entity type biology 2002 for the service 2004 shown in Figure 20.

D.4 Using Predefined Personalities And Knowledge Levels

[0212] In yet a further embodiment, a relative ability or existing knowledge level in a field may be specified as shown at 1516 in Figure 15. The specified knowledge level 1516 can be used for example to create new personalities that access different levels of service providers from predefined personalities specified at 1504. For example, with a personality directed at medical information, if knowledge of someone is novice (i.e., a layman) then more basic information providers are specified and more basic definitional services are specified in the personality. In addition, the knowledge level can be used to either include or exclude entities from an entity database that is used to create a personality (as set forth above in section D.3). For example, a expert in the medical field may not be interested in the same entities that a novice in the medical field would be.

[0213] Besides providing a knowledge level of desired personality, a hint (i.e., subject hint) is given to the type of personality that is desired as shown at 1514 in Figure 15. Upon receiving a hint, the meta-document server relates the hint of the desired personality to a set of actions that are specifically related to subject matter of the hint. Generally, the hint 1514 can be used to improve any of the methods for creating personalities that may be specified in Figure 15. The hint 1514 and knowledge level may be used individually or in combination.

[0214] In one specific example, if a hint 1514 of a medical personality is specified to the meta-document server along with document content referenced by the hyperlinks at 1508 or name at 1510, then the meta-document server 200 creates a personality by identifying services that enrich the identified content relating to the following: (a) an access to a general pharmaceutical guide for drugs mentioned in the document content; (b) medical records related to the user and to the items mentioned in the document content; (c) images, video clips, etc., associated with items mentioned in the document content from a medical database; (d) links to a community of sufferers for any illnesses mentioned in the document content; (e) alternative products to those mentioned in the document content; (f) connections to online drug stores; (g) connections to current research in any of the areas mentioned in the document content; (h) information on any companies mentioned in the document content; and (g) any other medical information related to the items found

in the document content.

[0215] In another specific example, given a hint 1514 that is a construction personality, the personality is created by the meta-document server 200 by identifying document content referenced by the hyperlinks at 1508 or name at 1510, and identifying services that enrich the identified content relating to the following: (a) building codes, zoning laws, property evaluations and other legal documents concerning the items (e.g. addresses) identified in the document content; (b) images (photos, diagrams, blueprints) of the items (e.g., buildings, materials) mentioned in the document content; (c) history (e.g., social, constructor, tenancies, etc.) relating to the document content; (d) similar buildings in the world, architects building such buildings; (e) neighboring buildings, tenants, etc.; (f) simulations of the areas/buildings, mentioned under certain conditions (e.g., earthquake, fireproof); (g) maps of the areas mentioned in the document content; (h) sensor devices (e.g., web cams, thermometers, etc.) of the areas mentioned in the document content; and (i) costs, suppliers, retailers, delivery rates, technical specifications, tutorials, etc. for materials mentioned in the document content.

D.5 Using Information Extraction Techniques

[0216] The meta-document server as described above enriches (e.g., marks up) document content with results from different services. Typically these results are list of documents, lists of summaries, extracted information typically of a very simple structure nature. For example, results may include stock quotes and biographic entries. In this section a method is described that extracts information of a more sophisticated nature from unstructured text. This is achieved using information extraction techniques such as question answering.

[0217] In one information extraction technique, personalities can also be created and/or modified using predefined questions that can be used in conjunction with a lexicon or lexicons associated with a personality to create one or more question forms. Each question form is used to create a new document service request that is satisfied using a known question answering system that uses a combination of information retrieval and syntactic or pattern matching techniques.

[0218] In one embodiment, question forms are created automatically using an input question defined by a user at 1520 in Figure 15. For example, if the question is "What is the procedure for ablation of the liver?" and the specified personality at 1504 includes a lexicon that is body organs, which includes the word "liver", then the meta-document server would identify the body organ found in the question 1504 (e.g., liver) and replace it with a generic symbol representative of the identified lexicon. In this specific example, the word "liver" would be replaced with the generic symbol <BODY_ORGAN> to produce the question form "What is the procedure for ablation of the <BODY_ORGAN>?" Alternate question forms can be defined using the same question for the example given above depending on how many alternate lexicons are defined in the specified personality.Thus, with the same question, but with a different lexicon, for example of surgical procedures, the question form can be defined: "What is the procedure for <SURGICALPROCEDURE> of the liver?"

[0219] Yet another question form can be produced using the same question if the personality included both lexicons for body organs and surgical procedures. This would produce the question form: What is the procedure for <SURGICAL_PROCEDURE> of the <BODY_ORGAN>?" Once all possible question forms are generated, each question form is added to the personality as a new document service. Each document service added instantiates (i.e., creates a specific instance of) the question form with any entities found in the document content 102 or markup 108 that is also in the lexicons identified by the generic symbol in the question form. For example, assuming the document content included the entity "kidney", which was also part of the body organ lexicon. The instantiated question in this instance would be: "What is the procedure for ablation of the kidney?" In one embodiment, these resulting question forms are evaluated for their usefulness.

[0220] When a document is enriched with a personality that includes an instantiated query, the document service request that includes the instantiated query satisfies it with a question answering technique to produce an answer or result. The answer in the example above would be "a nephrectomy". An example of a question answering technique is described by Cooper et al. in "A Simple Question Answering System," published in proceedings of the Ninth Text REtrieval Conference (TREC-9) held in Gaithersburg, Maryland, November 13-16, 2000, which is incorporated herein by reference.

[0221] Once the document service satisfies an instantiated query with an answer, the document service enriches the document by linking the entity in the document with the instantiated query and the answer. In the example given above, the entity "kidney" is linked to the instantiated query (i.e., What is the procedure for ablation of the kidney?) and the answer (i.e., a nephrectomy). In one embodiment, the instantiated query and the answer are displayed in a pop-up window 1028 as shown in Figure 10 when a user locates a pointer 1030 in the vicinity of a recognized entity 1032 (e.g., recognized entity Xerox and the instantiated query of "What is <COMPANY NAME>'s stock price?").

[0222] In the event multiple generic symbols can be added to a question specified by a user, the user may be given the option that only a document service request be specified for only the most generic question form (e.g., What is the

procedure for <SURGICAL_PROCEDURE> of the <BODY_ORGAN>?). Alternatively, the user may be given the option that document service requests be specified for all or selected ones of the identified question forms. In addition, the user may be given multiple answers and multiple information sources to select from.

[0223] Figure 24 is a flow diagram that depicts one embodiment for identifying an answer of an instantiated question. Initially at 2402, the meta-document server 200 receives the instantiated question. The type of question is determined at 2404 and converted to a query at 2406. At 2408, the query is submitted to an information service adapted to handle questions of the type identified. At 2410, passages of the top N results of the query are extracted using for example a summarizer. At 2412, the passages of the extracted top N results of the query are assigned part of speech tags and shallow parsed. At 2414, weights of relevance are calculated for each word in the passages of the extracted top N results of the query using the substantiated question and the determined question type. At 2416, sentences or part of sentences of the extracted passages with words having highest computed weight of relevance are selected as proposed answers to the instantiated question.

D.6 Using Learning Personalities

[0224] The meta-document server 200 provides an e-learning personality that may for example be available in the personality window 1014 in Figure 10. When an e-learning personality is applied to a document, each service in the personality analyzes the contents of the document, recognizing entities and concepts and combinations specific to that service. Each service then links these entities, concepts, or combinations to new content found by a possibly web-based database search, or prepares the search and inserts a link, that when activated, performs the search. Personality services are not limited to simple search, but can perform any actions depending on the content analyzed.

[0225] Figure 25 illustrates a list of services 2502 available when an e-learning personality is selected to enrich document content. E-learning service 2504 and 2506 link words or multi-word expressions found in the document to their definitions and/or translations, respectively. This service may perform lemmatization or stemming before accessing a dictionary. In addition, this service may use the context of the words or multi-word expressions surrounding an element in the content to limit the number of definitions and/or translations displayed. Another e-learning service 2508 links each text unit (i.e., document, paragraph, phrase, word) to a tutorial concerning that element. Yet another e-learning service 2510 links each text unit to a tutorial concerning the text unit. Yet further e-learning services 2512, 2514, and 2516 link each text unit to interactive courses, available online courses, or online resources concerning the subject of the text units, respectively.

[0226] Advantageously, personalities prepare and perform a multiplicity of independent language learning tasks on a specified document(s). When the personality is applied to the document content, each selected service in the personality analyses the contents of the specified document(s), recognizing entities and concepts and combinations specific to that service. The service then links these entities, concepts, or combinations to new content found by a possibly web-based database search, or prepares the search and inserts a link, that when activated, performs the search.

[0227] In one variation, the e-learning personality may also include a service that tracks the user's past action (or access a user profile) to provide new information when the same entity is linked to other documents. In one specific embodiment the e-learning personality is specifically directed at learning languages. In this embodiment, the meta-document server 200 provides computer assisted language learning through using the herein-described document enrichment mechanisms. Figure 26 illustrates an example list of services 2602 available when a language learning personality is selected to enrich document content.

[0228] More specifically, the language learning personality is defined using a personality that performs two or more of the services defined in Figure 26, which include: (a) service 2604 and 2606 that link words or multiword expressions found in the document to their definitions and/or translations, respectively (possibly performing lemmatization or stemming before accessing the dictionary and possibly using the context of the element to limit the number of definitions displayed); (b) service 2608 that links each sentence, or phrase, to a grammatical description of the structure of the sentence or phrase (possibly linking to a textual explanation of the structure in the reader's native language, or to a textual, audio or video grammar lesson corresponding to that structure); (c) service 2610 linking each word, multiword expression, phrase or sentence to other instances of the same in different contexts from the present (e.g., by retrieving similar but differing text segments possessing the same word, multiword expression, phrase or sentence; the retrieved elements could be presented, for example, in a format that brings the similar structure to the center of the field of vision of the user for easy comparison of the differing context); (d) service 2612 that links each word, multiword expression, phrase or sentence to a one or more interactive grammar exercises concerning that element; and (e) services 2614 and 2616 that link to content specific language teaching resource that corresponds to the document content. A similar approach can be followed for other topics of learning.

E. User Controlled Enrichment

[0229] This section describes additional properties that can be specified for personalities and services. Deciding what to enrich and how to enrich content can vary depending on the personality and/or service specified. In one form, a personality annotates any phrase or word identified in its associated list of lexicons (e.g., sports figures), pattern matching using POS tagging, and/or regular expressions (e.g., proper names, noun phrases), or some linguistic processing variant of the two. In another form, a personality provides a global document service that annotates an entire document with for example citations and related documents. This section describes different techniques for providing users with more control over what and how personalities annotate content in a meta-document (e.g., footnotes).

E.1 Automatically Inserting and/or Linking Content

[0230] Figure 12 illustrates at 1220 a mechanism for selectively specifying at a personality level whether to insert enrichment as links 1222, or content 1224, or automatically determine whether to link or insert content at 1226. In either case, links are drawn from entities recognized in document content 102 to either content or services located at a remote location (in the case of 1222) or content located in document markup 108 of a meta-document.

[0231] In an alternate embodiment shown in Figure 27, the user is given the ability to selectively specify personality and/or service behaviors to recognized entities in specified content or documents. In this embodiment, a user for example can select a portion of the enriched document 1018 shown in Figure 10 and select for example the stock quote global service results 1026. This series of actions using known pointer selection techniques causes the display of stock quote options window 2710 shown in Figure 27.

[0232] In the options window 2710, a user may specify that a particular service behavior be applied to all selected documents, a currently selected document, or a selection at 2712. In addition, the options window 2710 permits a user to statically or dynamically update linked information at 2714 that is inserted in a specified form at 2716. For example, information may be inserted as links or content as described above. Content that is inserted can be inserted as for example footnotes or as a list of content at the end of a document. Content that is accessed dynamically is recalculated each time a link or content is accessed (e.g., using Microsoft OLE-like techniques). Content that is accessed statically is done so at a frequency specified at 2718 (e.g., monthly, daily, hourly, etc.).

[0233] Advantageously, a user is given the ability to modify a default behavior of a service while specifying whether changes apply to all documents the user controls, the current document only, or the current selection of a document that contains one or more recognized entities. Depending on the level of change, they are either stored as properties of a particular meta-document or as part of a user's profile.

[0234] Whether to link or retrieve and insert content in a meta-document may be specified for each personality or it may be performed automatically if specified at 2724 in Figure 27 or at 1226 in Figure 12. Determining whether to link or insert content automatically is performed using information from a user's past history of interaction with the meta-document server 200. If specified to automatically link or insert content to a specific personality at 2724 or as a property of a personality at 1226, then the decision whether to insert information as links or content will depend on whether the information is inside or outside a user's interaction history. If outside a user's interaction history, then links are inserted; otherwise, if inside the user's interaction history, the content is retrieved and inserted into a meta-document.

[0235] A user's interaction history can be specified using a history of links accessed by the user and/or a list of interesting concepts to the user. A list of interesting concepts to the user can be determined using for example frequently followed links or from a user profile developed by recording email history or using a recommender system such as Knowledge Pump developed by Xerox Corporation. In this mode of operation, information from a user's interaction history from entity browsing patterns is used to determine whether to enrich document content.

[0236] In yet another embodiment, an annotation property can be set for a specific service as shown or more generally for a personality. In Figure 14, each service has a defined entity type 1412 with an annotate property 1414. The annotate property operates in one embodiment as defined in window 2800 shown in Figure 28 that is made available when selecting a specific annotate property for a service. In one mode of operation 2802, any identified entity is annotated according to an annotation that is predefined for a particular entity type.

[0237] In two other modes of operation 2804 and 2806, a filter function is applied to a list of words. The filter function determines whether to annotate an entity based on predefined filtering criteria such as the frequency the word is used in a reference document (e.g., a document identified to be linked to an entity) or the usage of the entity in the reference document as compared to the document content in which the entity was identified (e.g., using POS tagging).

[0238] In the "expert" mode of operation 2804, only those entities that occur in referenced document(s) or database (s) 2805 with a frequency below a predefined threshold are annotated. In the "novice" mode of operation 1206, only those entities that are identified in referenced document(s) or database(s) 2807 with a frequency above a predefined threshold are annotated. Alternatively or in conjunction with these modes of operation, an entity with few dictionary

senses, or synonyms (e.g., as determined from an online thesaurus) might be discerned as a domain specific entity and therefore either annotated or not annotated. In one embodiment, categories in services are used to form a vocabulary to evaluate dictionary sense.

[0239] A variation of this embodiment allows a user to specify frequency of occurrence at 2801 and 2803 and the reference document(s) and/or database(s) 2805 and 2807 at 2810 (i.e., referenced corpus). For example, in one embodiment this variation would provide when in expert mode, if the frequency of an entity identified in a document is less than a first predefined threshold and the frequency of the entity in identified corpora is less than a second predefined threshold, then the entity in document content of a meta-document is enriched.

[0240] It will be appreciated that the subject of a referenced corpus may relate to a specific subject or a plurality of subjects. Also in this embodiment, the user is also given the ability to specify at 2810 in Figure 28, whether to limit the annotation of words in the document content 102 and/or document markup 108 to only those words that appear once or more than once in the document. This provides that only terms appearing in the document content 102 more than a certain number of times will be annotated as specified at 2811.

[0241] In operation, when a particular document service request 106 is invoked by the meta-document server 200, entities are searched in reference document(s) and/or database(s) and/or document content 102 and/or document markup 108 for their frequency of occurrence. If outside the range of the predefined threshold values, then the entity identified in the document content is not annotated, thereby advantageously limiting document markup in a user specifiable and intelligible manner.

E.2 Propagating Enrichment Between Documents

[0242] Enrichment of a document or meta-document can also be controlled by automatically propagating markup there between as each document or meta-document is accessed by a user. This information can be used as a first pass to enrich documents in real-time while at the same time provide enrichment that may be contextually related to a user's current work in process. This enrichment can be distinguished from other document enrichment using formatting such as font color or the like. In addition, since this enrichment can be tagged for later identification, it can be easily removed from or reinserted into a particular meta-document similar to a track changes function in a text document.

[0243] In one embodiment, enrichment is propagated between meta-documents in the meta-document server 200 as shown in Figure 2. The propagation of enrichment between documents is a user settable property that can be selected in personalities window 1210 at 1230 shown in Figure 12. In operation, if enrichment is selected to be propagated between meta-documents, then entities identified by the meta-document server during enrichment are associated with their annotations and stored together in an entities propagation list. When a new meta-document is enriched by the meta-document server, it first searches through the document content looking for entities that are identified in the entities propagation list. If found, the similar entity is annotated as stored and defined in the entities propagation list. Subsequently, the document service continues with other enrichment functions associated with the service as described above.

[0244] In an alternate embodiment for automatically propagating enrichment between documents, functionality for propagating enrichment can be included in a plug-in to any browser and need not be integrally coupled to the meta-document server 200 as shown in Figure 2. The plug-in in this instance would propagate markup (e.g., hyperlinks) seen on each document during a current session between fetched content (e.g., web pages and/or documents). The markup could be recorded from a predetermined number (i.e., one or more) of previously fetched (or browsed) documents or by session in a markup propagation list that associates strings in fetched content with their markup.

[0245] For example, a plug-in to browsers such as Netscape or Internet Explore can be added that marks up document content as a user browses from one document to the next. That is, every page that is viewed on the browser during a current session (e.g., starting from a first identified document) is analyzed and all strings that are marked up (e.g., everything between the HTML <a> and) are stored by a plug-in with an expanded URL (e.g., base URL plus relative URL found in the 'href' field within the <a> tag).

[0246] Thus, when Xerox is identified in the browser window, then the plug-in would store "Xerox" http://www.xerox.com in its list of seen markups (i.e., the markup propagation list). When the user moves onto a new page, any text segment found in the list of "seen markups" would inherit the annotations found there. For example, if a newly browsed web page contained the string "Xerox" then this string would then be identified by the plug-in in the markup propagation list and annotated with its associated link http://www.xerox.com in addition to any other links it might already possess.

[0247] Figure 29 presents a flow diagram that sets forth the steps for propagating enrichment between electronic documents of different embodiments. In the first embodiment, acts at 2902, 2904, and 2912 are performed. In a second embodiment, acts 2902, 2904, and 2905 are performed. In a third embodiment, a combination of the acts performed in the first and the second embodiments are performed.

[0248] A system for performing propagating enrichment in accordance with acts in Figure 29 waits for a next docu-

ment to be accessed by a user at 2902. Once accessed at 2902, the document is referred to as "the accessed document". The accessing of a document includes any task conceivably performed on a document by the user of a computer. For example, a user has accessed a document when it is displayed (i.e., viewed), printed, emailed, stored, edited, recommended, deleted, processed, had a personality attached to it, etc.

[0249] At 2904, the accessed document is enriched with entities in an interaction history. The creation of the interaction history is described below. The interaction history associates each entity therein with information that identifies a link identifying a location of a document for which the entry was created. This may take the form of a link to which it refers, or if no markup, then the document from which it originates. In addition, the interaction history, includes: the purpose for which the document was accessed (e.g., print, store, email, etc.), the time at which the document was accessed, the POS of the entity, the entity type (e.g., personality entity type), and the number of times and the last time the document was visited.

[0250] The act of enrichment performed at 2904, defines a document with propagated enrichment. Propagated enrichment can be represented in a form to be distinguished from other content in the document, such as using a different font color, format, highlighting, redlining or the like. In one embodiment, all information concerning any entity in the interaction history that is identified in the accessed document is enriched with the contents of the interaction history. This list can be displayed in chronological order or in a ranked order according to relevancy (e.g., determined using a Cosine similarity calculation using entity and surrounding context). Alternatively, that which is displayed to the user is filtered to identify those most relevant if more than one entry in the interaction history is provided for any one entity. Relevant results can be determined by measuring utility of an entry as described in section D.3.2 above for entities. Others entries can be accessed by the user by requesting additional enrichment results. In another embodiment, a similar result is achieved by displaying only those results that are filtered.

[0251] The enrichment information can be filtered using any number of known techniques. For example in one embodiment, enrichment information is filtered with respect to a domain specific corpus using Zipf's Law. Zipf's law, which is known in the art, concerns the distribution of different words in a text object and states that the product of a feature's frequency (where Zipf's law is generalized from words to text features other than and including words) in a text object, **f**, and its rank, **r**, is a constant, **c** (i.e., $f*r = c$). Bearing in mind this law, words having a low frequency will not be that interesting to the reader. In addition, words that have a high frequency will not be of interest either. Consequently, enrichment information is reduced by eliminating information that occurs frequently or very rarely.

[0252] The organization for a user of enrichment information when more than one entry exists in the interaction history for the same entity can be performed using one or a combination of more than one of the following ranking heuristics (i.e., ranking techniques): (a) order information by the most recent time that the document containing the entity was last accessed; (b) order information based on the size of the document to which the markup refers; (c) order information based on whether the document to which the entity refers is an authority or a reference (i.e., hub) as described by Kleinber in "Authoritative Sources In A Hyperlinked Environment," IBM Technical Report RJ 10076, May 1997; (d) order information using a similarity metric to identify the document to which the markup refers and the accessed document; and (e) a ranked list based on actions to the document to which the markup refers. In one embodiment, an equal weight is assigned to each action. Alternatively, higher weights are assigned to certain actions that are deemed important (e.g., printing or recommending). In addition, the ranked information can be displayed in the context of original content. For example, the ranked information can include the closest one hundred words surrounding the ranked information.

[0253] If the system performing propagating enrichment is communicatively coupled to an enrichment system such as the meta-document server 200, acts 2905 are performed. At 2906, a determination is made as to whether a personality is attached to the document to which enrichment is being propagated. If a personality is attached then acts 2908 and 2910 are performed; otherwise, act 2912 is performed.

[0254] At 2908, the document accessed at 2902 is enriched using the attached personality to define a document with personality enrichment. The personality can be specified to be active for only a specified period of time, after which services of the personality that are persistent cease to operate. A document with personality enrichment is not displayed to the user in this instance; its creation is for the purpose of identifying additional markup to propagate to subsequent markup of documents. Entities from the document with personality enrichment are then used to supplement the interaction history at 2910. If no personality is identified, then entities from the document accessed at 2902 are used to supplement the interaction history at 2910. Upon completing either acts 2910 or 2912, act 2902 is repeated. Documents used to markup the document as determined by the meta-document server can also be examined for markup to update the interaction history. Markup coming from the meta-document server may receive a low weight that is used during ranking.

[0255] Figure 30 sets forth a flow diagram with acts for creating and updating an interaction history that are performed at 2912. It will be appreciated by those skilled in the art that the acts performed at 2912 are similar if not identical to those performed at 2910 except that the document with personality enrichment is used to supplement the interaction history and not the accessed document.

[0256] Initially at 3002, a set of rules for identifying entities in the accessed document is defined. The set of rules in one embodiment identify all of the entities in the accessed document that have links or hyperlinks associated therewith. In another embodiment, the set of rules is used to identify entities with specific part of speech tags such as a noun phrase. In yet another embodiment, the set of rules is defined using a personality of the meta-document server 200. In yet a further embodiment, the set of rules used to identify entities that occur within the accessed document with a predefined frequency.

[0257] Using the set of rules defined at 3002, entities in the accessed document that satisfy the set of rules are identified at 3004. At 3006, a determination is made whether to filter the entities identified at 3004. If a determination is made to filter the entities identified at 3004, then those entities are filtered at 3008. Filtering at 3008 involves identifying the overall frequency of entities in the accessed document. Those entities with the lowest frequency pass through the filter. Such a filter assumes that words that occur less frequently are harder to identify (and are therefore more important) than those that do occur more frequently. Finally at 3010, either those entities that exist after acts 3006 (if no filtering is performed) or 3008 (if filtering is performed) are used to update the interaction history for subsequent use at 2904 (shown in Figure 29) to enrich documents accessed by the user at 2902.

[0258] The determination to filter entities at 3006 can be made for example using a maximum threshold number to limit redundant, superfluous, or surplus information. In one embodiment, when an entity has been marked up previously in a document which a user has recently accessed using for example a predefined window of time or usage, or a combination thereof, then the entity should be marked up regardless of any filter.

[0259] Exceptions to filters at 3004 prevent removal of information specified by the system and/or user to be the most significant to propagate between documents. While the filtering performed at 3008 is to avoid excessive markup from being propagated between documents, checks should be put in place to ensure that certain of the markup which is deemed most important is almost always propagated.

[0260] Figure 31 presents a flow diagram for identifying what entities to markup at 3008. At 3150 actions are performed to generate a set of query strings. Initially at 3152, words in the accessed document are tokenized. At 3154, stop words are eliminated from the tokenized words at 3152. At 3156, one of three methods is selected before continuing using the words remaining ("the remaining words") after eliminating stop words at 3154. In a first method at 3158 and 3160, entities are identified in the remaining words and query strings are generated using the identified entities. In a second method at 3162 and 3164, part of speech tagging and shallow parsing is performed on the remaining words to identify chunks and query strings are generated using the identified chunks. In a third method at 3166, n-gram or n-word (e.g., when the number of words is greater than or equal to one) query strings are generated using the remaining words.

[0261] Once the query strings are generated at 3150, they are optionally stemmed at 3168. At 3170, a query is formulated using the query strings (as stemmed at 3168) and context information (e.g., from POS and/or DMOZ categorization information). The resulting query may, for example, be of the form of an SQL type query. At 3172, the query is performed on the interaction history. Finally at 3174, the results are ranked using the ranking techniques set forth herein. A fixed number or a percentage of the highest ranked entities identified are used to markup the identified entity at 3010. Should the query yield no relevant results then the entity is left unchanged.

[0262] Figure 32 illustrates an example in which enrichment is propagated between accessed documents. In the example, browser window 3210 illustrates two entities 3212 and 3214 that have been identified to update interaction history 3218 at 3220 and at 3225 (e.g., act 2912). In this example shown in Figure 32, entity 3212 is recorded in the interaction history 3218 at 3220 to refer to the markup associated with it (i.e., hyperlink www.xerox.com) and not the document from which it originates (i.e., hyperlink www.nytimes.com), unlike the entity 3214 which is not associated with markup and is therefore linked in the interaction history 3218 at 3220 to the document to which it originates (i.e., hyperlink www.nytimes.com).

[0263] The interaction history 3218 which is shown in detail in Figure 33 records entries in a database that include: an entity column containing the particular entity being referenced; a location column that identifies the location of the entity in the identified document; a POS column for storing part of speech information to help determine what context the markup was applied in the accessed document and what context the entity should be applied when propagating it to newly accessed documents; a text category column that contains for example a DMOZ category categorizing the markup; and an entity type column for use in providing additional context information and for its use in propagating it to newly accessed documents.

[0264] In addition the interaction history 3218 includes: a markup column that contains the location of the document that entity refers to (e.g., entity Xerox 3212) or if no markup to the document itself (e.g., entity Xerox 3214); a visits column that identifies how many times the user has accessed the document; a last visit column that identifies the date (and time) the document was last visited; and an action list that identifies an ordered list of actions that have been performed on the document referenced by the link as well as the number of times each action was performed (these statistics may be used for ranking purposes when marking these entities up).

[0265] Referring again to Figure 32 that also shows an entity 3232 (e.g., Xerox) to which enrichment markup 3234

from the interaction history 3218 at 3228 has been propagated to a document accessed using a browser window 3230. The enrichment markup 3234 is displayed, for example, after selecting the entity 3232 by any conventional means such as pointer 3236. Each entry in the propagation markup 3234 can be expanded to show detailed information concerning the entity (e.g., www.xerox.com). Additional entities that are not shown in the propagation markup 3234 can be shown in an expanded selection (not shown) by selecting, for example, button 3238.

[0266] In an alternate embodiment, the interaction history can be used to create a personality as defined above in section D.3. Alternatively, propagating enrichment between documents can be performed by a service of the meta-document server.

[0267] In yet another embodiment, documents accessed can be used to create an information space as set forth in section G.2 below. The information space can then be used to create the interaction history.

E.3 Automatically Completing Citations

[0268] Meta-document enrichment can also be controlled by automatically completing citations set forth in document content 102. In one embodiment, personalities are defined with a property that allows for citations to be automatically included as shown in Figure 12 at 1240. If a personality is specified to have citations automatically completed (i.e., citation mode) then a document service request 106 is added as part of the personality that references an automatic citation document service in services database 210.

[0269] As a result entering citation mode at 1240, the meta-document server will automatically generates a set of citation parentheses ([]) and move a cursor to the middle of the generated parentheses when a user edits meta-document content and types a left square bracket "[" (followed by a capital letter, and followed optionally by a string of letters). In another embodiment, the user could enter citation mode through a menu system of a meta-document editor (not shown).

[0270] In one embodiment, the automatic citation document service assists with the generation of citations and bibliographic entries (i.e., references) and/or cross-references. Content that is referenced or cross-referenced in document content 102 may be included as part of the document markup 108 or linked through document markup 108. This embodiment may operate in either edit mode or non-edit mode as described below to develop entries in a bibliographic database used for automatically generating citations.

[0271] In edit mode and once the system is in citation mode, the user can type the initial letter of the author's name, whom the user wishes to cite, or press a designated key/button, such as the tab key, to indicate to the system to auto-complete the citation from a bibliographic database that meta-document server constructs as a user inputs and/or edits a meta-document. Auto-completion may involve the user selecting from a collection of possible citations or verifying that the proposed citation is actually the citation required by the user. In addition, the auto-completer inserts a bibliographic entry as part of the document markup of a meta-document. The location and style of the entry in the document content is determined from the personality of the document. For example, if the personality is scientific, then the entry will be placed at the end of the document in a standard scientific citation style.

[0272] In non-edit mode, entries for citations in the bibliographic database are identified by: (a) scanning meta-document content and markup of a user for bibliographic entries using known techniques such as Hidden Markov Models; and (b) scanning document categorized by the text categorizer described in section F.1 that is categorized as citable material. Citable material can include any document that contains a title, a list of authors, a date and/or place of publication. The documents cited in the bibliographic database are searched for on the Internet or other fee-for-service content providers. Once content of a citation is located, it is referenced in the bibliographic database using a link or by inserting the content directly into the database.

[0273] Subsequently, after retrieving document content used to build the entries in the bibliographic database in non-edit mode, entries in the database for this content are identified and the content fetched as set forth above, leading to a new collection of bibliographic entries that are added to the bibliographic database used for auto-filling of citations. This process can be repeated to a pre-specified depth, which depth may be defined as part of a personality as shown at 1242 in Figure 12.

E.4 Combining And/Or Intersecting Entities

[0274] As set forth above, the meta-document server 200 recognizes entities in document content 102 as contiguous strings and annotates them either by linking to content that has or has not been retrieved. This feature of the meta-document server 200 enables annotation of combinations of entities within a meta-document.

[0275] In one embodiment shown in Figure 14, the watch business personality includes two services that combine or intersect a set of entities (i.e., an entity type) at 1420 and 1422, respectively. The service at 1420 uses entities of the entity type products 1424 to identify pairs of entities from an identified entity type. The service at 1422 uses entities of the entity type "companies" 1426 to identify pairs of entities and make available 3rd party entities at for example

1060 in Figure 10.

[0276] Figure 34 illustrates the manner in which the two services 1420 and 1422 are performed. Initially, document content 3410 is processed at 3412 using identified entity type 3411 to recognize corresponding entities using known techniques (e.g., using a list, regular expressions, etc.). Without specifying entity pairs at 3420, entities 3414 identified in the document content 3410 of type 3411 are processed by a specified document service at 3416 that, for example, locates documents 3418 with the identified entities 3414. The document service results at 3418 are then used to annotate the entities 3414 in document content 3410 at 3436.

[0277] Unlike typical services 3416, the service at 1420 (shown in Figure 14) specifies entity pairs 3422 using identified entities 3414 at 3420. Specifically, at 3420 an entity pair (I,J) is created for each entity (I) identified at 3414, which is not equal to any other identified entity (J) at 3414. Subsequently, a specified service, for example, locating documents 3428 is performed using the entity pairs 3422. In alternate embodiments, the entities 3414 are specified in other arrangements than pairs 3422 at 3420.

[0278] For example, for any service applicable to the entities 3414 that permits a conjunction of terms (e.g., the information retrieval system Altavista.com allows the prefix "+" to search terms that will be found having at least those terms), a retrieval request is then created that involves all the entity pairs found at 3422. If the entities identified were the proper names "Jack" and "Jill" at 3414 and these formed an entity pair at 3422, a service operating with the entity pair at 3424 may then identify at 3428 documents that refer to the entity pair in some form. The result of the request 3428 is then used to annotate document content 3410 at 3436. In one embodiment, after highlighting an entity from an entity pair in the annotated document content, a pop-up window appears with the entity pair identified with links to the documents 3428 identified as having the entity pair.

[0279] Because computational sources may be limited, the number of pairs at 3422 may need to be reduced. Selection rules 3423 reduce an identified set at 3423 by selecting a combination of entities based on frequency of occurrence: (a) in a corpus of documents (e.g., World Wide Web); and/or (b) in documents accessed (e.g., opened, printed, emailed, etc.).

[0280] A variant of the service 3420 is to specify that entities of the same type are considered if they are predefined in a hierarchical structure, as shown in Figure 35. For example, if the entity type specified with the document service is "people" at 3520 then entity pairs would only be found for each subclass node such as "scientists" at 3522. Another variant of the service 3420 is to pair only those entities in a specified entity type with entities that appear with those entities that also appear within a selection of text (e.g., 500 characters, 10 words, same sentence or paragraph, etc.). Yet another variant of the service 3420 is to pair only those entities that share a syntactic relation (e.g., subject-object).

[0281] Referring now to the service 1422 (shown in Figure 14) for which a service 3424 has been performed with pairs of entities 3422 that may identify documents 3428 with paired entities. These documents are then analyzed at 3430 to identify (i.e., intersect) documents that occur with third party entities (i.e., entities not identified at 3414). For example, if "Jack" and "Jill" are the entity pair, then those documents 3428 identified with "Jack" and "Jill" involving a third party entity (e.g., John, Bob, Simon, etc.) are identified at 3430. These identified third party entities 3432 (which can be viewed by selecting button 1060 shown in Figure 10) and their associated documents 3434 resulting from service 1422 are use to annotate document content 3410 at 3436.

E.5 Using Entity Types Defined In A Hierarchy

[0282] To permit annotation to be applied at different granularities of content, the meta-document enrichment server 200 may organize entity types in a hierarchy 3500 as shown in Figure 35. This permits the specification of document service requests 106 to take place at different levels of the entity type hierarchy 3500. To formulate the hierarchy 3500, entity types are classed in an ontology. The ontology can be formulated using for example the DMOZ ontology (published on the Internet at dmoz.org). The ontology may be global (i.e., available to any user of the system) or local (i.e., available only to a select set of users of the system).

[0283] In one mode of operation, a document service request is applied to selected document content. Associated with that service is an entity type. When an entity in the entity type is recognized in the selected document content, that service activates all services related thereto in the hierarchy 3500 from the node at which it is classified up to the root 3502. That is, all parent nodes of the entity type with the recognized entity are identified, and services associated with each node are applied to the selected document content.

[0284] In this mode of operation, each entity type in the hierarchy has associated therewith a service. For example, if "Michael Jordan" were identified by a "Player Statistics" document service request that referenced the entity type at the level 3508 of the hierarchy 3500, then different services would be invoked for each of the entity types at levels 3506 (e.g., a Team Statistics document service request), 3504 (e.g., a Web Page document service request), and 3520 (e.g., a Vital Statistics document service request).

[0285] In another mode of operation, a user is given the ability to modify each service to set a depth 1430 to which a service identifying an entity type 1412 in the hierarchy 3500 should rise as shown in Figure 14. By right clicking on

a depth 1430, a menu 1432 allows the viewing and editing of the hierarchy 3500 shown in Figure 35. In addition, the menu 1432 allows a user to set the depth 1430 to which different services in the hierarchy should be invoked.

[0286] In an alternate embodiment, the hierarchy 3500 shown in Figure 35 defines all the services that are associated with a node at which an entity type is found in the hierarchy along with the respective depths at which each associated service is found. In this alternate embodiment, a user is given the ability to modify the service applied to each entity type in the hierarchy as well as what entities define each entity type. In yet another embodiment, the ontology is organized using a lattice instead of a hierarchy.

F. Services and Utilities Using Text Categorization

[0287] Services in the services database 210 and utilities such as personality recommender 216 may perform a variety of functions relating to the enrichment of document content that utilize a text categorizer 3610 shown in Figure 36 that forms a utility integrated with or accessed by the meta-document server 200.

F.1 Text Categorizer

[0288] The goal of a text classification system, such as text categorizer 3610, is to classify a document 3612 into a set of one or more classes 3620, which are also referred to as categories. In operation, the text categorizer 3610 assigns a document one or more classes in a set of classes that are defined in an ontology represented in knowledge base 3622. An example of an ontology is the DMOZ ontology (published on the Internet at dmoz.org).

[0289] In addition, the text categorizer 3610 includes a pre-processing module 3614 and an approximate reasoning module 3618. The purpose of the text pre-processing module 3614 is to transform the document 3612 into a representation that facilitates the text categorizer 3610 to perform the task of document classification in an accurate, automatic, efficient and effective manner. Document representations 3624 produced by the pre-processing module 3614 include a set of features and associated weights.

[0290] Different combinations of known techniques from natural language processing such as translation of HTML to text, tokenization, stemming, stop word removal, parsing techniques, and entity recognition can be used to generate the sets of features 3624. Accordingly, the text pre-processing module 3614 may include a number of components such as an HTML to text converter, a tokeniser, a stemmer, a grammar-based feature generator, a feature generator, a word frequency analyzer, and a noun phrase analyzer (or extractor) to produce a set of features 3616 from the document 3612. A commercial application that may include some or all of these functions is Thingfinder™ offered by Inxight Software, Inc.

[0291] The weight value associated with each feature is calculated using any of a number of well known techniques, varying from a normalized frequency count to a more sophisticated weighting scheme which is calculated based upon an aggregation of a number of measures such as the frequency of each term in the document, its location in a document, the frequency of each term in a reference corpus, and the inverse document frequency of the term. The textbook by Manning and Schutze, "Foundations Of Statistical Natural Language Processing", published in 1999, MIT Press, Cambridge, MA, provides a more detailed presentation of text pre-processing performed by module 3614, the contents of which are incorporated herein by reference.

[0292] The approximate reasoning module 3618 processes the categories as represented in terms of rules (or other knowledge forms) stored in the knowledge base 3622, in conjunction with the document representations (e.g., features and associated weights) 3624, to assign a class label 3620 to the input document 3612. In one embodiment, the pre-processing module transforms a document 3612 into lists of tokens that are delimited by spaces, punctuation characters, or the like. Tokens that correspond to stop words (i.e., words that do not improve the quality of the categorization) are subsequently eliminated from this list of tokens. The remaining tokens in the list are then stemmed using Porters stemming algorithm. Then, stop words are removed from the stemmed word list, resulting in a list of terms/words. Finally, this list of terms is transformed to a frequency distribution consisting of **<term, frequency>** tuples where frequency denotes the number of occurrences of that term in the document to define the set of terms (i.e., document representations 3624).

[0293] Subsequently, the approximate reasoning module 3618 accesses a knowledge base 3622 that records variables (i.e., document features and associated frequencies) that are used to define a function that models the mapping from the document 3612, or its transformed representation 3624, to a class in an ontology. One specific embodiment of such a knowledge base is represented using a set of rules that describe relationships between the recorded variables. Typically each class is represented by one rule. In mapping the function, the inference engine 3618 matches the document with each class rule stored in knowledge base 3622 and uses a decision maker for drawing conclusions as to which action to rely on.

[0294] The function as represented by the knowledge base 3622 and approximate reasoning module 3618 can be defined using a variety of model types including the following: probabilistic models; fuzzy set/logic models; Boolean-

valued logic models; nearest neighbor approaches; and neural networks; some of which are described in more detail below. For background relating to some of these algorithms see the following publications by: Shanahan, "Soft Computing For Knowledge Discovery: Introducing Cartesian Granule Features", Kluwer Academic Publishers, Boston (2000); and Mitchell "Machine Learning", Mc Graw-Hill, New York (1997).

[0295] In addition to the elements shown in Figure 36, the categorizer 3610 can include a learning module. The exact make up of the learning module will depend on the model (e.g., probabilistic, fuzzy, etc.) used by the approximate reasoning module 3618 to map a set of documents to the list of categories. Generally, the learning module takes as input classified document examples for each class and generates a corresponding knowledge base.

F.1.1 Probabilistic Model

[0296] In one embodiment, the approximate reasoning module 3618 can use a probabilistic representation. The learning of probabilistic models involves determining the probabilities of various events. These are usually estimated from a labeled training dataset. More formally, a training dataset is a collection of labeled documents consisting of tuples $<D_i, L_i>$ where D_i denotes the document and L_i denotes the label associated with D_i.

[0297] In describing one specific type of probabilistic model, namely, a Naive Bayesian model, first it is described below how to represent models and perform inference approximate reasoning in such a framework, then it is described below how to learn Naive Bayes models from labeled example documents. The naïve Bayes approach to systems modeling has been demonstrated in a variety of fields varying from text classification to disease prediction as disclosed in: Good (1965), "The Estimation Of Probabilities: An Essay On Modern Bayesian Methods" M. I. T. Press; Duda et al. (1973), "Pattern Classification And Scene Analysis", Wiley, New York; and Langley et al. (1992), "An Analysis Of Bayesian Classifiers", in the proceedings of Tenth National Conference on A1, 223-228.

[0298] To simplify the description of the text categorizer 3610, it is assumed that documents 3612 will be assigned to no more than one class. However, it will be appreciated by those skilled in the art that the text categorization method described herein may be readily extended to assign documents to more than one class.

[0299] More formally, the problem of text classification can be represented as a text classification system **S** that assigns a document (or body of text) class labels drawn from a discrete set of possible labels **C**. Mathematically it can be viewed as the mapping: S:Doc $\rightarrow$ {label | label $\in$ C} (i.e., the target function $c = f(\vec{w})$ that models a dependency between a target variable **C** and a set of input features $\mathbf{f_1}, ..., \mathbf{f_n}$). The target variable **C** is discrete, taking values from the finite set $\{\mathbf{c_1}, ..., \mathbf{c_c}\}$. The naïve Bayes classifier accepts as input a document "Doc" and predicts the target value **C**, or a classification, for this tuple. It uses Bayes' theorem in order to perform inference:

$$\Pr(C_i|Doc) = \frac{\Pr(Doc|C_i)\Pr(C_i)}{\Pr(Doc)}.$$

[0300] Consequently, this problem can be represented in terms of class probability distributions Pr(C) and class conditional probability distributions Pr(Doc | C).

[0301] In one specific embodiment, a document Doc is represented in terms of features such as words that occur in the document Doc. Consequently, the above class conditional probability distributions can be rewritten as follows: Pr $(f_1, ..., f_n | C)$.

[0302] Within the naïve Bayesian framework a simplifying assumption is introduced, sometimes known as the naive assumption, where the input variables (in this case the terms) are assumed to be conditionally independent given the target classification value. As a result, the class conditionals reduce to: $\Pr(f_i | C)$.

[0303] Thus, inference (calculation of the posterior probabilities given evidence) using Bayes' theorem simplifies from:

$$\Pr(Class=C_i|<f_1,...,f_n>) = \frac{\Pr(<f_1,...,f_n>|Class=C_i)\Pr(Class=C_i)}{\Pr(<f_1,...,f_n>)}$$

[0304] to the following (and hereinafter referred to as "the simplified inference equation"):

$$\Pr(Class=C_i|<f_1,...f_n>) = \frac{\prod_{j=1}^{n}\Pr(f_j|Class=C_i)\Pr(Class=C_i)}{\Pr(<f_1,...,f_1>)}.$$

[0305] Decision making consists of taking the classification value C_{max} whose corresponding posterior probability is the maximum amongst all posterior probabilities $Pr(C_i | <f_1, ..., f_n>)$ for all values $C_i \in \Omega_c$. This can be mathematically stated as follows:

$$Class(<f_1,...,f_n>) = C_{max} = \underset{C_i \in \Omega_c}{\operatorname{argmax}} Pr(C_i | <f_1,...,f_n>).$$

[0306] Since, in this decision making strategy, the denominator in the simplified inference equation is common to all posterior probabilities, it can be dropped from the inference process. This further simplifies the reasoning process (and the representation also) to the following:

$$Class(<f_1,...,f_n>) = \underset{C_i \in \Omega_c}{\operatorname{argmax}} \left(Pr(Class = C_i) \prod_{j=1}^{n} Pr(f_j | Class = C_i) \right).$$

[0307] As a result of making the naïve assumption, the number of class conditional probabilities that need to be provided reduces from being exponential in the number of variables to being polynomial. This assumption, while unlikely to be true in most problems, generally provides a surprisingly high performance that has been shown to be comparable to other classification systems such as logic systems (decision trees) and neural networks (see Wiley cited above; and Langley et al. (1992), "An Analysis Of Bayesian Classifiers", in the proceedings of Tenth National Conference on A1, 223-228).

[0308] In other words, each class is represented by a series of word conditional probabilities for each word and a class conditional that are used in the calculation of the posterior probability for a class given a new document to be classified.

[0309] Naïve Bayes classifiers can quite easily be learned from example data. The learning algorithm operating in a learning module consists of estimating the class conditional probabilities and the class probabilities from a training dataset **Train** (a labeled collection of documents) for each possible document classification **Class**, where the class conditionals correspond to the following: $Pr(f_i | Class = c_j)$ " $i \in \{1, ..., n\}$ and the class probability distribution corresponds to: $Pr(Class = \mathbf{c_j})$.

[0310] The class probability $\mathbf{Pr(Class = c_j)}$ corresponds to the fraction of documents having the classification of $\mathbf{c_j}$ in the training dataset **Train**.

[0311] Each class conditional $\mathbf{Pr(f_i | Class = c_j)}$ can be estimated using the m-estimate (see Mitchell cited above):

$$Pr(f_i | Class = c_j) = \frac{\sum_{j=1 Doc}^{|c_j|} Freq(f_i, Doc_j) + 1}{\sum_{j=1}^{|c_j|} |Doc_j| + |Vocab|},$$

[0312] where: $\mathbf{Freq(f_i, Doc_j)}$ denotes the number of occurrences of the feature $\mathbf{f_i}$ in the training document $\mathbf{Doc_j}$; $\mathbf{|Vocab|}$ denotes the number of unique features considered as the language of the model (i.e., the number of variables used to solve the problem); and $\mathbf{|Doc_j|}$ denotes the length of the document $\mathbf{Doc_j}$ (i.e., the number of terms, words, or features in the document).

F.1.2 Fuzzy Model

[0313] In another embodiment, the text categorizer 3610 uses a fuzzy model to categorize document 3612. In this embodiment, the pre-processing module 3614 includes a feature extractor 3615, a feature reducer 3617, and a fuzzy set generator 3621 as shown in Figure 36. The feature reducer 3617 is used to eliminate features extracted by the feature extractor 3615 that provide little class discrimination. The fuzzy set generator 3621 generates either fuzzy sets or granule fuzzy sets depending on the fuzzy model used. Associated weights of features generated by the preproc-

essing module 3614 are interpreted as fuzzy set memberships or probabilities.

[0314] More specifically in this embodiment, the approximate reasoning module 3618 computes the degree of similarity (i.e., match) between the unlabelled text object 3612 that is represented in terms of: a feature vector produced by feature extractor 361 5, a document fuzzy set produced by the fuzzy set generator 3621, and one or more categories as specified by the approximate reasoning module 3618. The approximate reasoning module 3618, which contains matching, filtering and decision making mechanisms, accesses the knowledge base 3622 to classify the unlabelled text object 3612.

[0315] In a first embodiment, the knowledge base 3622 contains rules for each class (i.e., category), where each rule is made up of a class fuzzy set and an associated class filter. During operation of this embodiment, the approximate reasoning module 3618: (1) calculates the degree of match between the document fuzzy set 3624 and a fuzzy set associated with each class (i.e., each class fuzzy set); (2) passes the resulting degree of match through a respective filter function (i.e., class filter); and (3) determines a class label to assign to the unlabelled text object based upon the filtered degrees of match (e.g., the class label associated with the highest degree of match is assigned to be the class label of the text object).

[0316] In a second embodiment, each rule is made up of a granule fuzzy set. Similar to the categorizer of the first embodiment that uses fuzzy set models, this categorizer uses granule feature based models. In operation, the categorizer of this second embodiment performs a functional mapping from the set of features to a set of class values. Further details of the a text categorizer that uses fuzzy models is described by Shanahan in U.S. Patent Application Serial No. 09/928,619, entitled "Fuzzy Text Categorizer", which is incorporated herein by reference.

F.1.3 LSI Model

[0317] In yet another embodiment, the text categorizer 3610 uses LSI (Latent Semantic Indexing) to categorize document 3612. Text classification and learning can be performed using LSI and similarity metrics in the resulting feature space. The LSI model is used to translate feature space into latent concepts space that can be used to explain the variance-co-variance structure of a set of features through linear combinations of these features. Subsequently these transformed features can be used as input to any learning algorithm. In addition, LSI classification can be used with K nearest neighbor and a fuzzy classifier. Having identified the latent concepts they can be used for classification (such as fuzzy classifier defined above or K nearest neighbors) or similarity metrics (Cosine metric that can be used for ranking or re-ranking). Additional background relating to the generation of context vectors is disclosed by Deerwester, in "Indexing By Latent Semantic Analysis", Journal of the American Society for Information Science, 41(6): 391-407, 1990.

F.1.4 Vector Space Model

[0318] In yet a further embodiment, the text categorizer 3610 uses a vector space model to categorize document 3612. Under the vector-space model, document and queries can be conceptually viewed as vectors of features, such as words, noun phrases, and other linguistically derived features (e.g., parse tree features). Typically a feature extraction module transforms a document (or query) into its vector of features, $\mathbf{D} = \langle f_1, \ldots, f_n \rangle$, where each f_i denotes the statistical importance (normalized) of that feature. One common way to compute each weight f_i associated for document Doc is as follows:

$$f_i = freq(f_i, Doc) * idf(f_i),$$

[0319] where **freq(f_i, Doc)** represents the frequency of feature f_i, in document **Doc** and **idf(f_i)** represents the inverse document frequency of the feature f_i, in a document collection DC. The **idf(f_i)** factor corresponds to the content discriminating power of i^{th} feature: i.e. a feature that appears rarely in a collection has a high **idf** value. The **idf(f_i)** factor is calculated as follows:

$$f_i = \log\left(\frac{|DC|}{DF(f_i)}\right),$$

[0320] where |DC| denotes the number of documents in the collection **DC** and **DF(f_i)** denotes the number of documents that contain f_i. Typically, a normalized document vector, $D = \langle nf_1, \ldots, nf_n \rangle$ is used in the vector space model of

information retrieval, where each nf_i is obtained as:

$$nf_i = \frac{f_i}{\sqrt{\sum_{j=1}^{n}(f_j)^2}}.$$

[0321] Queries can also be represented as normalized vectors over the feature space, $Q = <q_1, ..., q_n>$, where each entry indicates the importance of the word in the search.

[0322] The similarity between a query **q** and a document **d**, **sim(q, d)**, is defined as the inner product of the query vector **Q** and document vector **D**. This yields similarity values in the zero to one range [0, 1].

[0323] Additional background relating to the generation of vector space models is disclosed in U.S. Patent No. 5,619,709, and by Salton et al., in "A Vector Space Model For Information Retrieval", Journal of the ASIS, 18:11, 613-620, Nov. 1975.

F.2 Recommending Personalities

[0324] The meta-document server 200 provides a service for recommending personalities at 216 in Figure 2. In one instance, personalities are recommended for each document after a user uploads to the meta-document server 200 and the user has selected the personality property 1214 shown in Figure 12. After a user selects the personality property 1214, the personality recommender 216 automatically recommends a personality for each document uploaded by the user. By recommending a personality, the personality recommender 216 aids a user to decide which of a plurality of document enrichment themes are to be applied to an uploaded document by analyzing document content and other contextual information (e.g., actions carried out on the document) of the uploaded document.

[0325] In one embodiment, personalities that are recommended by the personality recommender 216 are automatically attached to the uploaded document without requiring user acknowledgment and these documents are immediately enriched by the meta-document server. Alternatively, the personalities that are recommended by the personality recommender 216 are attached to a meta-document only after the user provides an acknowledgement that the recommended personality is acceptable to the user.

[0326] In order to decide which personality (or personalities) to recommend to attach to a document, the meta-document server 200 uses an uploaded document 3712 as input to the personality recommender system 216, an embodiment 3700 of which is shown in detail in Figure 37. Generally, the personality recommender system 3700 shown in Figure 37 is similar to the document categorizer 3610 shown in Figure 36 except that the personality recommender assigns a list of one or more personalities 3720 instead of a list of one or more categories as specified in section F.1 for the categorizer. The personality recommender 3700 can learn rules for recommending personalities and for developing a personality ontology using documents previously uploaded to the meta-document server 200 and assigned a personality by a user.

[0327] More specifically, the personality recommender system 3700 shown in Figure 37 is a variant of the text categorizer described above in section F.1 and shown in Figure 36. The knowledge base 3722 can be defined manually using data from personality database 212, which may contain user specific personalities or generally available personalities (e.g., using features and weightings chosen manually for each personality that could be applied) and documents that were previously assigned to those personalities in the meta-document database 202.

[0328] Alternatively, the knowledge base can be defined semi-automatically or automatically using features and weightings chosen by machine learning techniques. In the case of automatically learning the features and weightings, the learning module 3730 may use meta-documents existing in the meta-document database 202 to train the knowledge base 3722. Subsequently, the learning module 3730 validates the knowledge base 3722 using user profile database 3708. The user profile database 3708, which includes portions of the meta-document database 202 and the personality database 212, includes references to meta-documents that users have already applied a personality thereto.

[0329] In operation, the pre-processing module 3614 (described above in section F.1) of the personality recommender 3700 extracts features 3616 from an uploaded document 3712. Subsequently, the approximate reasoning module 3618 (described above in section F.1) derives a list of personalities 3720 using knowledge base 3722. These extracted features would then be exploited, again using standard techniques (using for example, Bayesian inference, cosine distance, as described above), to classify the new document and rank the possible list of personalities 3720 to recommend enriching specified document content. Every personality ranking above a certain threshold or just the top N (N>=1) personalities can be recommend by the approximate reasoning module 3618.

[0330] In a variant of the personality recommender 3700, the personalities ranked for a new document are re-ranked using the profile of the user. For example, if the approximate reasoning module 3618 attaches to a document a business

and a sports personality, but the user's own profile in 3708 reveals that this user has never applied a business personality then the ranking can be altered in 3701 so that only the sports personality is proposed, or applied with greater priority, before the business personality. Accordingly, personality recommendations can be tailored for a particular user using the user's interaction history with the meta-document server 200 (e.g., an example interaction history is shown in Figure 33 and described in section E.2).

F.3 Generating Queries Using Identified Entities

[0331] Traditional searches for information are invoked when an information need exists for an identified task. From this information need a query is formulated and a search performed, generally directed by a user. In accordance with searches performed by services of the meta-document server 200, one or more documents relating to a task are identified and uploaded to the meta-document server 200. From these documents queries are generated for specified services automatically (and optionally as specified by a user).

[0332] As set forth above, a document service request in a personality associated with an uploaded document identifies entities that are used to perform other document service requests such as queries. The manner in which to automatically formulate queries given an identified entity and its associated document content is the subject of this section. This technique for automatically formulating a query aims to improve the quality (e.g., in terms of precision recall) of information retrieval systems.

[0333] Figure 38 illustrates the elements and flow of information for generating a query 3812 by query generator 3810. The query generated may include some or all of the following elements as discussed in more detail below: (a) a set of entities 3808 identified by, for example, a document service request 106 performed by entity extractor 3804 or manually by a user, (b) a set of categories 3620 generated by the categorizer 3610 (as described above in further detail while referring to Figure 36), (c) an aspect vector 3822 generated by categorizer 3610 or short run aspect vector generator 3820, and (d) a category vocabulary 3621 generated by the categorizer 3610.

[0334] In operation as shown in Figure 38, the document content 3612 or alternatively limited context (i.e., words, sentences, or paragraphs) surrounding the entity 3808 is analyzed by categorizer 3610 to produce a set of categories 3620. It will be appreciated that although the description is limited to document content it may in also include enriched document content. In addition, the document content 3612 is analyzed by short length aspect vector generator 3820 to formulate a short length aspect vector 3822. In an alternate embodiment, the aspect vector generator 3820 forms part of the categorizer 3610.

[0335] In one embodiment, the query generator 3810 coalesces these four elements (i.e., entity 3808, category 3620, aspect vector 3822, and category vocabulary 3621) to automatically formulate query 3812. Advantageously, the query 3812 may be contextualized at different levels: first, the query is set to be directed in a specific category of an information retrieval system that may, for example, be hierarchically organized; second, the query may be augmented with additional terms defined in aspect vector 3822; third, the query may be further augmented with additional terms related to the category vocabulary 3621. In alternate embodiments described below a query can be contextualized using just one of the category 3620 and the aspect vector 3822.

[0336] After generating the query, in one example embodiment, it is used by the meta-document server 200 to access content provided by networks services 206 (introduced in Figure 2). The content provided as a result of the query can then be used by the content manager 208 to enrich the original document content 3612. In another embodiment, the content is provided to a user as a result of performing a search on a specified entity 3808.

F.3.1 Category Generation

[0337] In generating the set of categories 3620, the categorizer 3610 classifies input document to generate classification labels for the document content 3612. Terms and entities (i.e., typed terms, such as people organizations, locations, etc.) are extracted from the document content. For example, given a classification scheme such as a class hierarchy (e.g., from a DMOZ ontology that is available on the Internet at dmoz.org) in which documents are assigned class labels (or assigned to nodes in a labeled hierarchy), a classification profile is derived that allows document content to be assigned to an existing label or to an existing class, by measuring the similarity between the new document and the known class profiles.

[0338] Document classification labels define the set of categories 3620 output by the categorizer 3610. These classification labels in one embodiment are appended to the query 3812 by query generator 3810 to restrict the scope of the query (i.e., the entity 3808 and the context vector 3822) to folders corresponding to classification labels in a document collection of an information retrieval system. In an alternate embodiment, the classification labels are appended to the terms in the aspect vector to formulate a more precise query. Adding terms in the aspect vector adds constraints to the query that limit the search to a set of nodes and/or sub-nodes in a document categorization structure (e.g., hierarchy, graphs). In yet a further embodiment, the classification labels are used to identify the characteristic vocab-

ulary (i.e., category vocabulary) 3621 associated with the corresponding classes. The terms of the characteristic vocabulary 3621 in this embodiment are appended to the aspect vector to again formulate a more precise query.

[0339] After processing the query by submitting it to an information retrieval system (e.g., Google, Yahoo, Northern-Lights), the query can be refined by filtering and/or ranking the results returned by the query mechanism using the classification labels or its associated characteristic vocabulary in a number of ways. For example, results can be ranked from most relevant to least by matching returned document profiles against the classification labels or the characteristic vocabulary of the predicted class by: using a categorizer; or using a similar metric in the case of the characteristic vocabulary, such as the cosine distance or similarity measure base on an LSI transformation of the original feature space. The results of these more precise queries are used to enrich original document content. In one embodiment, documents are enriched by the meta-document server 200 described above, the operation of which involves automatically executing the query, for example, on the Internet, and retrieving the query results and linking these results to the original terms and entities in document content.

[0340] Figure 39 illustrates an example of a query 3930 contextualized using classification labels 3920 of document categorization hierarchy 3900. Using document content 3902, the categorizer 3610 identifies classification labels 3920. These labels identify nodes 3910, 3912, and 3914 of the top-level node 3904. Specifically in this example, the entities "seven" and "up" are determined by categorizer 3610 to relate most appropriately to the class of documents found in the directory science>biology>genetics. As specified at 3930, the search is focused on documents found in the single node of the document hierarchy genetics, at 3910.

F.3.2 Aspect Vector Generation

[0341] As set forth above, personalities recognize certain entities in a document and search for information concerning them in personality-specific data sources. Aspect vectors add a small amount of context to the entity to restrict a search for information, thereby making the search more precise.

[0342] In operation when an entity is found in document content by a document service request, that entity will be used by another document service request to gather and filter information concerning that entity. Producing an aspect vector contextualizes queries related to the entities by examining a portion of the document content that may range from all of it to one or more paragraphs and/or segments around the entity.

[0343] The aspect vector is produced by analyzing a document's textual content using natural language processing in order to extract different facets of the document. In one embodiment, three facets of document content are examined (i.e., tokens (i.e., words), phrases, and rare words) to identify terms to retain. The retained terms are added to the recognized entity, in order to increase the precision of the query.

[0344] Tokens from the document are identified using words that are normalized using, for example, techniques such as mapping uppercase characters to lower case, stemming, etc. These tokens are divided into two parts: words appearing in a list of stop words (e.g., in, a, the, of, etc.); and all other words. Tokens identified in the list of stop words are discarded and the remaining words are sorted by decreasing frequency to define a sorted list of words. From the sorted list of words, the N (e.g., N=3) most frequent words are retained. In addition, some of these N (e.g., N=2) words are specially marked so that their presence becomes mandatory in documents retrieved by the query.

[0345] Phrases in document content are defined either using a language parser which recognizes phrases, or approximated by some means (e.g., taking all sequences of words between stopwords as a phrase). Only phrases consisting of two or more words are retained. These remaining phrases are sorted by decreasing frequency. The top M (e.g. M=3) most frequent phrases, possibly fulfilling a minimum frequency criteria (e.g. appearing more than once in the entire document), are retained.

[0346] Rare words are defined as those (non-stopwords) appearing with a low frequency in some reference corpus (e.g. The British National Corpus of 100 million English words). All non-stopwords are sorted by their frequency in the reference corpus in ascending order. The top P least frequent words (e.g. P=3), possibly fulfilling a minimum frequency criteria (e.g. appearing more than three times in the entire document), are retained.

[0347] Variants of this method include limiting the number of context words used by a certain number of words or characters, for example, certain information retrieval systems accept queries up to a length of 256 characters in length, while others information retrieval systems accept queries that have a maximum of ten words. Another variant includes using additional lists of ranked items extracted from other facets of the text such as: (a) proper names (e.g., ranked by decreasing frequency), (b) rare phrases (as with rare words, calculating rareness by frequency in a reference corpus, for example, an image of the WWW), (c) dates, (d) numbers, or (e) geographical locations. Advantageously, mixing terms from different facets of the document content to extracted entities improves precision of query related to marking up the entity.

[0348] For example, assume a web page mentions a professor named Michael Jordan. Further assume that the entity identified by the meta-document server 200 is Michael Jordan. Sending the query "Michael Jordan" to an information retrieval system such as AltaVista identifies approximately 1.2 million documents, with the 10 top-ranked doc-

uments about the basketball player Michael Jordan. By augmenting the entities "Michael Jordan" of the query with the aspects such as "computer science", "electrical engineering", and "faculty members" extracted from the document content, a more precise query can be formulated for identifying information relating to a professor named Michael Jordan.

F.3.3 Example

[0349] Figure 40 sets forth a flow diagram which depicts one embodiment in which both categories and aspect vectors can be used to improve the accuracy of an information retrieval system. At 4002, one or more entities are extracted from a document. Entity identification or extraction can be performed: (a) manually by a user, (b) automatically by entity extractor 3804 shown in Figure 38 using for example a method as described in section B.4, or (c) by the categorizer 3610. At 4003, the extracted entity at 4002 is added to a query at 4003.

[0350] At 4004, the document from which the entity is extracted is categorized. Categorization involves producing a category 3620 and a category vocabulary 3621. The category vocabulary for a category consists of one or more terms that describe the category. In one embodiment, the category vocabulary is generated a priori and associated with each category in an ontology. At 4007, for the particular category identified at 4004, a node in the organizational structure of the categories is located.

[0351] At 4008, if the node located node has not been searched with the query, then the query as it is defined is directed to the located node in the category organization at 4010. At 4009, if the root node has not already been searched using the defined query, then the node in the category organization at which the category is defined is changed at 4014 to its parent node. The parent node in a category organization is generally less descriptive than the child node. The root node defines the least descriptive category in the category organization.

[0352] At 4012, if search results are obtained at 4010, then they are evaluated for accuracy at 4016. If no results are obtained at 4012, the node in the category organization at which the category is defined is changed at 4014 to its parent node and act 4008 is repeated. Note that if there is no parent of the located node at 4014, then the node remains unchanged and is by definition the root node.

[0353] At 4016, if the search results are determined to be accurate (e.g., by user approval), then the process terminates at 4030. At this point the results of the query may, for example, be displayed to a user or used to automatically enrich document content.

[0354] At 4018, if the results are not accurate at 4016, then a determination is made whether a short run aspect vector has already been added to the query. If it has not already been added then a short run aspect vector using the document content and the entity as described above in section F.3.2 is generated at 4020. At, 4022 the aspect vector is added to the query and the node to which the query is pointing in the category organization is reset to the node that corresponds to its original categorization at 4024. Subsequently using this augmented query, act 4008 is repeated.

[0355] Furthermore, if the query should need to be further augmented at 4026 with the category vocabulary because of inaccurate results found at 4016, then the category vocabulary is added to the query at 4028, thereby further augmenting the query. The node to which this augmented query is pointing in the category organization is reset to the node that corresponds to its original categorization at 4024 and act 4008 is repeated.

[0356] Figure 41 illustrates a client interface 4110 similar to the client interface 1010 shown in Figure 10. Unlike the client interface 1010, the client interface 4110 displays an augmented query that can be performed using a recognized entity 1032 in a pop-up window 4102. The pop-up window 4102 appears when a user locates the pointer 1030 in the vicinity of the recognized entity 1032. The pop-up window 4102 illustrates one or more category organizations 4104 used in defining a query, as well as, classification aspects 4106 and contextual aspect 4108 that are associated with the query, each of which can be viewed and edited as shown in window 4112. To manually invoke a search based on an entity, the user selects the desired level in the category organization and whether one or more aspects should be used to augment the query.

F.4 Finding An Expert For An Enriched Document

[0357] In order to help a user understand a document, an expert service provides help finding experts for subjects mentioned in a meta-document. In one embodiment, a user selects button 1036 in Figure 10 after a document is uploaded. Once invoked, the expert service uses as input whatever content (e.g., text, hyperlinks, graphics) that is available in the current state of the document (e.g., the user may be composing the document) to find an expert about the subject. Advantageously, a document text segment can be used by the expert utility to generate the query to access a database of experts, and manage the exchange of responses or documents, within the context of a the meta-document system shown in Figure 2.

[0358] In one embodiment, the expert utility operates by performing the following steps: (a) the current state of a meta-document is input to the expert utility; (b) a profile is created for the meta-document (or for a document segment

selected by the user) either by traditional indexing means, or by creating short query context as disclosed in section F.3 above or by categorizing as described in section F.1 above (Note that the profiles can be created for the entire document or for any segment of the document depending on the number of segments of the meta-document selected by a user.); (c) this profile is used to query a known website for experts (e.g., http://www.exp.com) or by finding the most active rater for topics in that profile in some recommendation system such as Knowledge Pump developed by Xerox Corporation; and (d) pointers to and/or content regarding the experts found are referenced and/or brought back as annotation for the document segment selected.

G. Additional Meta-Document Services

[0359] This section sets forth additional services and embodiments that in one embodiment may operate separate from or integral with the meta-document server.

G.1 Notification Of Enrichment

[0360] As set forth above, when a personality 104 is attached to document content 102, the personality consists of many document service requests 106 identifying document services that are periodically initiated by scheduler 204 to examine the document content 102 and the document markup 108. By examining content and markup of a document, a document service may recognize a certain number of entities inside the document. The document service also may link these entities to a multiplicity of data sources on the World Wide Web (i.e., WWW) or fetch the content of the link, as provided in section 1220 of Figure 12. In addition as part of the document service, the service may also filter and/or transform retrieved document content.

[0361] If desired, a user may specify whether to invoke a notification service that will notify a user upon completion of a document service. It will be appreciated by those skilled in the art that document services may be able to be performed in real time and therefore not require notification of its completion to a user. In the event notification is required to perform actions that cannot be performed in real time, as part of the properties 1210 of a personality shown in Figure 12, a user may specify at 1204 whether to be notified by email 1205, voice mail 1206, or SMS (Short Message Service) text messaging over GSM 1207 upon completion of the service. An example of a service that requires significant processing time is a combinatorial search of a list of words.

[0362] When a notification mechanism is selected at 1204, a notification document service request is added to the specified personality to alert the user who applied the personality to the document when significant changes appear on the web or in a local database concerning any of the entities mentioned in the document. The threshold amount of change that invokes a notification service can be predefined by the user and/or system. In addition, the user may be provided with a mechanism (not shown) for specifying a specific entity to be watched for changes.

[0363] Advantageously, the notification system is not based on specifying a URL or a document repository to be watched for changes. Instead this notification system is initiated by specifying a document service request of a meta-document. Consequently, the notification of changes to information involves only that information which the user is concerned about. In addition, this form of notification provides a level of indirection, since the user is alerted about new information concerning entities in a document even if the document content 102 or markup 108 never changes.

[0364] More specifically, this change alerting document service request is packaged in a personality that can be activated in the meta-document server 200 (i.e., attached to a document) by the user. Initially, document service requests analyze a document by linguistically processing the document to recognize entities within the document. These entities can be strings from a list (e.g., list of medicine names), or regular expressions describing a multiplicity of entities (e.g., a proper name recognizer, a chemical formula recognizer, etc.), or elements recognized by linguistic processing (e.g., noun phrases, words in a subject-verb relations, etc.). Entities may also have keys associated with them in another list or database (e.g., Xerox as an entity with stock key XRX).

[0365] Another document service accepts these entities, their associated keys, a procedure for accessing information for each entity, an update period, information about the user requesting the notification and a change significance level (e.g., Any Change, Minor Change, Major Change, etc.) as input. This document service request then performs the information access (e.g., local database access, accessing a content source on the Internet, etc.) for each entity at the beginning of every update period.

[0366] The document service request compares the data retrieved for each entity at the current and at the previous update period (i.e., the data retrieved for each entity in the previous update is stored and accessible to the document service request). If the stored information is significantly different, as described below, from the newly retrieved information the user is notified (e.g., via e-mail or any other notification mechanism) that new entity-specific information is available and the user is also given a description of the change. The document service request decides on significance using a change significance parameter that measures how much the new information differs from the stored information (e.g., by comparing the number of characters, etc.).

[0367] In one embodiment, the change significance parameter has a plurality of settings (e.g., high, low). For example, if the information retrieved for an entity previously was a web page, and the change significance parameter was set high, then the user may be notified only if the length of the web page length changes by more than 30%. If the change significance was set low, then the user would be notified if the page length changes by more than 5%. If the change significance parameter was set to any change, then any change in the page length would cause the user to be alerted. In an alternate embodiment, the change significance parameter is computed by storing any reduced description of the accessed pages (e.g., hash function, significant words, all non stop words, etc.) in the system and comparing the stored representations of the page to newly accessed representations in order to determine change.

G.2 Document-Centric Suggestions

[0368] This section describes a mechanism that uses an information space surrounding a document to provide an improved (e.g., more accurate and more stylish) document-centric auto-completion system and auto-correction system that can be used during content creation. Document auto-completion saves a user from having to retype text (and other document content such as graphics) and related markup such hyperlinks, bibliographic entries etc., by providing suggestions of words that have been used previously in a contextually similar manner. Document auto-correction provides a textual correction system that dynamically updates the information space as corrections are made or accepted.

[0369] The meta-document server 200 described above is an example of one embodiment that can be used to create an information space surrounding a document, thereby creating a document-centric view of the world. An information space includes document content, document markup, and information relating to additions and/or changes relating to document content (e.g., additions, changes, keystroke order etc.). For example, Figure 42 illustrates an information space 4200 that surrounds meta-document 4202. The meta-document 4202 includes content and markup. The markup enriches content of the information space of the meta-document 4202, for example, by linking identified entity 4204 in the meta-document content 4203 to a set of meta-documents 4208.

[0370] In addition, the markup of meta-document 4202 grows the information space 4200 on a document level (as opposed to an entity level) at 4216 using similar documents 4206. The similar documents 4206 links to a set of meta-documents 4210 that relate to the content 4203 as a whole and not to any single entity of the content 4203. Also, the document level markup of the information space 4200 includes a reference 4212 to an entity database 4214 of extracted entities, an example of which is shown in Figure 48 and discussed in more detail below.

[0371] It will be appreciated by those skilled in the art that the elements making up the meta-document information space 4200 (e.g., document content 4203, the sets of meta-documents 4208 and 4210, and the entity database 4214) need not be collocated together in a single space and/or machine. Instead, the elements making up the meta-document information space 4200 may be located physically distant from each other on different computer systems and/or file storage systems that operate independently across the network 221 shown in Figure 2.

[0372] The construction of the information space 4200 surrounding a document can begin at document creation time by, for example, creating a document on the meta-document server 200. Once the information space surrounding a document is created, the user or the system can exploit it during the knowledge management cycle. The system in case of an auto-completion service uses the information space of a particular meta-document(s) to aid in creating suggestions for completing input for a user.

[0373] Auto-completion involves the process of automatically completing one or more words without manually typing all the characters that makeup that word(s). In one embodiment, the user types the first few characters of a word, presses a special request key to invoke completion, and the rest of the word is filled in. The completed word may also be rejected with the aid of another special key. If multiple alternatives exist, the user may be prompted to select one from a displayed list of alternatives or to reject the proposed completions.

[0374] In one embodiment illustrated in Figure 43, an auto-completion module 4302 operates with a text editor 4314 and the meta-document information space 4200. The auto-completion module 4302 provides document-centric suggestions to entity fragments (e.g., string fragments) added to document content 4203 using the text editor 4314. With the aid of the entity database 4214 in the information space 4200 suggestions for expanding the entity fragments are defined. As illustrated in Figure 43 the auto-completion-module 4302 includes a tracking module 4304, a query formulation module 4306, an information retrieval system 4308, a suggestion module 4310, and an insertion module 4312.

[0375] In the embodiment shown in Figure 43, the entity database 4214 in information space 4200 stores one or more text objects (i.e., a word or collection of words that may take the form of a string) that could be used to auto-complete users textual input at editor 4314 destined to form part of document content 4203. Exactly what text objects define the entity database 4214 depends on the content 4203 and the personality used to define the information space surrounding the content.

[0376] The tracking module 4304 interacts with the text editor 4314. An example of a text editor is the Microsoft® Word editor. The tracking module 4304 monitors a user's input for auto-completion requests (e.g., via designated keystrokes) or for partially input words (e.g., characters string of 2 or more characters). In one embodiment, the tracking

module 4304 is integral with the text editor 4314. In another embodiment, the tracking module 4304 operates independent from the text editor 4314, for example, as an optional plug-in to text editor 4314.

[0377] The query formulation module 4306 translates an auto-completion request received from the tracking module 4304 into a query that is passed onto the information retrieval module 4308. The information retrieval module 4308 accepts the query derived from the auto-completion request and searches the entity database 4214 for possible auto-completions that would be best used to auto-complete (i.e., match) the string currently input by the user.

[0378] The suggestion module 4310 either selects the most appropriate string match (i.e., high confidence completion) or presents a list that is ranked or otherwise ordered in a predefined form of the most appropriate alternative completions to the user of the text editor 4314. The user subsequently selects one or none of these alternative strings. If one of the alternative strings is selected, the insertion module 4312 takes the selected string for auto-completion and auto-completes the current input string by inserting the remaining characters of the selected string after the string fragment.

[0379] Although the example discussed herein is limited to query construction in a text auto-completion context, it will be appreciated by those skilled in the art that a similar analysis can be used for other types of objects that need be auto-completed. For example in alternate embodiments, entity fragments in the auto-completion system may include other objects for completion besides text objects, such as multimedia type objects. Multimedia type objects include any input sequence (e.g., from an input device such as a keyboard, mouse, interaction device such as a gesture recognition system, etc.), graphics object, sound object, and images object.

[0380] As such, the database of auto-completions 4302 is no longer just a list of text strings but list of tuples consisting of an access key (e.g., entity fragment), and of an object such as a string of words, a graphics object, and/or an input sequence, that is used to auto-complete a user's input. This object may have associated with it various attribute descriptions that make up other fields in the database tuple. For example, an auto-completion system for graphics would suggest the completion of a fourth side of a square once three or even two sides have been drawn.

[0381] Unlike traditional auto-completion systems, which typically use a static database of entities to auto-complete user's input and provide facilities for the user to add one's own auto-completion entities, the auto-completion system 4300 dynamically builds an auto-completion database of entities (text or otherwise) from the information space that can be created around a document using information space creation systems such as the meta-document server 200. Which entities are extracted, how they are extracted, and indexed in the auto-completion database is determined by the personality associated with the document. For example, bibliographic entries may only be important for scientific personalities.

[0382] Figure 44 illustrates an alternate embodiment in which the auto-completion module 4302 operates integrally with elements of the meta-document server 200 described above and shown in Figure 2. This embodiment includes a document initialization module 4404 in the user manager 214 for initializing a meta-document with a name, a personality and other meta-values (e.g., access privileges etc.). For example, user operating computer 226 inputs and/or edits document content using text editor 4314 that forms part of the content manager 208 (or alternatively, part of computer 224).

[0383] While receiving input and/or edits to document content, the meta-document server 200 anticipates the information needs of the user creating and/or editing the document content by creating an information space around the document content that might be useful for the creator (and ultimately the reader) of the meta-document. As described above, this information can be linked to the document or inserted into the document. The meta-document server 200 dynamically maintains the information space 4200 such that newly inserted input by the user causes the system to update the meta-document's information space. Furthermore, some of the services of a personality used to create the document information space maybe be periodically carried out thereby resulting in new markup/content for the document as new content is added to the document.

G.2.1 Creating And Updating Auto-Completion Database Entries

[0384] In particular, the auto-completion system shown in Figure 44 illustrates the manner in which the entity database 4214 is used for auto-completion, as well as, service 4406 for carrying out the process set forth in Figure 45 for creating and updating the entity database 4214 dynamically from the document information space 4200. In one embodiment, the service 4406 accessed by scheduler 204 begins at 4504 by initializing the database with entities from lexicons associated with the personality that has been assigned to the meta-document. In alternate embodiments, the database is either initialized using an empty database or it is initialized using a database of domain specific lexicons. In operation, the lexicons are used to identify entities in the document content that are to be enriched by predefined services (see for example Figure 4).

[0385] Subsequently at 4506, the module 4406 waits for a signal from text editor 4314 that document content 4203 has been added and/or edited. At 4508, the information space is updated based on the added and/or edited document content. At 4510, the updated information space (i.e., added and/or edited document content and enrichment associ-

ated therewith) is processed for entities that could potentially be used for auto-completion. At 4512, if extracted entities are deemed to be appropriate for auto-completion, then they indexed and inserted into the database of entities 4214; otherwise, or upon completion of 4512, the service 4406 waits for additional signals from the editor 4314.

[0386] As illustrated in the flow diagram shown in Figure 45, populating the auto-completion database is an ongoing process, which involves scanning the dynamic information space of the document for entities that could prove useful for auto-completion. The process of entity extraction for auto-completion varies according to the type of entity extracted. Considered first is text based entity extraction. A text-based entity is defined as a word or collection of words that appear contiguously in the document information space.

[0387] An entry that is inserted into the auto-completion database for a text entity as shown for example in Figure 48 includes: (a) a key or multiple keys (e.g., all possible n-grams, such as bi-grams or tri-grams, that make up a word or phrase) for specifying entity fragments to be searched; (b) the expanded entity relating to the entity fragment (i.e., word or words making up the entity, which may be delimited by punctuation characters such as spaces, fullstops etc. or using grammar rules which chunk words together into semantic entities such as noun phrases, verb phrases etc.); (c) any markup (such as hyperlinks, cross-references, footnotes etc.) that is associated with the entity; (d) any formatting (such as bold, italic, font size, etc.) that is associated with the entity; (e) the origin of the entity (e.g., location of the document containing the entity, segment containing the entity, etc.); (f) the position of the entity at its origin; (g) an identified part of speech of the entity at its origin; and (h) the context (e.g., categorization) of the entity at its origin.

[0388] Other types of information stored in the database that are useful for suggesting more accurate completions include bibliographic entries and related citations. Such entries and citations can be stored in the database as markup and recognized using known pattern recognition techniques and machine learning techniques such as hidden Markov models. Once recognized, this markup can be stored in the auto-completion database in similar fashion as the entities. The key in the case of a bibliographic entry could consist of the authors names, a subset of the characters that make up the authors names, or the citation associated with the bibliographic entry.

[0389] In addition, generic objects can also be recognized and recorded in the auto-completion entity database. A generic object can viewed as being made up of a sequence of inputs such as mouse movements, mouse clicks, keyboard inputs, human gestures as identified by a gesture recognition system, and facial expressions as recognized by facial recognition system. Such input sequences can be stored in the auto-completion database and be indexed by the first n inputs in the sequence. For example, consider an input sequence that consists of four straight lines that form a rectangle. This sequence could be retrieved and used for auto-completion of rectangles once the first one or two lines have been input, thereby alleviating the need for drawing the rest of the rectangle. These input sequences could be identified automatically using known data mining techniques, which search for general patterns in the input sequence.

[0390] It will be appreciated by those skilled in the art that when using the method outlined in Figure 45 for populating the auto-completion entity database, the entity database can grow to be prohibitively large, therefore, some entity selection algorithms should be used at 4510 to select which entities will provide the most benefit to the user in terms of time saved through auto-completion of these entities. For example, text based entities could be selected based on the length or the utility of the entity or combination of these.

[0391] Alternatively, utility measures such as Zipfs law could be used for entity selection. Zipf's law, which is well know in the art, concerns the distribution of different words in a corpus such as the information space surrounding a document, the online content available through the World Wide Web or some other domain specific corpus or a combination of the aforementioned. Zipf's law states that the product of a word's rank (r) and frequency (f) is a constant (C) i.e. $r*f = C$. Consequently, words/phrases that occur very rarely may ignored by the auto-completion system. In another embodiment, text terms could be selected based on the part of speech tags. For example, select only noun phrases from the information space.

[0392] Also, other factors such as the length of entities, highlighting information (i.e. are headings, bold, hyperlinked, etc.), markup information (such as hyperlinks, footnotes etc.), location of the entity in a document, its frequency in a document (or within a corpus) could used any combination to determine the utility of inserting the entity into the entity completion database. Those entities with a utility above a certain threshold are selected and inserted into the entity database. In one embodiment, the utility of an entity is determined using a weighted linear combination of factors as set forth below:

$$Utility(entity) = \sum_{factors} weight_{factor}(factor) =$$
$$weight_{bold}(bolded(true = 1; false = 0)) +$$
$$weight_{italic}(italic(true = 1; false = 0)) +$$
$$\vdots$$
$$weight_{uppercase}(upercase(true = 1; false = 0)) +$$
$$weight_{location}\left(1 - \frac{location\ of\ word}{document\ length}\right) +$$
$$weight_{frequency}\left(\frac{frequency\ of\ word\ occuring\ in\ document}{highest\ frequency\ of\ any\ word\ in\ document}\right) +$$
$$weight_{corpus}\left(\frac{frequency\ of\ word\ occuring\ in\ corpus}{highest\ frequency\ of\ any\ word\ in\ corpus}\right)$$

[0393] Additional factors include, heading information, footnoted, hyperlinks, comment. The weights associated with each of these factors in one embodiment be a uniform weight for each factor (i.e., uniform weight = one/number of factors). Alternatively, each weight can be set by a user, or determined automatically using known optimization techniques such as Powell's direction set minimization techniques or genetic algorithms.

G.2.2 Document-Centric Auto Completion

[0394] Figure 46 illustrates a logic flow diagram for selecting words using the auto-completion system shown in Figure 44. In box 4602, a request for word auto-completion is received. In one embodiment, a user types in the initial characters of a word (e.g., the first two, three, four, etc. characters of a word). The user may then invoke an auto-completion process in module 4302 by selecting a request key such as a right arrow key on a keyboard. In alternative embodiments, the system may automatically invoke auto-completion without having to be prompted by a user.

[0395] Once the auto-completion process is invoked, the string of characters typed by the user, hereafter referred as the string fragment or more generally referred to as the entity fragment, is used at 4604 to extract context information using content surrounding the entity fragment in the document content 4203 to which the entity fragment is targeted, herein also referred to as the target document. In one embodiment, portions of the target document (e.g., the paragraph preceding the text fragment in the target document, all text currently forming the document content, etc.) are categorized using the categorizer 3610 to define a category (e.g., from a DMOZ ontology) to which content surrounding the text fragment in the target document relates.

[0396] Subsequently, a query is formulated at 4606 using the extracted context information and string fragment. In one embodiment, the query can simply be the string fragment. In alternative embodiment, the query can be expanded using various contextual information that may lead to more accurate suggestions for completion. For example, the auto-completion system could process the sentence of which the string fragment is a member using linguistic processing tools such as XeLDA (Xerox Linguistic Development Architecture) described in U.S. Patent Application Serial No. 09/221,232, which is incorporated herein by reference.

[0397] This linguistic processing could lead to further requirements on the entities that could be considered for auto-completion. For example, linguistic processing could determine that a noun phrase is the most likely word(s) to be input next. This expectation can be incorporated into the query thereby limiting the search to noun phrases beginning with the string fragment. The query could be as follows: key ="dig" + Part-of-speech-tag = noun phrase.

[0398] As described at 4606, additional information, such as the classification of the document into categories using the DMOZ ontology, could be used also to reduce the search space. For example, if the currently input document is a document about SubjectX then the query can be further refined to stating the classification of the document where these candidate strings occur must be about SubjectX. Consequently, the query in this case could be the following: key ="dig" + Part-of-speech-tag = nounphrase + class_of_document = SubjectX.

[0399] At 4608, the formulated query is submitted to the information retrieval system 4308 in the auto-completion module 4302. In operation, the information retrieval system 4308 locates matches subject to the constraints specified in the query using known matching techniques. The matched items are retrieved and ranked based on their level of appropriateness for completion (i.e., how well they satisfy the query constraints and possibly additional constraints such how near each matched item is to a previously completed item) at 4612. The top ranked match that contains the same (or similar) initial characters typed by the user is displayed for user acceptance at 4614. In an embodiment, the

suggested completion is displayed in a fashion that is distinct from the text that the user has typed. For example, if the user's text is shown in black, the completion will be shown in gray.

[0400] If a user accepts the word match offered in decision box 4614, then that word is selected for copying into a target document as shown in box 4620. For one embodiment of the present invention, the user accepts the completion by continuing to type text into the target document. Once the word auto-completion process is completed the user may continue typing into the target document or may repeat this word auto-completion process.

[0401] If a user rejects the word match offered in decision box 4614, then the top ranked match is removed from the list of possible completions. For one embodiment of the present invention, the "up arrow" key is used both to reject the present completion and to request an alternative completion. If an alternative match is desired in decision box 4618 after removing top ranked results at 4616, then the alternative word match is displayed for user acceptance in box 4612. Boxes 4612, 4614, 4616, and 4618 are repeated as long as the user keeps rejecting the offered word match and additional matches are available.

[0402] Once no alternative word matches are available in decision box 4618, then the auto-completion of the string fragment terminates. The user may stop the system from cycling through the already rejected word matches by selecting an "END" key, or deleting the most recently offered word match.

G.2.3 Examples Of Document-Centric Auto Completion

[0403] Figure 47 illustrates an example of the auto-completion process performed using the auto-completion entity database presented in Figure 48. More specifically, Figure 47 illustrates an example of string fragment "dig" 4702 of a target document typed by a user into a window 4704 in which a text editor operates. Once the user invokes the auto-completion process using the example special character ">" 4706, the system in one embodiment identifies context using document selection 4705. Note that document selection 4705 can either be manually (i.e., by a user) or automatically (i.e., by the system) selected, as specified for example in user preferences of the text editor.

[0404] Subsequently as indicated by arrow 4710, a document category is identified using for example a DMOZ ontology using the document selection 4705 and a categorizer to define a context for the string fragment 4702. The context information (e.g., science) is used to limit the search for matches between the string fragment and entries in the key column of the auto-completion entity database shown in Figure 48. In addition, matches identified in popup window 4708 can be limited and/or ordered by comparing part of speech information relating to the string fragment 4702 using the document selection 4706 and the part of speech information relating to entities in the auto-completion entity database shown in Figure 48. After the auto-completion system identifies one or more matches, the sorted results are displayed for user acceptance as shown in popup window 4708.

[0405] A user may accept an offered word or phrase by selecting one of the available choices in the popup window 4708. In one embodiment, the popup window 4708 contains the highest-ranking matches, where each match can be expanded to find additional specifics on each match. A simpler interface is a popup window that displays only the highest-ranking match. In the example shown in Figure 47, the entity "digital" 4722 is expanded to include three different possible forms that the entry may take. Each subentry of the entry 4722 is derived from the information in the auto-completion database shown in Figure 48.

[0406] Generally, the entries and subentries in the popup window 4708 may contain a word or a word sequence that are displayed in a unique manner to indicate that it is being offered to the user for acceptance. If the offered word(s) are accepted, feedback is provided to the user that these words were selected for copying into the target document. Note that the entry or entries offered to the user may be one word, a phrase, all words until the end of a sentence, all words until the end of a paragraph, or some other grouping of words.

[0407] An alternate embodiment of presenting suggested completions of the entity fragment is to provide snippets/segments of the document from where the suggested completions were extracted using the origin information associated with each indexed entity in the auto-completion database possibly in a separate sub-window. These snippets/segments can be presented in their original format recorded in the auto-completion database or in more standard format such as the format of the target document. Furthermore, the first word/phrase match is displayed with additional emphasis. The first word/phrase match can be displayed with additional emphasis (e.g., highlighted) to indicate that it is being offered for user acceptance.

[0408] After a word or phrase is accepted by a user (e.g., simply by clicking with a pointer thereon) in the user in popup window 4708 as indicated by arrow 4716, then feedback is provided to the user in that the word match was selected and copied into the target document with appropriate formatting and/or enrichment (e.g., links to other content). If the user decides to stop the auto-completion process, then the auto-completion process is terminated. Otherwise, additional word(s) may be offered to the user for user acceptance by selecting RE-RANK at 4720. The auto-completion process terminates if the user decides to do so by selecting 4722 or automatically if no new words are offered to the user (e.g., it loops back to the first word matched).

[0409] In a variation of the auto-completion system, the auto-completion system is used with a voice recognition

system to complete word formulations such as individual names. For example, orally the computer can be asked, please suggest names with first name Bob and last name beginning with the letter "F". In yet another variation, words that have already been auto-completed in the target document can be cached and used for future auto-completions in the target document.

G.2.4 Document-Centric Auto Correction

[0410] Figure 49 illustrates a document-centric auto-correction system 4960 that iteratively corrects errors in meta-document 4202 using information space 4200. The original (i.e., non-corrected) document content 4203 in meta-document 4202 originates from input data 4952 which may be generated by a digital signal generator 4954 (e.g., scanner) and converted to text by converter 4956 or which may be generated by text generator 4958 (e.g., editor). The converter 4956 may for example perform OCR of scanned text, ICR (Intelligent Character Recognition) recorded handwriting, and speech to text recognition.

[0411] The auto-correction system includes modules for performing error correction. These modules operate in accordance with the process set forth in the flow diagram shown in Figure 50. Initially at 5000, control module 4964 receives a text object either from text generator 4958 or text converter 4956. Subsequently at 5018, the control module 4964 initializes the number of iterations of correction performed to the text object by the auto-correction system 4960 to zero, and in addition at 5020 a personality is attached to the text object.

[0412] Once a personality is attached to the text object, the control module 4964 spawns a thread for developing information space 4200 at 5022. The spawned thread runs in update module 4972 concurrent or in parallel with the subsequent action 5024 of waiting a predefined period of time for a signal that the entity database has been updated.

[0413] After the spawned thread in update module 4972 initially develops the information space around the attached personality at 5008, the entity database 4214 is extracted from the information space 4200 at 5012. These two actions are performed as set forth above in section for the auto-completion system described in section G.2. Subsequently at 5014, the thread raises a signal that the entity database 4214 has been updated and determines whether the information space has been modified at 5010.

[0414] If the information space has been modified at 5010, then the thread repeats action 5008 to further develop the information space 4200 and thereafter repeating actions 5012 and 5014. In the event the information space has not been modified at 5010, then a determination is made as to whether the text object has been corrected at 5002. If the text object has been corrected, then the thread terminates at 5004; otherwise, the action of determining whether the information space has been modified at 5010 is repeated.

[0415] After waiting the predetermined period of time for a signal that the entity database has been updated at 5024, errors in the text object (i.e., document content 4203) are identified and corrected at 5026. Subsequently, the number of iterations performed is incremented at 5028. At 5030, if the difference between the number of errors corrected at the current iteration and the number of errors corrected at the previous iteration is less than a threshold value, then the text object is determined to have been corrected at 5032 (which triggers a positive evaluation at 5002 terminating the thread); otherwise, the act of waiting a predefined period of time for a signal that the entity database has been updated at 5024 is repeated.

[0416] Figure 51 is a flow diagram depicting a process for identifying and correcting errors in the text object (i.e., document content 4203) using the entity database 4214, at 5026 in Figure 50. The process performed in Figure 51 is performed by the following modules in combination: error finder module 4966, query formulation module 4968, information retrieval module 4970, and error correction module 4962. Initially at 5140, the text object is tokenized to a list of words. For a next word in the list of words at 5142, it is examined to determine whether it has already been corrected at 5144. If it has been corrected, then a determination is made whether the word was the last word in the list at 5160. If it was the last word then this iteration of act 5026 terminates; otherwise, the next word in the list is processed at 5142.

[0417] When a word has not been previously corrected at 5144, then a query for the word is formulated at 5146. Formulating a query involves generating a query string that includes context information obtained from content that surrounds the word in the text object. At 5148, the query is submitted to the entity database 4214. The results from the query are ranked at 5150, from which the highest ranked results are selected at 51 52. Ranking techniques are described in section E.2 above.

[0418] At 5154, the highest ranked results are evaluated to determine whether any or one in particular satisfies or best satisfies evaluation criteria. The evaluation criteria include information associated with the entity or word being evaluated in the entity database 4214. The entity database includes such information at POS information, text category, and entity type, as shown in Figure 33. This information is matched against results using an evaluation criteria. An example of an evaluation criteria is the following: accept the highest ranking word if the word used in the act 5146 to formulate the query is not found in the first ten elements of the ranked results in 5150. If a result of the query is identified that satisfies the evaluation criteria, then it is used to correct the word in the text object at 5158. If no result of the query is found to satisfy the criteria at 5154 or the object word has been corrected at 5158, then the process continues at

5160 unless it is the last word in the list.

[0419] Correcting a word in the text object at 5158 modifies the information space 4200, which in turn causes the thread to determine at 5110 that the information space has been modified. In effect, each time the auto-correction system 4960 corrects at least one word, the text object in document content 4203 is modified. Each time the text object is modified, the information space around the document content can be further developed in accordance with personality attached to it. Modifying the information space in turn causes change to be made to the entity database.

[0420] In one embodiment, the information space is developed through multiple iterations, as auto-correction changes are added to it such as spelling corrections and/or content enrichment such as adding hyperlinks, copyright, and citation information to identified entities or words in the corrected textual content.

[0421] In an alternate embodiment, the auto-correction system 4960 in addition to textual objects processes image objects, and/or graphics objects. As set forth above, textual objects may be derived from handwriting, scanned textual content, converted audio. Image content may be in the form of scanned images, for example. In such an alternate embodiment, the document-centric auto-correction system performs graphics corrections. In this embodiment, graphics elements that have been drawn are evaluated and determined if they are have errors. Suggestions can be provided in the case when the shape comes close to a predefined shape such as a square. For example, if a foursided object has three sides that are the same length and a fourth side of different length, the system would suggest that the fourth (shorter length side) should be longer.

[0422] In one specific embodiment, a digital copier first scans a document and identifies text, graphic, and/or image content in the scanned document. This scanned information is automatically processed by the auto-correction system 4960 before being rendered on a hardcopy document. This embodiment permits scanned document content to be enriched and corrected before being rendered on a hardcopy document. Such corrections include performing image, copyright, citation, and spelling corrections.

[0423] Image corrections include identifying that an image (or any other content for that matter) set forth in the document content 4203 is not original by being, for example, degraded in some form. The auto-correction system 4960 would identify the original image in the information space 4200 surrounding the document content 4203. Once original content is identified, it is automatically put in place to correct the document content 4203.

G.3 User Directed Enrichment

[0424] This section specifies a further service offered by the meta-document server 200 after uploading and/or authoring a document thereon. In particular, this section discloses a method for formulating directed searches on heterogeneous sources of information such as the World Wide Web, and proprietary databases while authoring a document. The directed searches provide an alternative mechanism for gathering document enrichment that is formulated by the user. Advantageously, the document enrichment services allow a user to define a parameterized specification of enrichment and where the enrichment is to take place in the authored document content. That is, in authoring document content the service enables a user to request enrichment of identified content so that enrichment identified by the meta-document server may be integrated in real-time as a document is being authored.

[0425] Figure 52 illustrates a block diagram of the elements for forming a directed search. A user authoring the meta-document 100 using the text editor 4314 initiates directed search service 5206 that forms part of services database 210 in the meta-document server 200 shown in Figure 2 using a control keyboard function key, a window control button, or the like for interrupting the text editor 4314. The text editor 4314 may be any word processor adapted to edit textual content as well as possibly other content such as graphics data and image data. In addition, the text editor 4314 may be integrated with applications that provide and/or perform a variety of functions such as spreadsheet applications and database applications, and need not be strictly limited to word processing.

[0426] Once invoked, the directed search service 5206 makes available to the user an interface as shown for example in Figure 53 that provides the user with the ability to qualitative and quantitative specify criteria for the search. Once specified, the content manager 208 inserts the specified search criteria into a new document service request 106 of the meta-document 100. Scheduler 204 operating in meta-document server 200 as described above then initiates the new document service request 106.

[0427] Figure 53 illustrates one embodiment of a user interface 5300 for specifying a directed search, which can be invoked by selecting window control button 1040 in client interface 1010 (shown in Figure 10). In the interface 5300, the user is given the ability to specify a plurality of criteria for the search. At 5302, the user is able to select search criteria for performing the search. The search criteria in one embodiment may be specified using a service from services database 210. More generally, service requests may be launched while editing or viewing a document in any application program enabled with directed search capabilities. In this more general embodiment, a directed search may be invoked by selecting with a pointing device a particular section of a document while the document is being edited or viewed. In this general embodiment, settings of the directed search may be specified in a popup window similar to interface 5300 or automatically using a set of user and/or system specified default settings.

[0428] At 5304, the user is able to specify parameters of the selected search criteria (e.g., service) at 5302. At 5306, the user is able to specify the form of enrichment, which include links, content, or metadata. Links can be passive links or active links (e.g., a URL to a CGI-BIN script). Content can be formatted using the format of the original document content. Metadata can include predefined formatting specified by for example the user or the content service invoked. At 5308, the user is able to specify how frequently the search should be performed (e.g., once, daily, weekly, monthly, etc.).

[0429] In addition at 5310, the user is able to specify and/or select from one or more information service providers (i.e., information source). In the example shown in Figure 53, Thomson and Hoover as shown as selected information service providers, whereas Reuters is not. In the event the user does not specify an information service provider, one or more service providers will be automatically selected at runtime using other search criteria specified by the user in the request. Furthermore, of the information providers selected at 5310, the user is able to rank the order in which these selected information providers are to be used to carry out the directed search at 5318. For example in Figure 53, the user is provided with the option of either specifying that information service providers should be searched by cost (i.e., use those selected information providers that charge less first) at 5320 or by speed (i.e., use those selected information providers that will carry out the search the fastest) at 5322.

[0430] In alternate embodiments, ranking criteria may also include an alphabetical ordering, a predefined user specified ordering, a quality ordering (i.e., rank those information providers that provide the highest quality service first, independent of cost), preferred customer ordering, and privacy ordering (i.e., rank those information providers with the best privacy policy first). In yet another embodiment, multiple ranking criteria can be selected and ordered at 5318 and later used to select which of the information providers specified at 5310 to first carry out the directed search.

[0431] At 5312, the user is able to specify parameters for the content identified by the information source(s) selected at 5310. These parameters include specifying the language of the content, whether the content is premium (i.e., whether the user is willing to pay for it), the quality (i.e., higher quality content is more expensive), whether copyrighted (i.e., from a copyrighted source), whether free or fee-based, the maximum expenditure of a fee-based search, and the maximum length (e.g., in words) of retrieved content. Automatic summarization can be used to reduce content that exceeds the maximum number of words specified by a user, such as described in U.S. Patent No. 5,384,703, which is incorporated herein by reference.

[0432] At 5314 the user is able to specify exactly where and how the enrichment is to occur in the authored document content. In the embodiment shown in Figure 53, the user is able to specify that results in the form specified at 5306 are to be inserted at for example: (a) after the current selection of the user; (b) before the current selection of the user; (c) as a comment to the current selection; or (d) as a footnote to the current selection. The current selection is the selection of document content last specified before invoking the interface 5300. To aid the user, the current document selection is reproduced at 5315. If no specific word is selected, then a location is identified using some indicia 5317. In either case, whether the content identified before invoking interface 5300 are words, images, and/or graphics, context surround the content is provided at 5315.

[0433] To formulate output of the directed search specified in the interface 5300 for the meta-document, the insert button 5316 is selected by the user. Figure 54 illustrates an example of the output of the directed search specified in Figure 53. Specifically, Figure 54 shows XML tagged translations 5400 of the information entered in the interface 5300. This document service request once invoked by the scheduler 204 will lead to the execution of a query. As set forth above the meta-document server as shown in Figure 4 executes a document service request that may fetch, filter and/or summarize content. The results of the query in the form of additional document markup 108 are then added to the meta-document 100.

[0434] In the event the service provided by a content service provider is fee-based, then the rights of material are purchased within the constraints specified by the user. Such purchases may for example a rights management services such as ContentGuard™. Document content 102 that is inserted and/or annotated can be color coded to indicate that it is fee-based content. In addition, color-coding can be used to indicate that a copyright to the content has or needs to be acquired. For example, if the maximum set price by a user is exceeded for document content, a notice can be inserted in the document content indicating that enrichment exists for the particular location by a particular content provider at an identified price.

G.4 Exporting/Importing Enriched Documents

[0435] This section describes a service for exporting and/or importing enriched meta-documents. This service allows user to exchange meta-documents that have been enriched at different meta-document servers. Generally, a meta-document is exported using an exchange format that includes specifying and/or describing all or portions of the meta-document 100 shown in Figure 1. The exchange format allows binding between identical or dissimilar meta-document formats. The exchange formats can be represented using one or more exchange files using any appropriate media (e.g., email). In the event an imported exchange file is dissimilar to the exchange formats of the importing meta-document

server, a method is provided for binding any dissimilarities with services and/or personalities available at the importing meta-document server. The binding of dissimilarities are tracked so that the meta-document can be exported to the originally exporting meta-document server and accurately re-mapped to the importing meta-document server.

[0436] The exchange process is initiated when a command is received to either import and/or export a meta-document. The exchange process can be used for exchanging meta-documents with another meta-document server and/or for archival or backup. In one embodiment, the exchange process is invoked by selecting import/export button 1050 at client interface 1010 shown in Figure 10. It will be appreciated, however, that a meta-document exchange can be initiated either manually, semi-automatically, or automatically by a user and/or a system depending on the particular purpose for using the exchange process. In response to a command to export/import one or more selected meta-documents, the exchange process is invoked. In one embodiment, the exchange process operates as part of content manager 208 in meta-document server 200 (shown in Figure 2).

[0437] Figure 55 illustrates one embodiment of an interface 5500 for specifying a meta-document exchange, which can be invoked by selecting button 1050 in client interface 1010 (shown in Figure 10). In exporting a meta-document, the exchange process creates an export package that may be composed of a set of one or more export files. The set of export files in the export package encodes fully or partially the meta-document 100 specified at 5502. In exporting a meta-document, the exchange process is not concerned whether other meta-document servers have knowledge of the format used to export the meta-document. Accordingly as described below, the exchange process is adapted to bind information in an export package to its system when importing a set of export files.

G.4.1 Exporting

[0438] As shown in Figure 55, the exchange process is adapted to export selected components of a meta-document at 5504. The general features of a meta-document 100, which is shown in Figure 1, include: original document content 102, document markup 108 (i.e., document enrichment); and one or more personalities 104, each identifying one or more document service requests 106 and an entity database 111.

[0439] In one export format, a personality of a meta-document is represented using a set of services $S_{1,1}$ to $S_{1,j}$ and services providers $SP_{1,1}$ to $SP_{1,k}$. In another export format, the services and service providers are grouped into one or more personalities from available personalities $P_{1,1}$ to $P_{1,i}$. In this alternate export format, a personality identifies a named set of services and associated service providers. A detailed example of one possible export format is described in section G.4.2 below.

[0440] These different export formats may permit a set of dictionaries (i.e., entity database) $D_{1,1}$ to $D_{1,\lambda}$ that is used by the services to be recorded with a personality. Each dictionary is a static list of terms or regular expressions for identifying concepts within document content to be enriched (i.e., entities). Also these different export formats may permit a set of strategies $St_{1,1}$ to $St_{1,m}$ that are used to identify key concepts within the document to be specified within a personality. Strategies encode the order and mode in which entities in dictionaries are applied (see Figure 61 and description thereof below). The mode of entities in a dictionary can be negative or positive, depending on whether a dictionary is used to identify concepts to be enriched or to eliminate concepts from the enrichment.

[0441] In addition, advanced export control features can be specified at 5506 shown in Figure 55. One advance export control feature allows the exchange history (i.e., import and/or export) of a meta-document can be specified. The exchange history specifies whether a meta-document has already been exported from and/or imported to another meta-document server. In one embodiment, each meta-document server is identified by a unique set of properties that identify itself and possibly the services it uses (e.g., server id, service descriptions, providers, etc.).

[0442] Also the results of the last enrichment performed on a meta-document can be specified as an advanced export feature. The last enrichment can include additional content, links, and metadata. This advanced feature need not be specified to accurately export a meta-document since the importing meta-document server will recreate an enriched document using the exported document markup in any case. However, exporting this data in addition to document markup provides an importing meta-document server with additional information to bind to services in its services database that do not map to services defined in the imported meta-document. In the event no direct mapping between services exists, the results associated to the unavailable service(s) are displayed in the imported meta-document are marked with a status "frozen". A service can also be "frozen" if it cannot be accessed, which may occur when a service cannot be accessed because it is either unavailable because it is down or the meta-document server trying to access it cannot (e.g., it is off-line).

G.4.2 Exchange Format

[0443] Figures 56, 57, 58A, and 58B illustrate a detailed example of an export format. In this example, an exported meta-document is represented using a single file, although it will be appreciated that multiple files may be used to accomplish the same result. In one embodiment, multiple files are compressed and packaged into one file to represent

the export file. Whenever this export file includes other documents, for example, the original document content of these documents needs to be included in the export file. For example, if original document content is formatted in HTML, then all the necessary HTML frames, images, style sheets, and JavaScript files needs to be included in the export files so that the original document content can be rebuilt at the importing meta-document server.

[0444] Generally the export file includes everything needed to rebuild a meta-document document at the importing meta-document server. More specifically, as shown in Figures 56, 57, 58A, and 58B, the export file includes all information relating to personalities, services, providers, dictionaries and strategies used to enrich document content. Advantageously, a meta-document can be completely or partially rebuilt at an importing meta-document server depending on how well services in the export file match with services available at the importing meta-document server.

[0445] It will be appreciated by those skilled in the art that although the export file format shown in Figures 56, 57, 58A, and 58B is described using XML, any equivalent format or syntax could be used to describe the export format. It will also be appreciated that although two types of results are illustrated in the export file format (i.e., entities and global results), the export file format may instead include additional information produced during document enrichment (e.g., keywords, categories, etc.).

[0446] In addition, this export format can also be used to export and then import complete services or even full personalities between meta-document servers provided that additional data is recorded in the export file(s). The additional data includes data describing connectors to information providers (i.e., wrappers) and the format of dictionaries and strategies.

G.4.3 Importing

[0447] Referring again to Figure 55, which illustrates one embodiment of an interface 5500 for specifying a meta-document exchange invoked by selecting button 1050 in client interface 1010 (shown in Figure 10). At 5508, the name of an exported meta-document file produced by the exchange process operating on a meta-document server (i.e., the exporting meta-document server) is specified by an importing meta-document server. In requesting that an exported meta-document file is imported, the user is provided with the ability to specify importing all or selected portions of the exported meta-document file into a new meta-document at 5510.

[0448] More specifically, the exchange process is adapted to import an exported meta-document file into a new meta-document, where the exported meta-document file may include one or more of the following elements: Personalities $P_{2,1}$ to $P_{2,n}$; services $S_{2,1}$ to $S_{2,0}$; service providers SP2,1 to $SP_{2,p}$; dictionaries $D_{2,1}$ to $D_{2,q}$; and strategies $St_{2,1}$ to $St_{2,r}$.

[0449] When personalities, services, providers, dictionaries and strategies of the exported meta-document file do not match or strictly include the personalities, services, service providers, dictionaries and strategies used by the importing meta-document server, then the importing meta-document server may be requested to bind unmatched elements (e.g., personalities, services, dictionaries, etc.) with elements existing in the importing meta-document server. In one embodiment shown in Figure 55, the user is provided with the ability to specify whether to bind un-matched elements at 5512. Binding un-matched elements requires the importing meta-document server to identify a mapping between un-matched elements and compatible elements in the importing meta-document server.

[0450] In one embodiment, the following six actions are performed by the importing meta-document server to recreate a new meta-document given an exported meta-document file. It will be appreciated by those skilled in the art that this method is just one possible way to import an exported meta-document file and that alternate methods including the order of the actions and what is performed at each action may be varied to produce identical or similar results.

[0451] The first action involves extracting original document content forming part of the exported meta-document file and inserting the extracted original document content into the new meta-document.

[0452] The second action involves evaluating whether the personalities (i.e., $P_{1,1}$ to $P_{1,1}$) specified in the exported meta-document file are standard personalities. This second action is performed using properties that identify the personalities (e.g., name, creator, version number, unique identifier defined, for example, using the Digital Object Identifier standard, etc.). Standard personalities may, for example, be provided by software vendors and may be used by multiple systems. For all standard personalities, the exchange process matches the equivalent standard personalities from $P_{2,1}$ to $P_{2,n}$ to the personalities specified in the exported meta-document file. Identified matches are inserted into (or attached to) the new meta-document. Consequently, services and service providers associated with standard personalities are also inserted (or attached to), provided they correspond services available at the importing meta-document server.

[0453] The third action involves identifying "standalone services" that are specified outside a personality (i.e., $S_{1,1}$ to $S_{1,j}$). Similar to personalities, the third action matches standard services available at the importing meta-document server using properties that identify the services in the exported meta-document file. Subsequently at this third action, these identified services are inserted (or attached to) in the new meta-document file. Consequently, any local dictionaries and strategies associated with these services are also inserted (or attached to), provided they correspond dictionaries and strategies available at the importing meta-document server.

[0454] The fourth action involves creating at the importing meta-document server similar personalities for all un-

matched personalities (i.e., non-standard personalities). This fourth action includes: examining properties of services, service providers, dictionaries and strategies that are needed to implement the un-matched personalities (hereinafter referred to as "the examined properties"); and creating a personality with services, providers, dictionaries and strategies present at the importing meta-document server that have properties similar to the examined properties. Subsequently, these identified matches are inserted into (or attached to) the new meta-document.

[0455] The fifth action involves creating at the importing meta-document server similar services for all un-matched standalone services (i.e., non-standard standalone services), which is described in detail in section G.4.4 below. When an exported meta-document file is successively imported and exported by several meta-document servers, any mapping performed at the second, third, forth or fifth actions is applied to the original service description and not to a service description of a mapped service.

[0456] The sixth action involves recording those non-standard personalities and non-standard standalone services that were not successfully mapped to personalities and services of the importing meta-document server. In the event there exists binding errors, entities affected by these binding errors in the imported document content are marked as "frozen" so that the user understands that they cannot be updated.

[0457] Also these binding errors can be recorded with the new meta-document and associated with a meta-document server identifier so that they may be identified in the event that the new meta-document is imported by the exporting meta-document server. At such time the new meta-document is imported by the exporting meta-document server, any binding errors are eliminated by activating the content previously marked as "frozen". In addition, these binding errors can be either presented to a user for error reporting purposes or for requesting manual mapping to personalities and services existing on the importing meta-document server. Binding errors can be further eliminated as new services are added to the importing meta-document server. In this embodiment, as a new service is added to the importing meta-document server, meta-documents with unresolved binding errors (i.e., frozen content) that match against the new service are activated.

G.4.4 Matching Method

[0458] This section describes the fifth action (introduced in section G.4.3) that involves creating at the importing meta-document server similar services for all non-standard standalone services. Specifically at this fifth action, the importing meta-document server attempts to map all non-standard standalone services onto an existing service with dictionaries and strategies present at the importing meta-document server that have properties similar to the examined properties of the exported meta-document file. Subsequently, these identified matches are inserted into (or attached to) the new meta-document. Specifically, this mapping to non-standard standalone services described below is performed by first attempting to satisfy a category match; if a category match is successful, then a dictionary match, and a key match are performed. The category match is necessary but not sufficient for a binding to take place. That is, both the dictionary match and the key match must be satisfied in addition to the category match to successfully bind two services.

[0459] In addition to performing a binding function, the method described in this section for performing a mapping to non-standard standalone services can be used to add related service providers available at the importing meta-document server to services (standalone or referenced by a personality) already bound to the new meta-document. In operation, the user may select this added feature at 5514 in the interface 5500 shown in Figure 55. More specifically, this feature allows elements forming part of the exported elements that are bound to service providers available at the importing meta-document server to be augmented with additional related service providers also available at the importing meta-document server.

[0460] To achieve a category match between service, providers, dictionaries and strategies, the importing meta-document server develops an ontology of namespaces that describes all entities in the exported meta-document file. Each class of namespaces allows compatible entities to be classified in the ontology. In developing an ontology, any service and entity specified in the exported meta-document file is matched with a service in the importing meta-document server if identifications and descriptions correspond strictly to a common namespace (i.e., fall in the same category in the ontology).

[0461] In one embodiment, the importing meta-document server enhances the ontology of namespaces with mapping information from previous exported meta-document files that were imported. If a meta-document has been exchanged several times between the exporting meta-document server and the importing meta-document server, mappings between respective services of the exporting meta-document server and the importing meta-document server is more readily defined.

[0462] Thus, in the event services from the exporting meta-document server do not map exactly onto the description of services from the importing meta-document server (i.e., step four in section 4.3 above), then partial mappings between descriptions of services is examined. If a partial mapping between descriptions of services exists (e.g., namespaces "stock quotes delayed" and "stock quotes real time") then a dictionary match and a key match are initiated.

[0463] More specifically, if a non-standard standalone service from the exporting meta-document server cannot be exactly mapped to a service from the importing meta-document server because no service from the importing meta-document server has an exact matching description, then two services can be bound only if they are identified by the same category in the ontology and satisfy a dictionary match and a key match.

[0464] Assuming that each dictionary has an associated category in an ontology of namespaces, and that S and S' are services from the exporting and importing meta-document servers, respectively, such that: category(S) equals category(S'); and $\{D_1,...,D_n\}$ and $\{D'_1,...,D'_n\}$ are the categories of the dictionaries associated with the services S and S'. Also, assuming each service is associated with a set of providers, and each provider use a "key" that describes the type of information needed to satisfy a query. This key may be "generic" to the service provider (e.g., "TickerSymbol" for a stock quote service). Alternatively, this key may be "specific" to the service provider (e.g. "BloombergKeyId") when the service provider cannot use a generic key.

[0465] To determine if service S can be mapped to service S', the importing meta-document server evaluates whether the intersection of the categories of dictionaries associated with services S $\{D_1,...,D_n\}$ and the categories of dictionaries associated with services S' $\{D'_1,...,D'_n\}$ yields any matching dictionaries, and whether any of the keys of any of the matching dictionaries match. A specific key that is associated with a generic key is said to match this generic key. Alternatively, a first specific key is said to match a second specific key if they both can reduce to an equivalent generic key. If at least one of the keys of matching dictionaries match then the service S can be mapped to the service S' in the new meta-document. That is, services available at the importing meta-document server may be specified in the new meta-document when the services partially map to a predefined categorization, and have at least one common dictionary and key.

[0466] In one embodiment, a service S can be mapped onto a service S' by the importing meta-document server when one or more of the following mappings can be deduced: an equivalent mapping, a specific mapping, and a generic mapping. An equivalent mapping occurs when all of the dictionaries associated with the services S and S' are equivalent (i.e., $\{D_1,...,D_n\} = \{D'_1,...,D'_n\}$); in this case, services S and S' are said to be equivalent. A specific mapping occurs when all of the dictionaries associated with the services S map to a subset of the dictionaries associated with the services S' (i.e., $\{D\text{-}_1,...,D_n\} \subset \{D'_1,...,D'_n\}$); in this case, service S is said to be more specific than service S'. A generic mapping occurs when all of the dictionaries associated with the services S' map to a subset of the dictionaries associated with the services S (i.e., $\{D_1,...,D_n\} \supset \{D'_1,...,D'_n\text{-}\}$); in this case, service S is said to be more generic than service S'.

G.5 Alternate Embodiments

[0467] This section describes an alternate embodiment of the meta-document and meta-document server.

[0468] Figure 59 illustrates a meta-document with document ID 5902, document content 5904, document annotations 5906. The document ID 5902 records information concerning the meta-document such as a unique identifier, and properties such as owner, permissions, etc. The document content 5904 identifies the original content associated with the meta-document. The document annotations 5906 includes any annotations (e.g., comments) added to the original content 5904 by a user. In addition, a meta-document may include document markup 5908 of the document content 5904 and/or document annotations 5906.

[0469] Figure 59 also illustrates that the meta-document 5900 may include a list 5910 of one or more personality references 5911. Also, the meta-document 5900 may include a list 5912 of one or more service reference requests 5913. Figure 60 illustrates the contents of a personality 6000 that is referenced by a personality reference 5911. The personality 6000 includes a list of one or more service request references 5913. In addition, the meta-document 5900 may include an entity database 5914 that sets forth entities that are related in a document-centric way to the meta-document (e.g., user defined entities specific to the document content or document markup).

[0470] Figure 61 illustrates a service request 6100 that is referenced by a service request reference 5913. Each service request includes a description 6102, properties 6104 (e.g., refresh period between enrichments, notification method when content is enriched), a list of service providers 6106, strategies 6108, and dictionaries 6110 (i.e., entity databases). The list of service providers 6106 specifies a service provider and whether the service provider anticipates receiving a keyword or a key and/or requires a login identifier (and/or a new session at each login). A keyword is almost any content that is used to submit to a generic service provider (e.g., yahoo), whereas a key is content in a specific format that is used to access certain specific information (e.g., a ticker symbol to access a stock quote).

[0471] The strategies 6108 are used to identify key concepts within the document to be specified within a personality. Strategies are rules that encode the order and mode in which entities in dictionaries are applied to document content or markup. The entities in a dictionary can be applied negatively or positively, depending on whether a dictionary is used to identify concepts to be enriched or to eliminate concepts from being enriched. In the example shown in Figure 61, first people names are removed from those entities to be marked up, subsequently plurals as normalized and eliminated, and business names are identified.

[0472] In an alternate embodiment of the scheduler 204 and content manager 208 shown starting in Figure 2, the scheduler is given alternative or additional functions of identifying meta-documents in the meta-documents database 202 that need periodic updates performed thereto and awakens the content manager 204 (i.e., builder) as necessary. In addition, the scheduler can be programmed with maintenance events and/or batch processing events. The programming of such events can be performed by the system and/or a user. An example of a batch processing event, is the processing of enrichment of an identified folder of document(s) and/or document reference(s). The functions of the content manager in this embodiment are to identify document services that need to be satisfied. This can occur once it receives a meta-document to operate on either directly by a user at a client interface or by the scheduler. For each of the services that needs to be applied, the content manager builds a service request, triggers it, waits for results, and packages the results into markup of the meta-document.

[0473] Figure 62 illustrates an alternate embodiment of the client interface shown in Figure 10. The client interface 6200 shown in Figure 62 includes a command board 6202, which provides a user with the option of uploading a specified URL at 6203 by clicking upload button 6204. Depending on the manner in which a user has defined marking options (discussed below with reference to Figure 67), the uploaded document is either automatically enriched with a specified personality or enriched once a user specifies a personality in personality corner frame 6212.

[0474] When enrichment is either automatically or manually invoked a status window 6300 shown in Figure 63 is presented to a user for the specified document 6302 and the specified personality 6304. The status window shows the percentage completion 6306 for each specified service 6308 of the personality. Services may be individually paused at 6310 or stopped at 6312. A service that is stopped is terminated for example if there is no response from the service or paused if that response is known to be temporary. Generally, enrichment is not shown on the uploaded document until all service requests are completed. However, get view button 6314 allows a user to temporarily view the enrichment of a document before all service requests have completed. In the example shown in Figure 63, all service request have completed and the window will subsequently disappear and the uploaded document enriched.

[0475] Once a document is uploaded and/or enriched, an enrich link button 6208 on the command board 6202 allows a user to specify whether or not to enrich links that are followed in the document currently viewed. These links may be links that are enriched by the meta-document server or links that previously existed in the linked document. For example, in the document view frame 6220 of the client interface 6200 highlighted text "Palo Alto" identified by reference number 6222 as well as highlighted test "PARC Web Site" identified by reference number 6224 have been enriched by the meta-document server. The text 6224 unlike the text 6222 included a hyperlink (as shown by the underlining) in the original document. When text 6222 and the text 6224 are selected by a user, popup windows 6400 and 6500 appear, respectively, as shown in Figures 64 and 65. As shown in Figure 65, the original link is accessible at 6502 and may be viewed in either the meta-document client window or in a separate browser window.

[0476] More specifically, each of the popup windows 6400 and 6500 set forth those services which can be applied to an identified entity. For example, the search engine service 6402 shown in Figure 64 may be invoked using either a query 6406 enhanced to focus on a particular category of a search engine as set forth in section F.3. In addition, a query 6408 may be invoked that is not enhanced in any form. For both queries 6406 and 6408, in one embodiment the highest ranked result is automatically displayed and if specified enriched. It will be appreciated that in alternate embodiments, other types of queries such as those discussed in section F.3 above may be specified at 6402.

[0477] Global results frame 6214 shown in Figure 62 provides direct access to services of entities identified in uploaded and enriched document content. Folders 6216 and 6218 effectively provide the same information as shown in popup windows 6400 and 6500 shown in Figures 64 and 65 respectively. Thus entities with services attached to them may be invoked either directly through document view frame 6220 by selecting the entity or through global results frame 6214 through services to which entities have been linked.

[0478] A store button 6210 on the command board 6202 provides the user with the option of storing an uploaded and enriched document as shown in Figure 62. Unless a document is stored, its enrichment will only exist for the current session a user is logged in. In addition, under the file tab 6205 the user is given the ability to manage stored files. Figure 66 illustrates a document storing management view 6600 of a user's files. In this view, documents uploaded for the current session can be viewed in folder 6602. In addition, documents stored by a user are viewed in the view 6600 such as folder 6604. The storing management view 6600 also allows a user to add to and view shared documents at 6606.

[0479] A user options window 6700 shown in Figure 67 may be accessed by a user in the configuration tab 6206 shown in Figure 62. The user options window 6700 provides a user with the ability to choose a default personality or always be asked to select a personality at 6702 and 6704, respectively. In addition, the user options window allows the format of marked up content to be specified at 6706, whether the default personality should be applied to each uploaded document at 6708, and whether to always enrich followed links at 6710 (as opposed to a specific uploaded document at 6208 shown in Figure 62).

[0480] A services configuration window 6800 shown in Figure 68 may be accessed by a user in the configuration tab 6206 shown in Figure 62. For each specified service under each category of service, a user may specify: the type

(standard, or special); the status (active, inactive); the priority (high, medium, low); the information provider; the list of dictionaries used by the service; its refresh period; and whether to attach links or content. By selecting a dictionary list for a service as illustrated by pointer 6802, a list of dictionaries appears as shown at 6804. Thus by specifying which dictionaries (i.e., entity databases) that can be applied by a service, the service can be made more specific or more general.

H. Miscellaneous

[0481] It will be appreciated by those skilled in the art that the meta-document server 200 and other computer systems described herein (e.g., computer 226, mobile phone 219, etc.) include software components and hardware components, and that such computer systems can be either a personal computer, workstation, a mobile/cellular phone, a handheld device etc.

[0482] The hardware components include a Central Processing Unit (i.e., CPU), Random Access Memory (RAM), Read Only Memory (ROM), User Input/Output ("I/O"), and network I/O. The User I/O may be coupled to various input and output devices, such as a keyboard, a cursor control device (e.g., pointing stick, mouse, etc.), a display, a floppy disk, a disk drive, an image capture device (e.g., scanner, camera), etc.

[0483] RAM is used by CPU as a memory buffer to store data. A display is an output device that displays data provided by CPU or other components in a computer system. In one embodiment, display is a raster device. Alternately, the display may be a CRT or LCD. Furthermore, user I/O may be coupled to a floppy disk and/or a hard disk drive to store data. Other storage devices such as nonvolatile memory (e.g., flash memory), PC-data cards, or the like, can also be used to store data used by computer system.

[0484] The network I/O provides a communications gateway to a network 221 (shown in Figure 2) such as a LAN, WAN, or the Internet. The network I/O is used to send and receive data over a network 221 connected to one or more computer systems or peripheral devices.

[0485] The software components includes an operating system software, application program(s), and any number of elements of the meta-document server 200. It should be noted that not all software components are required for all the described embodiments. The operating system software may represent an MS-DOS, the Macintosh OS, OS/2, WINDOWS®, WINDOWS® NT, Unix operating systems, Palm operating system, or other known operating systems. Application Program(s) may represent one or more application programs such as word processing programs, spreadsheet programs, presentation programs, auto-completion programs, editors for graphics and other types of multimedia such as images, video, audio etc.

[0486] It will also be appreciated that such application programs including any of the meta-document services of the meta-document server may be accessed through a "user interface" or "client interface" that may take one or more forms that include, graphical user interfaces, paper user interfaces, and application program interfaces (APIs). Thus, the use of the term "user interface" or "client interface" is defined herein to include access to services offered by a program invoked by a user through a graphical user interface, or the like, or by another program through an API.

[0487] The computer system may be implemented by any one of a plurality of configurations. For example, processor may in alternative embodiments, be defined by a collection of microprocessors configured for multiprocessing. In yet other embodiments, the functions provided by software components may be distributed across multiple computing devices (such as computers and peripheral devices) acting together as a single processing unit. Furthermore, one or more aspects of software components may be implemented in hardware, rather than software. For other alternative embodiments, the computer system may be implemented by data processing devices other than a general purpose computer.

[0488] Using the foregoing specification, the invention may be implemented as a machine (or system), process (or method), or article of manufacture by using standard programming and/or engineering techniques to produce programming software, firmware, hardware, or any combination thereof.

[0489] Any resulting program(s), having computer-readable program code, may be embodied within one or more computer-usable media such as memory devices or transmitting devices, thereby making a computer program product or article of manufacture according to the invention. As such, the terms "article of manufacture" and "computer program product" as used herein are intended to encompass a computer program existent (permanently, temporarily, or transitorily) on any computer-usable medium such as on any memory device or in any transmitting device.

[0490] Executing program code directly from one medium, storing program code onto a medium, copying the code from one medium to another medium, transmitting the code using a transmitting device, or other equivalent acts may involve the use of a memory or transmitting device which only embodies program code transitorily as a preliminary or final step in making, using, or selling the invention.

[0491] Memory devices include, but are not limited to, fixed (hard) disk drives, floppy disks (or diskettes), optical disks, magnetic tape, semiconductor memories such as RAM, ROM, Proms, etc. Transmitting devices include, but are not limited to, the Internet, intranets, electronic bulletin board and message/note exchanges, telephone/modem based

network communication, hard-wired/cabled communication network, cellular communication, radio wave communication, satellite communication, and other stationary or mobile network systems/communication links.

[0492] A machine embodying the invention may involve one or more processing systems including, but not limited to, CPU, memory/storage devices, communication links, communication/transmitting devices, servers, I/O devices, or any subcomponents or individual parts of one or more processing systems, including software, firmware, hardware, or any combination or subcombination thereof, which embody the invention as set forth in the claims.

[0493] The invention has been described with reference to particular embodiments. Modifications and alterations will occur to others upon reading and understanding this specification taken together with the drawings. The embodiments are but examples, and various alternatives, modifications, variations or improvements may be made by those skilled in the art from this teaching which are intended to be encompassed by the following claims.

Claims

1. A system for enriching document content, comprising:

   a user interface for specifying a personality defining a set of document service requests that identify enrichment themes;

   a user-manager for receiving document content and the personality; the user-manager forming a meta-document with the document content and the personality;

   a scheduler for selecting a document service request from the set; the scheduler initiating and managing communication with a service provider to satisfy the selected document service;

   a content manager for integrating results from the selected document service into the meta-document as document markup;

   wherein the user interface, the user-manager, the scheduler, and the content manager operate together to generate or recommend new personalities or modify existing personalities using one or more in combination of a set of algebraic computations, a set of document references, a set of predefined personalities, a set of learning personalities, a knowledge level, and a classification of personalities.

2. The system according to claim 1, wherein at least a first document service request in the set recognizes a set of entities, and at least a second document service request in the set accesses a portion of one of the document content and the document markup to carry out the second document service request.

3. The system according to claim 1, further comprising:

   means for displaying at a user interface a personality that includes a set of one or more services; each service in a personality having at least one of a set of service providers selected; each selected service provider having associated therewith one or more types of entities; and

   responsive to user input, means for modifying the personality to define one of a new personality and a modified personality by modifying the set of services.

4. The system according to claim 1, further comprising:

   means for defining a first personality having a first set of services;

   means for defining a second personality having a second set of services;

   means for applying an operator to the first personality and the second personality to create a third personality having a set of services that includes one or more of the services from the first set of services and the second set of services; wherein the third personality is a new personality.

5. The system according to claim 4, wherein the operator performs one of adding, subtracting, merging, composition, and intersection to services in the first set of services and the second set of services.

6. The system according to claim 1, further comprising:

an entity extractor for developing an expanded document with a set of identified documents and constructing an entity database with the expanded document; and

a service generator for creating a service request for each form identified in the expanded document and defining a new personality using the created service requests and the entity database.

7. The system according to claim 6, further comprising means for filtering the service request with no utility before defining the new personality using the entity database.

8. The system according to claim 6, wherein the set of documents is a set of links.

9. The system according to claim 1, further comprising:

means for receiving a user defined question;

means for identifying lexicons in the user defined question;

means for replacing the identified lexicons with generic symbols to define a generic question;

means for adding the generic question to one of a new personality and an existing personality.

10. The system according to claim 9, further comprising:

means for searching for entities that satisfy the lexicons in the generic question to define a specific question; and
means for answering the specific question using the entities that satisfy the lexicons in the generic question;

wherein the lexicons are defined in an existing personality.

11. A method for enriching document content, comprising:

specifying at a user interface a personality defining a set of document service requests that identify enrichment themes;

receiving at a user-manager document content and the personality; the user-manager forming a meta-document with the document content and the personality;

selecting with a scheduler a document service request from the set to initiate and manage communication with a service provider and satisfy the selected document service;

integrating with a content manager results from the selected document service into the meta-document as document markup;

operating the user interface, the user-manager, the scheduler, and the content manager together to generate or recommend new personalities or modify existing personalities using one or more in combination of a set of algebraic computations, a set of document references, a set of predefined personalities, a set of learning personalities, a knowledge level, and a classification of personalities.

12. The method according to claim 11, further comprising:

identifying a set of documents;

developing an expanded document with the set of identified documents;

constructing an entity database with the expanded document;

creating a service request for each form identified in the expanded document;

defining the new personality using the created service requests and the entity database.

13. The method according to claim 11, further comprising:

extracting features from the document content;

categorizing the features into a class of enrichment themes; and

recommending one or more personalities associated with the class of enrichment themes to enrich the document content.

14. The method according to claim 13, wherein the features are categorized into a plurality of classes of enrichment themes that are ranked using a user profile.

15. The method according to claim 13, further comprising:

creating a profile of the document content by categorizing the document content; and

using the profile for identifying experts to enrich the document content.

16. The method according to claim 11, further comprising providing at the user interface selections for one of inserting enrichment in the document content and linking enrichment to the document content.

17. The method according to claim 11, further comprising providing at the user interface selections for specifying how frequently to enrich the document content.

18. The method according to claim 11, further comprising providing at the user interface selections for specifying one of an expert mode and a novice mode to enrich the document content.

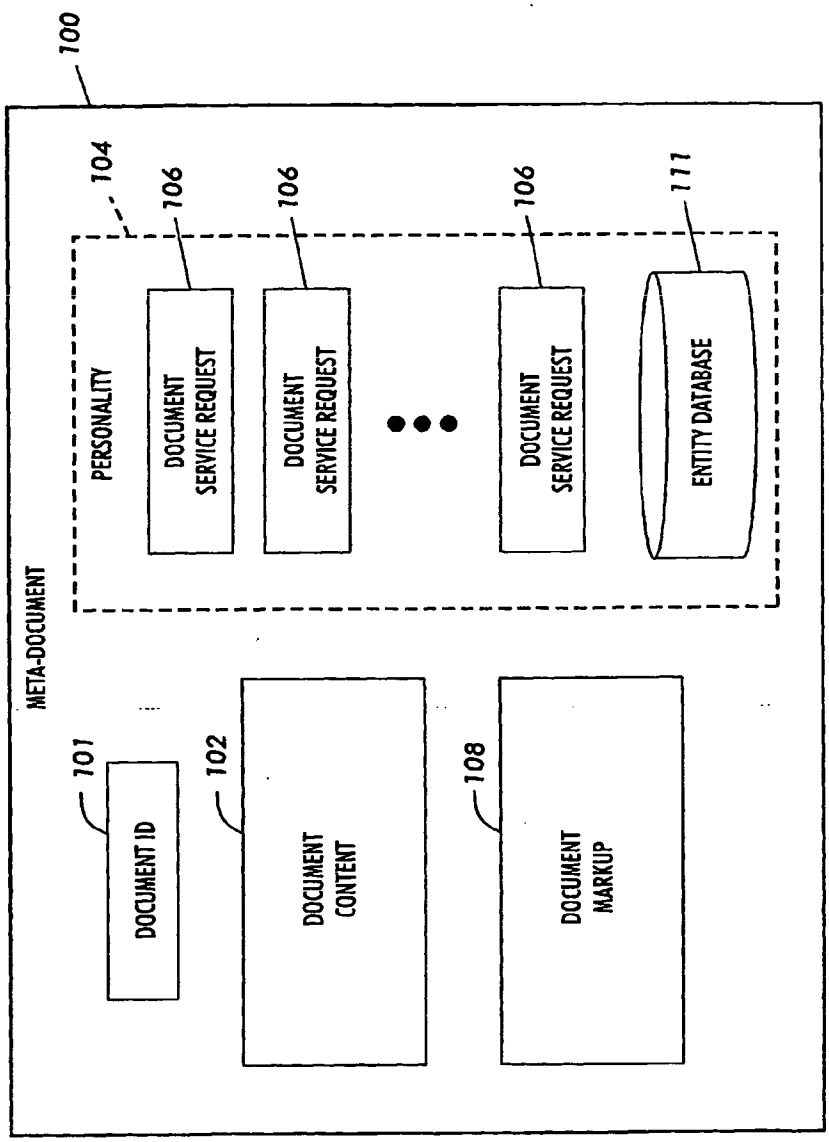


FIG. 1

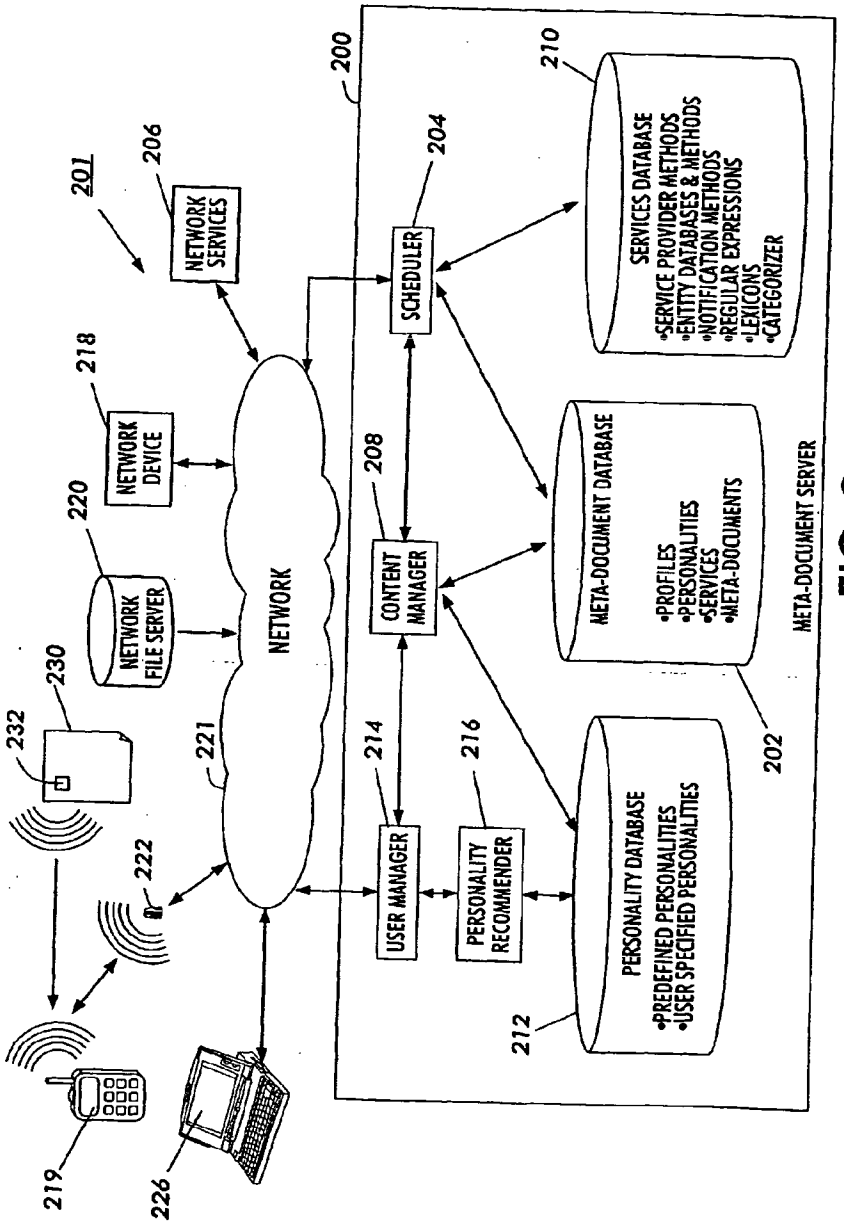


FIG. 2

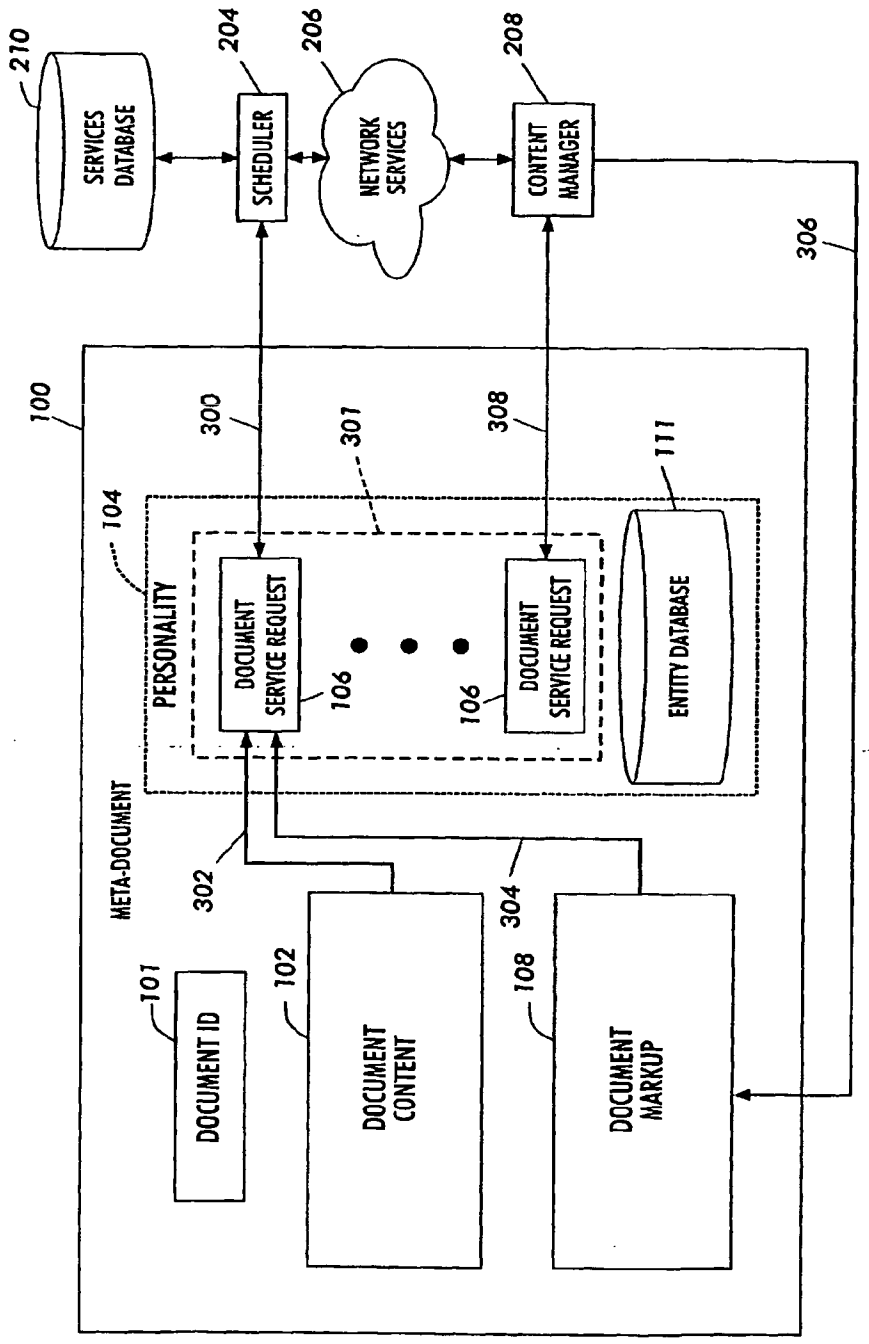


FIG. 3

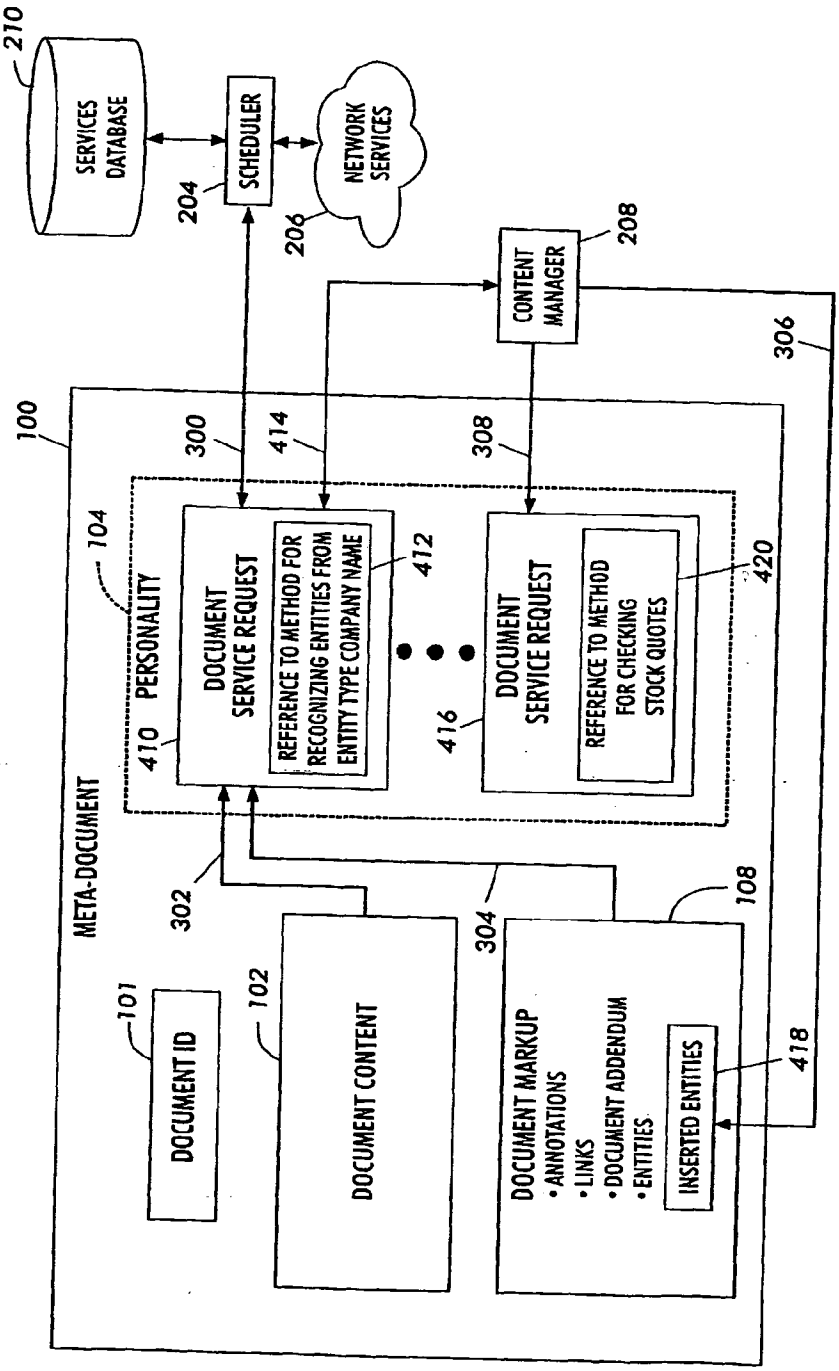


FIG. 4

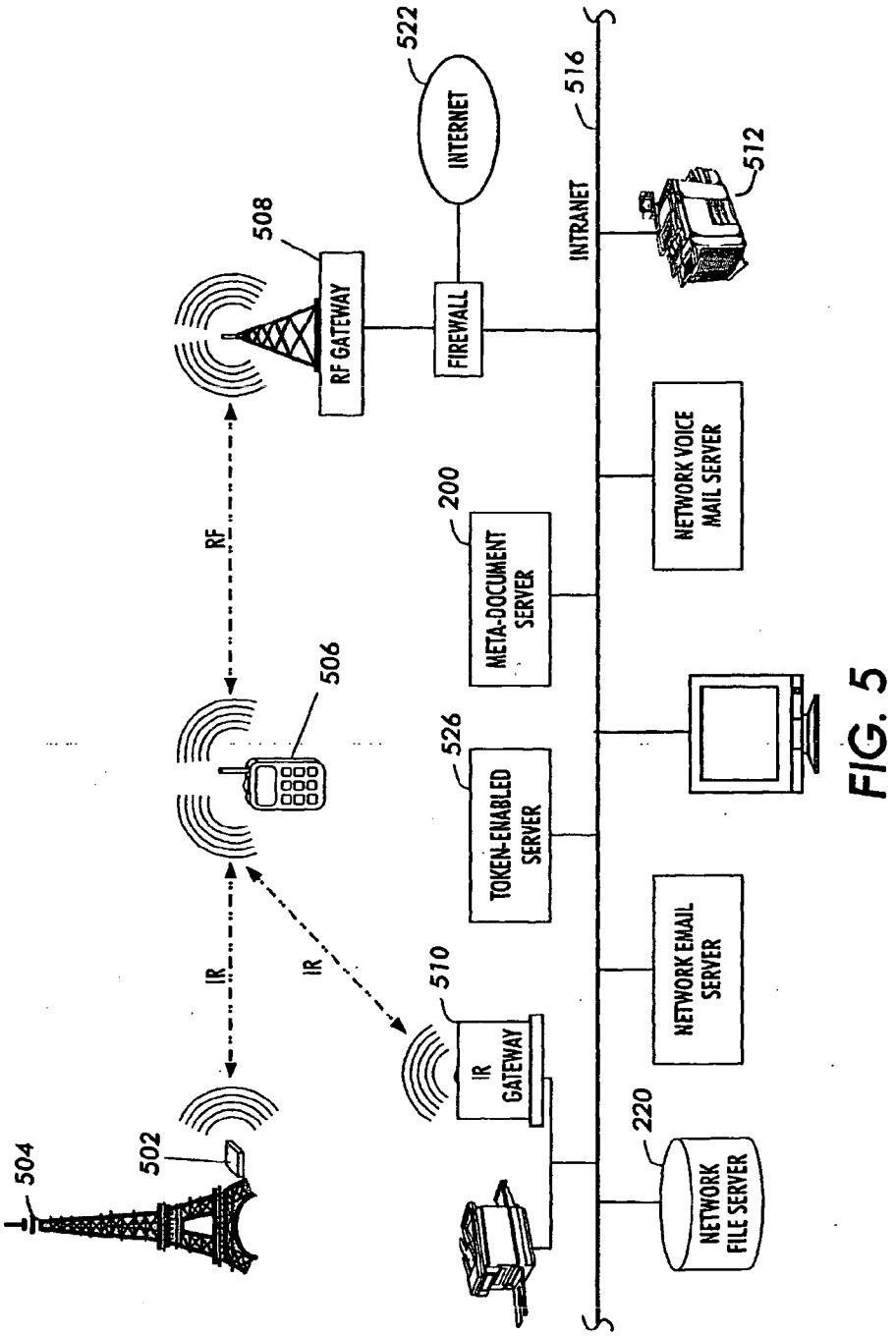


FIG. 5

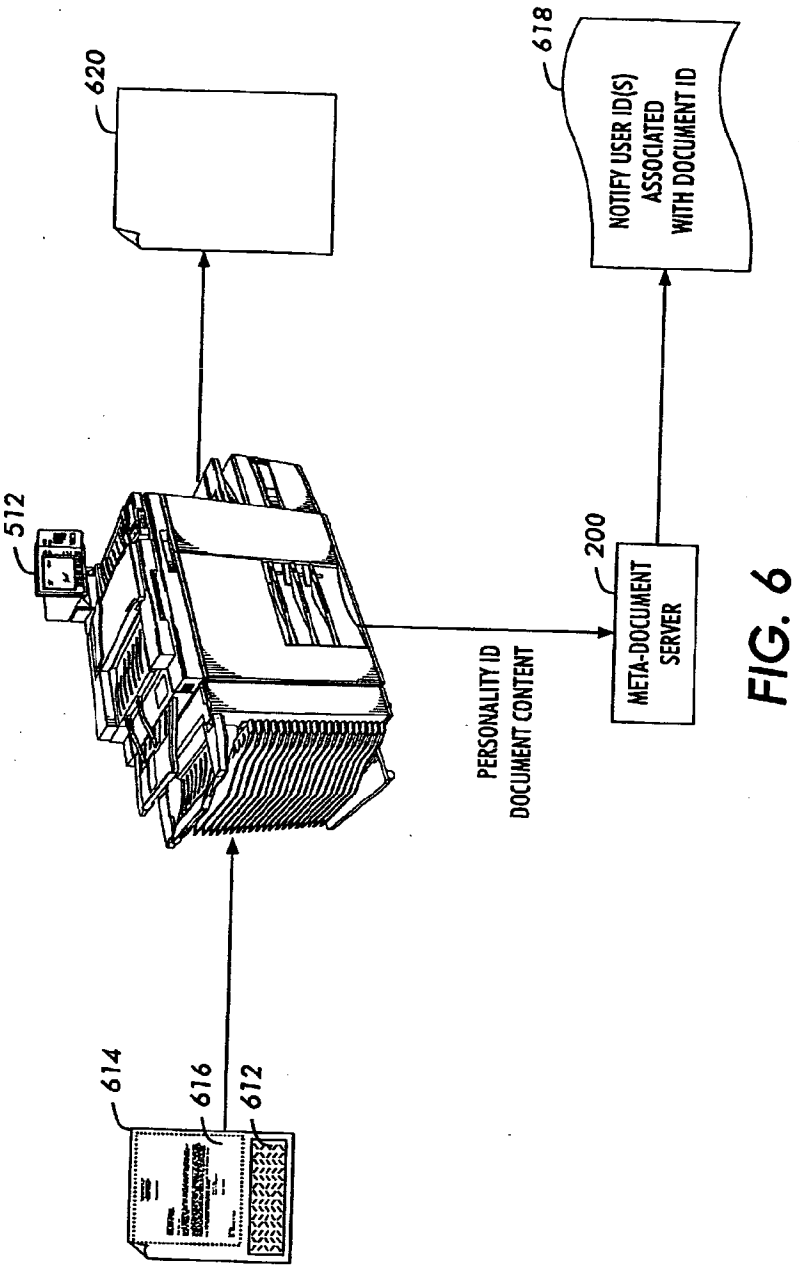


FIG. 6

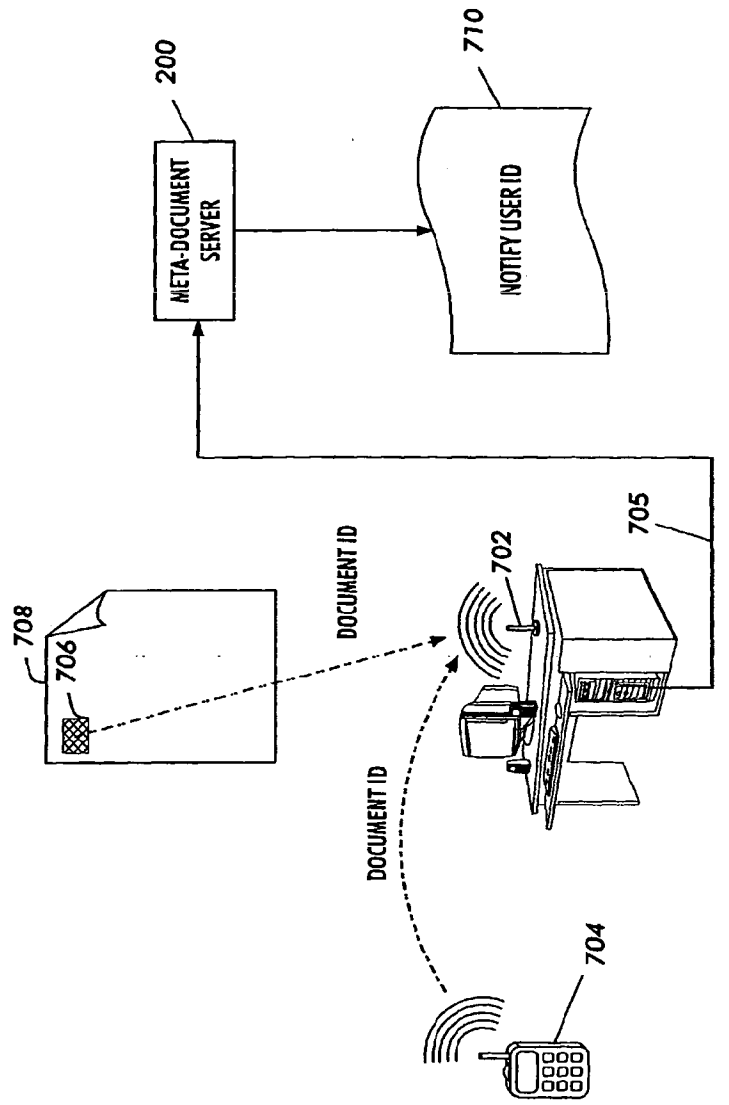

200
META-DOCUMENT SERVER
710
NOTIFY USER ID
705
• PERSONALITY/USERID
• DOCUMENT ID
702
DOCUMENT ID
708
706
DOCUMENT ID
704

FIG. 7

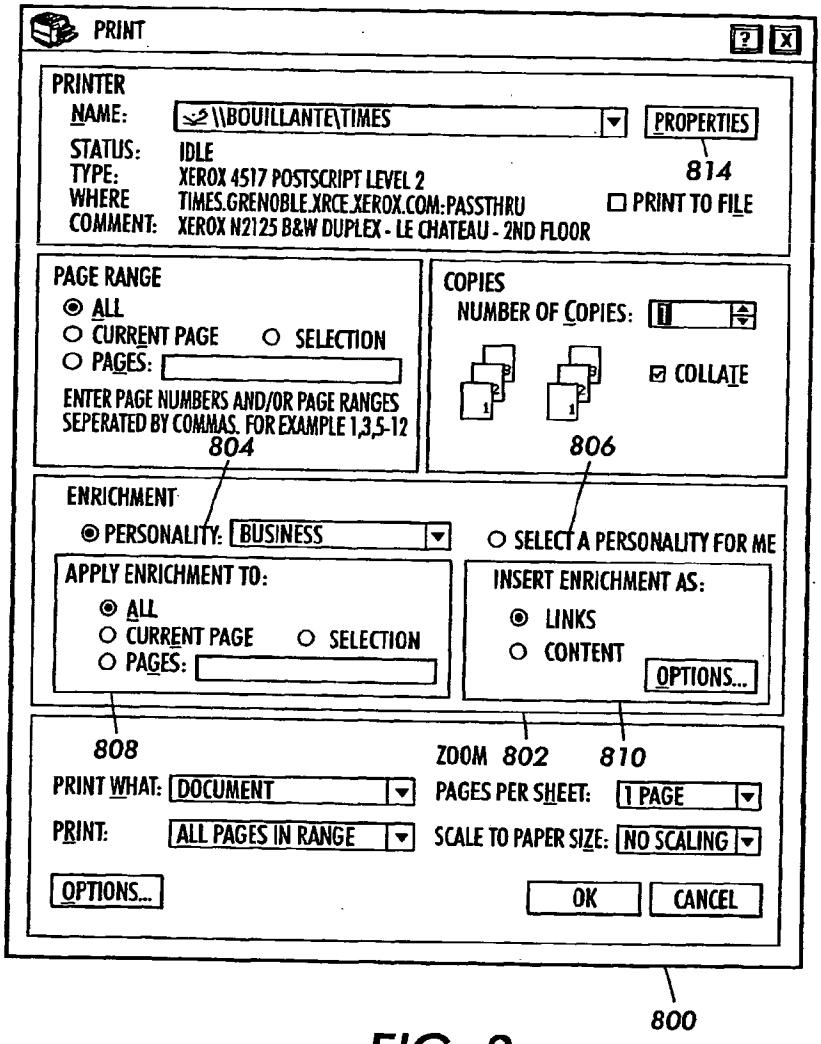


FIG. 8

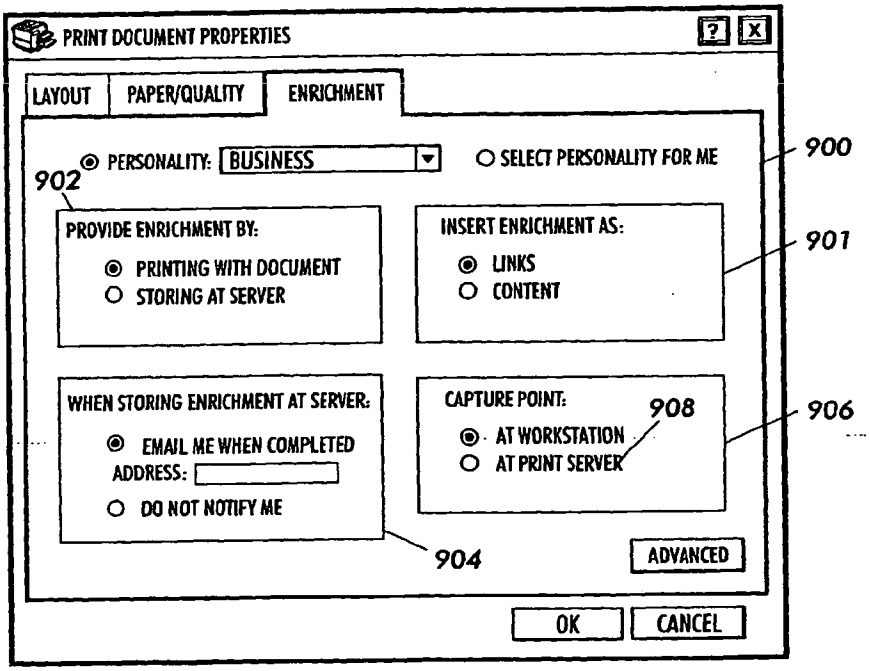


FIG. 9

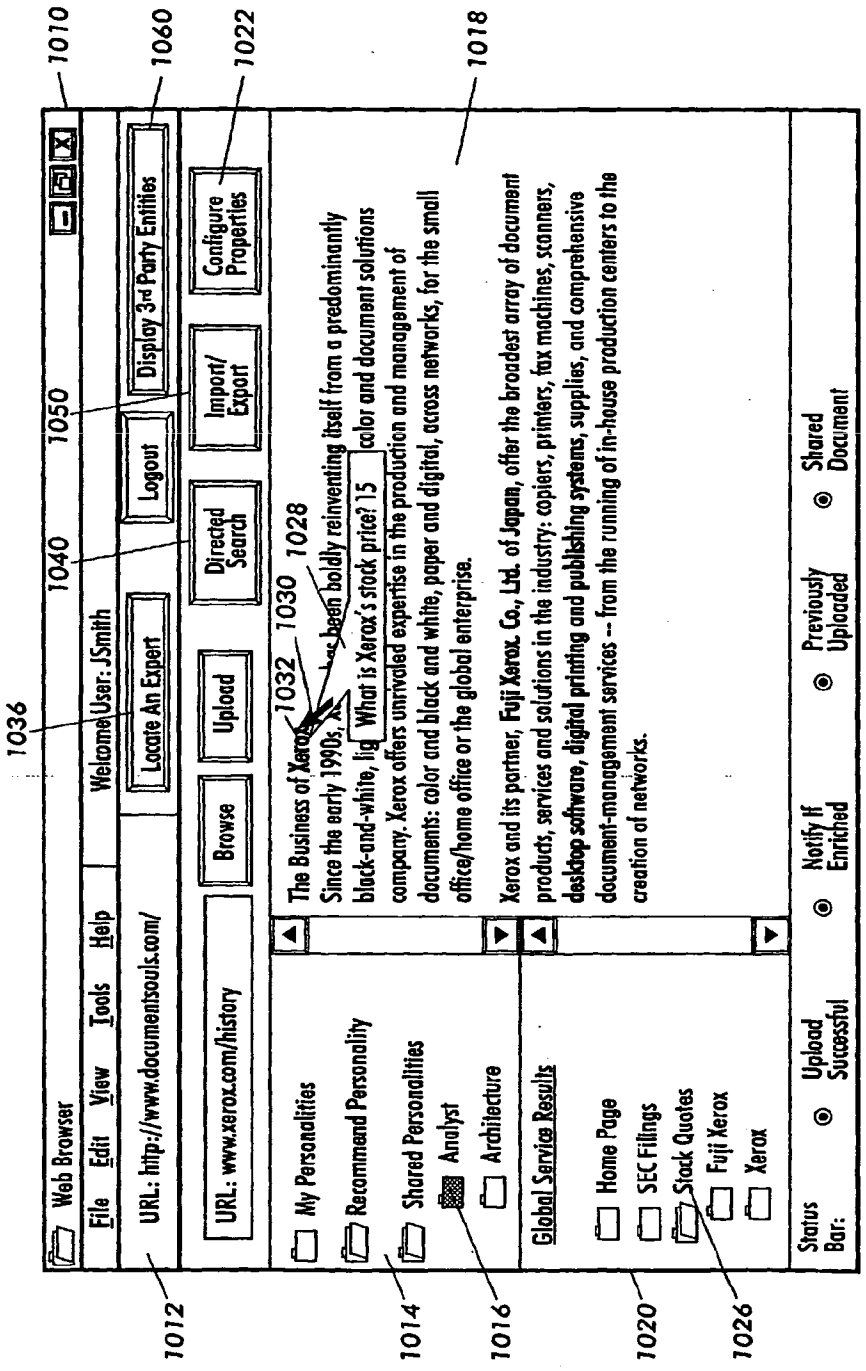


FIG. 10

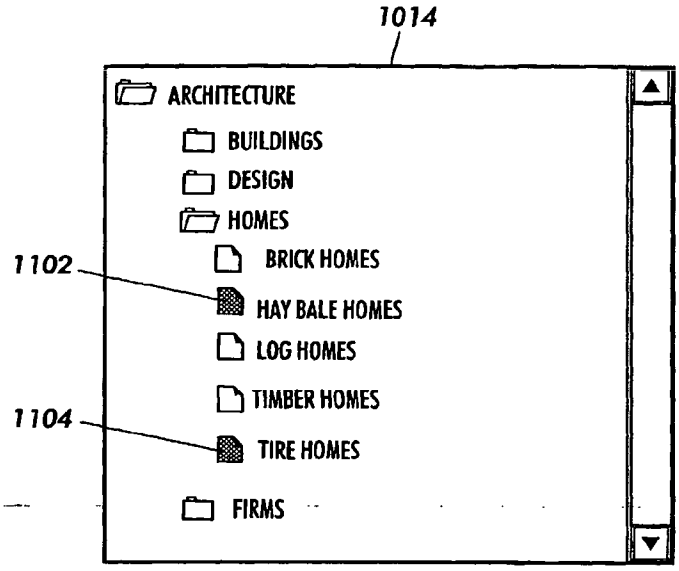

1014
ARCHITECTURE
BUILDINGS
DESIGN
HOMES
BRICK HOMES
1102
HAY BALE HOMES
LOG HOMES
TIMBER HOMES
1104
TIRE HOMES
FIRMS

FIG. 11

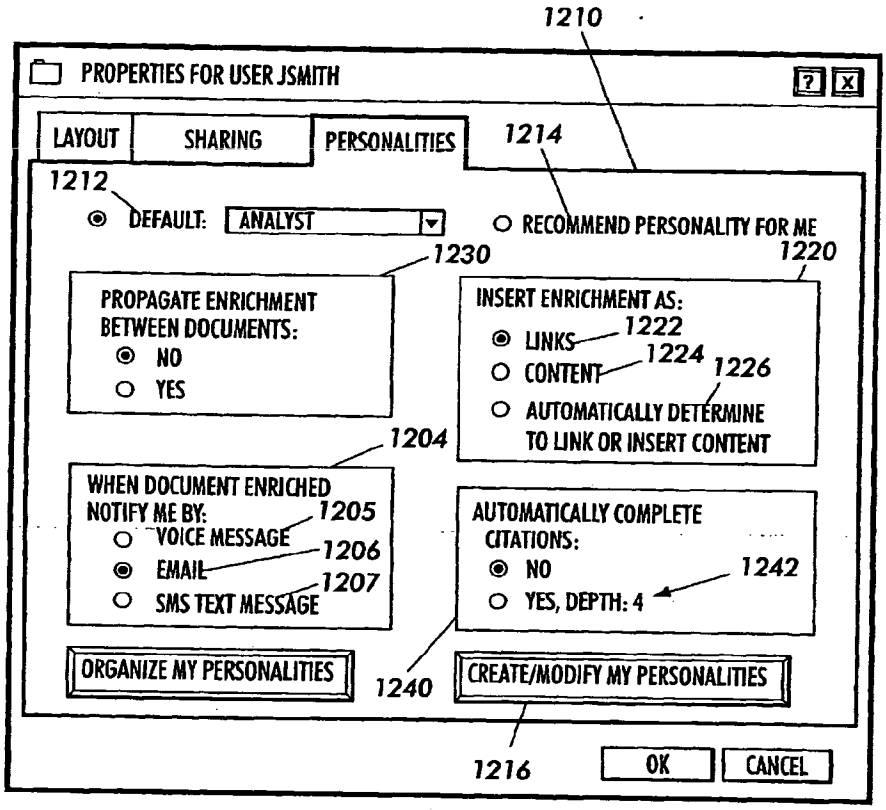


FIG. 12

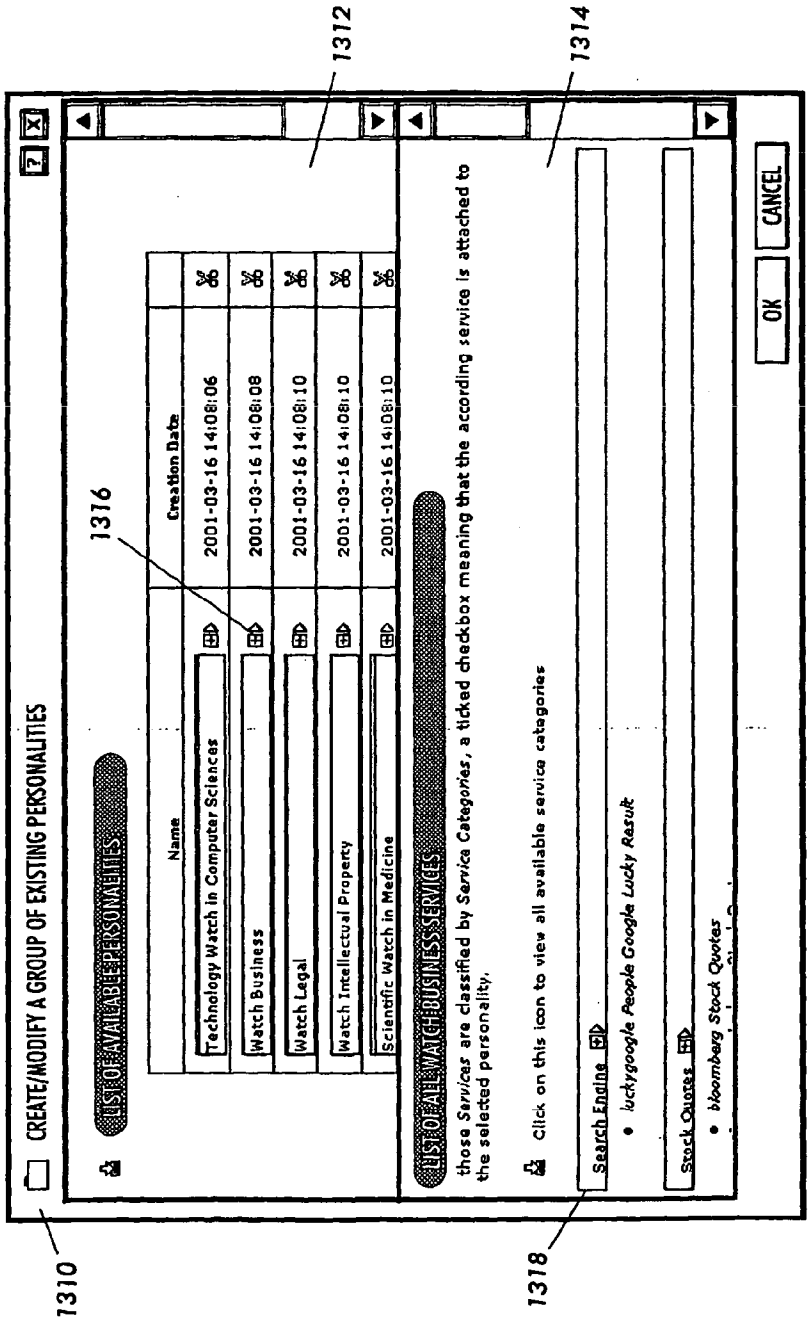


FIG. 13

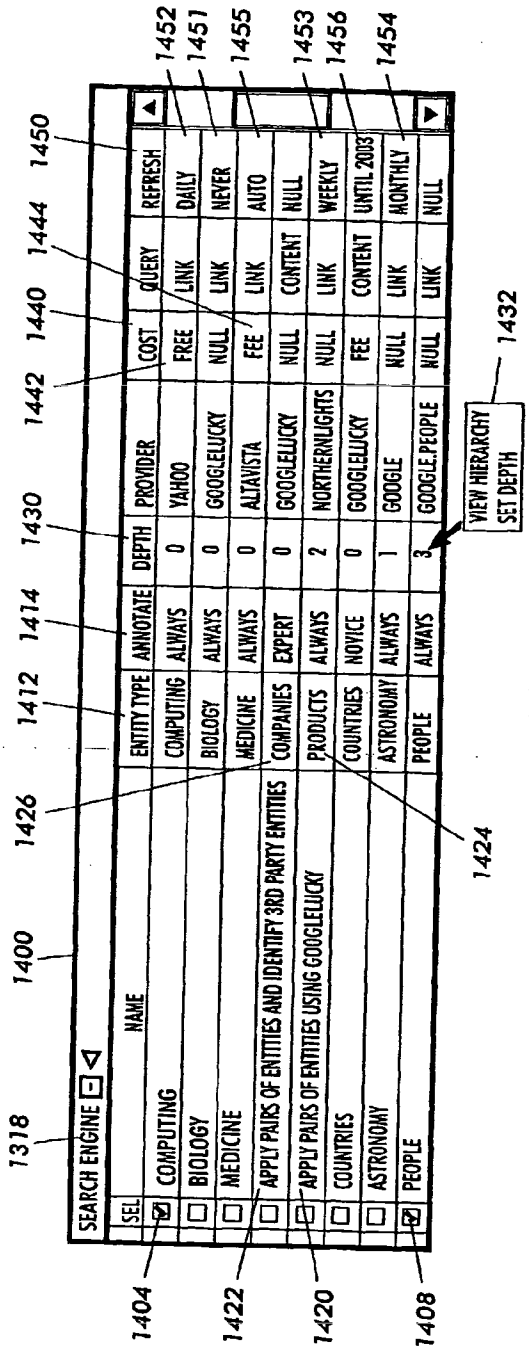

SEARCH ENGINE
SEL
NAME
ENTITY TYPE
ANNOTATE
DEPTH
PROVIDER
COST
QUERY
REFRESH
COMPUTING
ALWAYS
0
YAHOO
FREE
LINK
DAILY
BIOLOGY
ALWAYS
0
GOOGLELUCKY
NULL
LINK
NEVER
MEDICINE
ALWAYS
0
ALTAVISTA
FEE
LINK
AUTO
APPLY PAIRS OF ENTITIES AND IDENTIFY 3RD PARTY ENTITIES
COMPANIES
EXPERT
0
GOOGLELUCKY
NULL
CONTENT
NULL
APPLY PAIRS OF ENTITIES USING GOOGLELUCKY
PRODUCTS
ALWAYS
2
NORTHERNLIGHTS
NULL
LINK
WEEKLY
COUNTRIES
NOVICE
0
GOOGLELUCKY
FEE
CONTENT
UNTIL 2003
ASTRONOMY
ALWAYS
1
GOOGLE
NULL
LINK
MONTHLY
PEOPLE
ALWAYS
3
GOOGLE.PEOPLE
NULL
LINK
NULL
VIEW HIERARCHY
SET DEPTH
1318
1400
1426
1412
1414
1430
1442
1440
1444
1450
1452
1451
1455
1453
1456
1454
1432
1424
1404
1422
1420
1408

FIG. 14

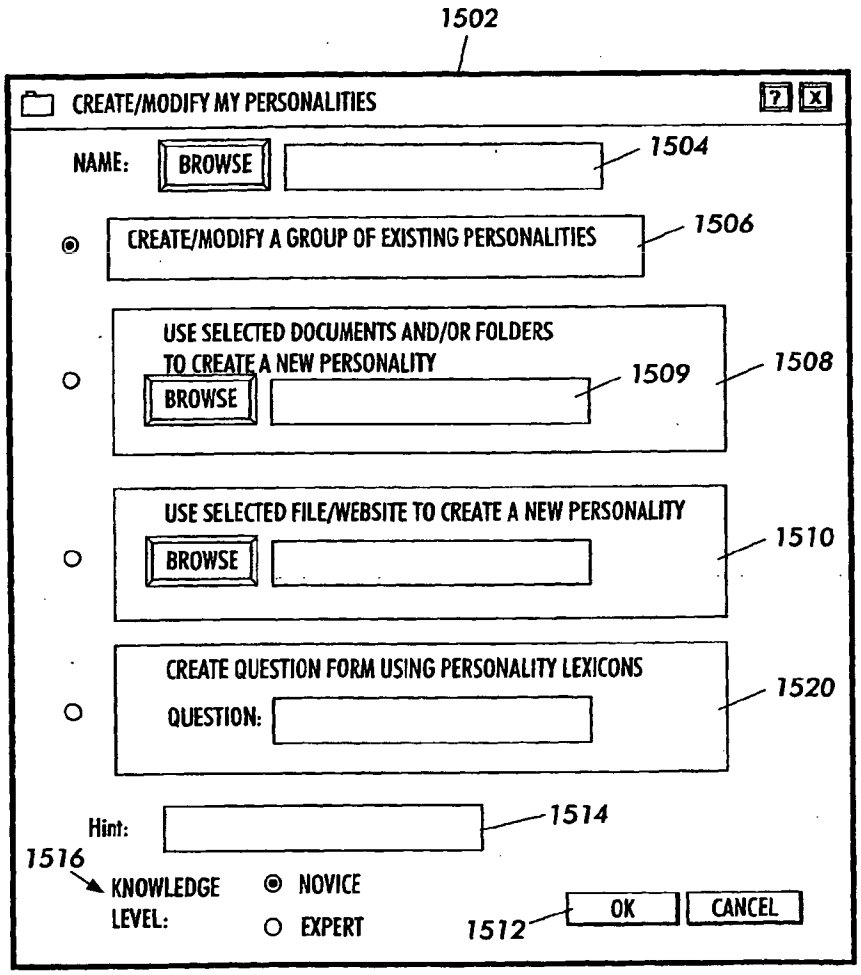


FIG. 15

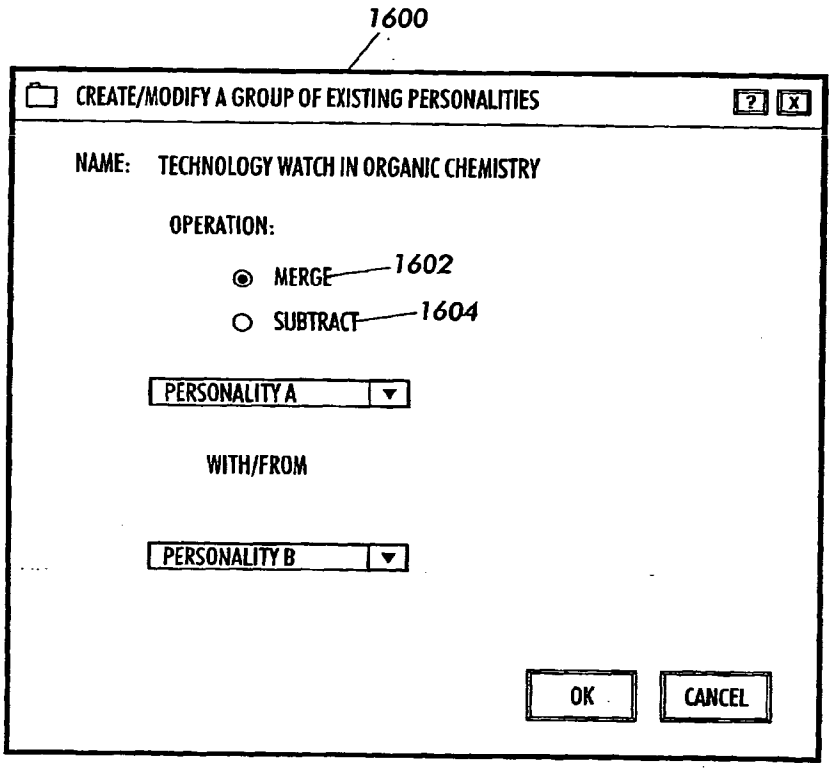


FIG. 16

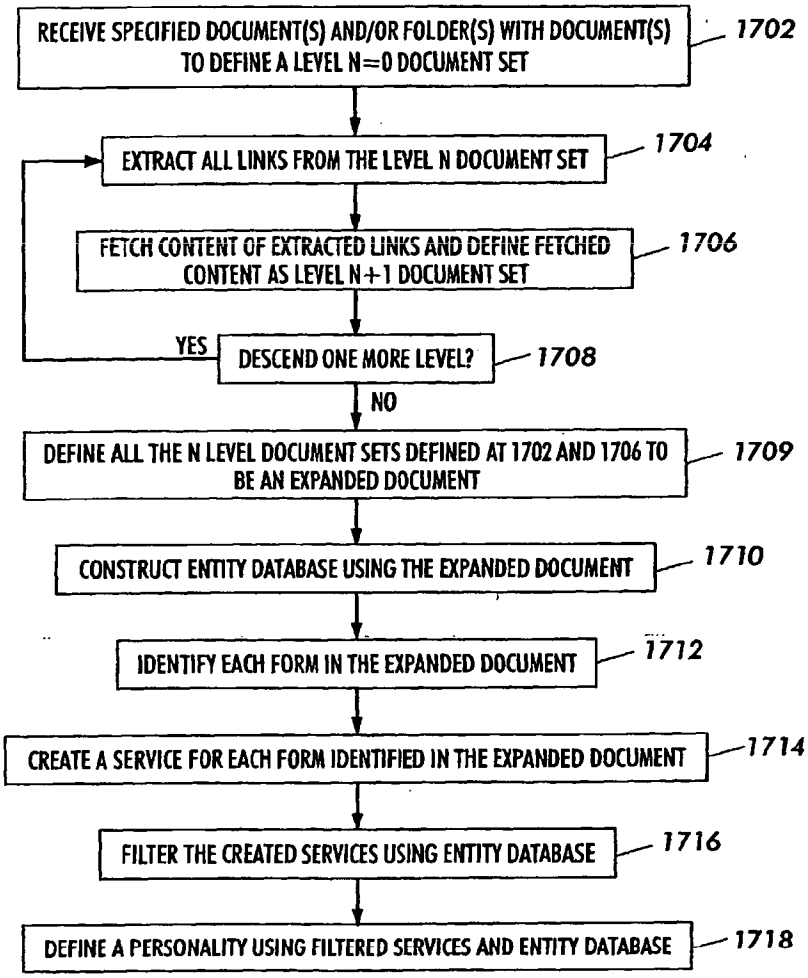


FIG. 17

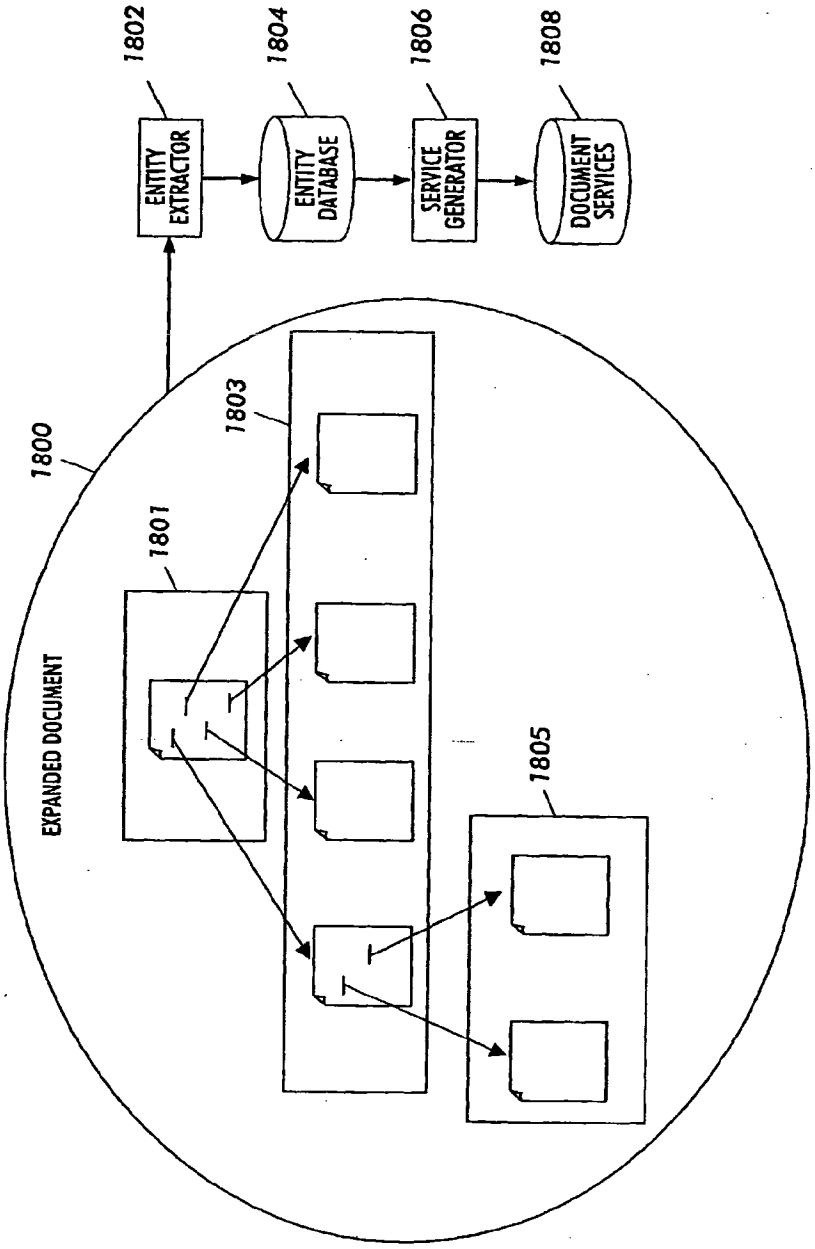


FIG. 18

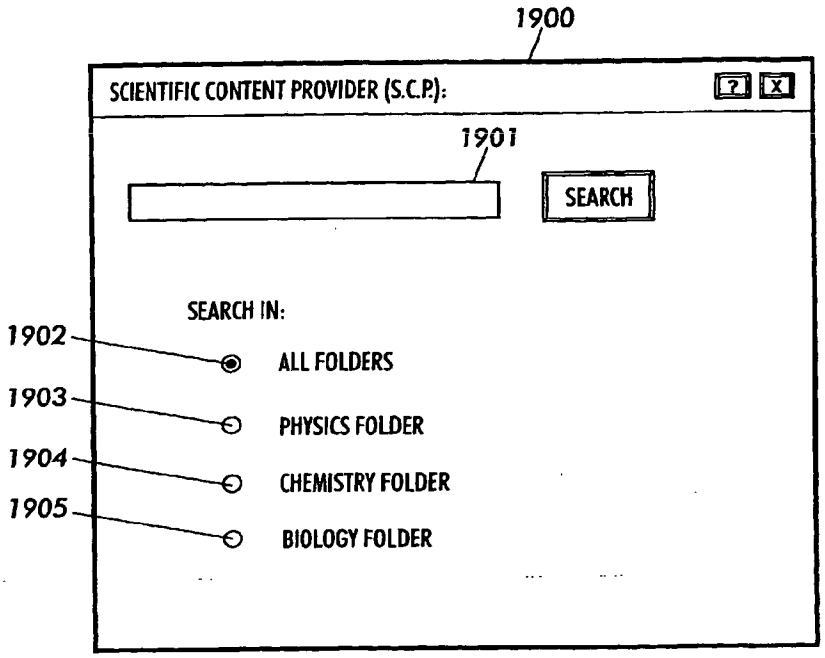


FIG. 19

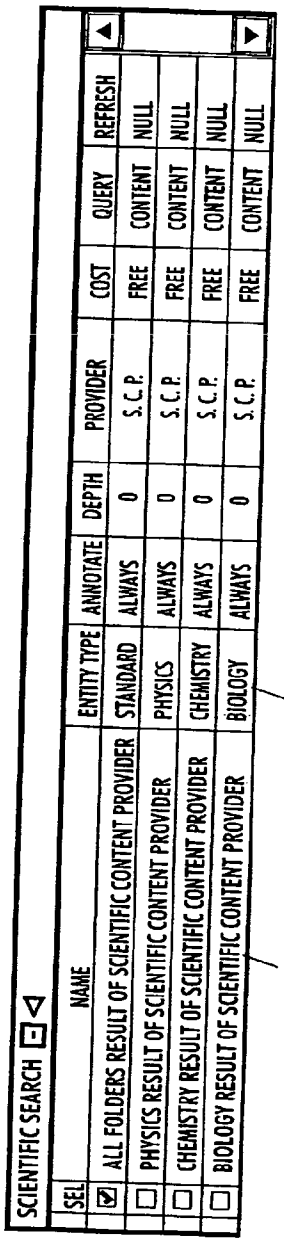


FIG. 20

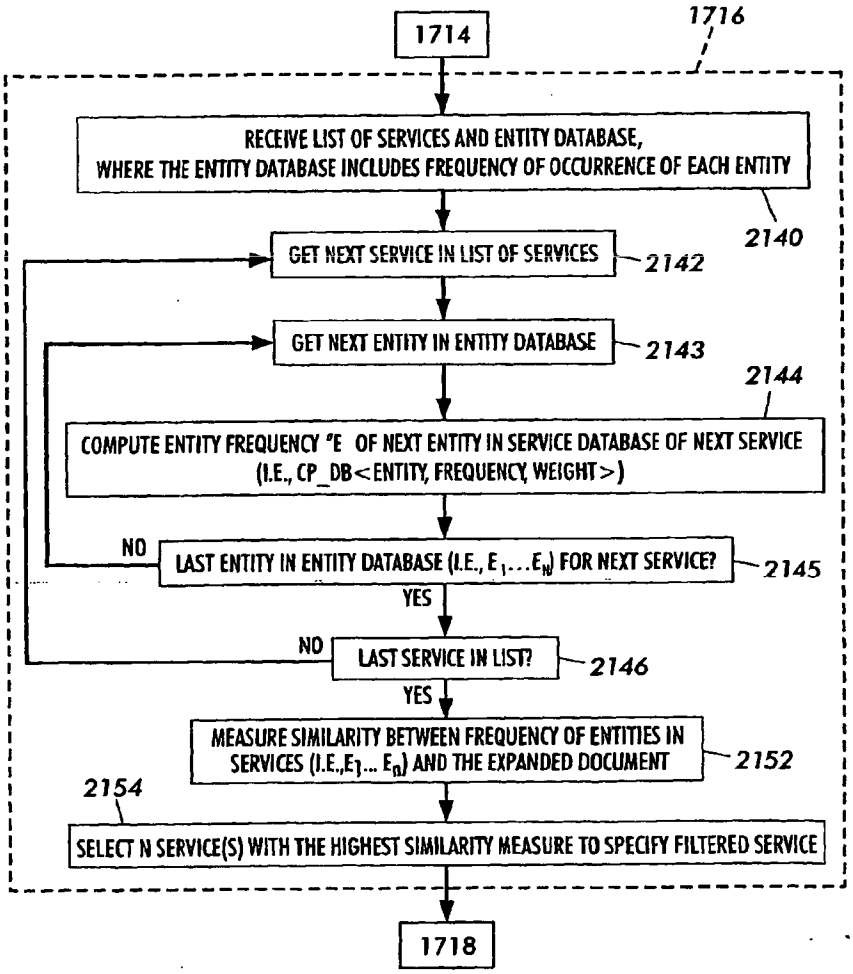


FIG. 21

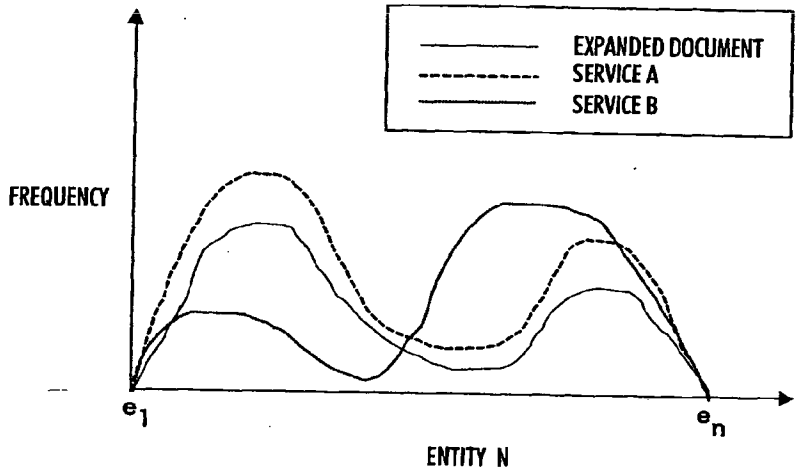


FIG. 22

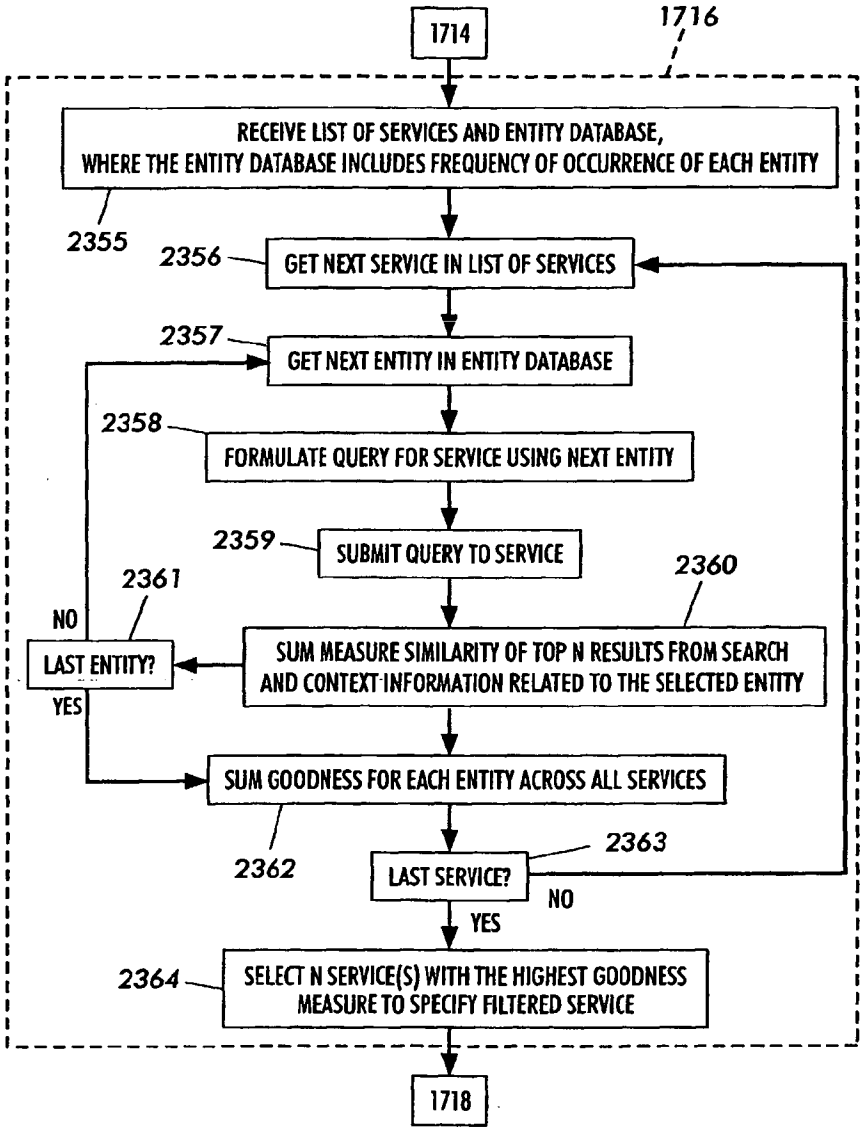


FIG. 23

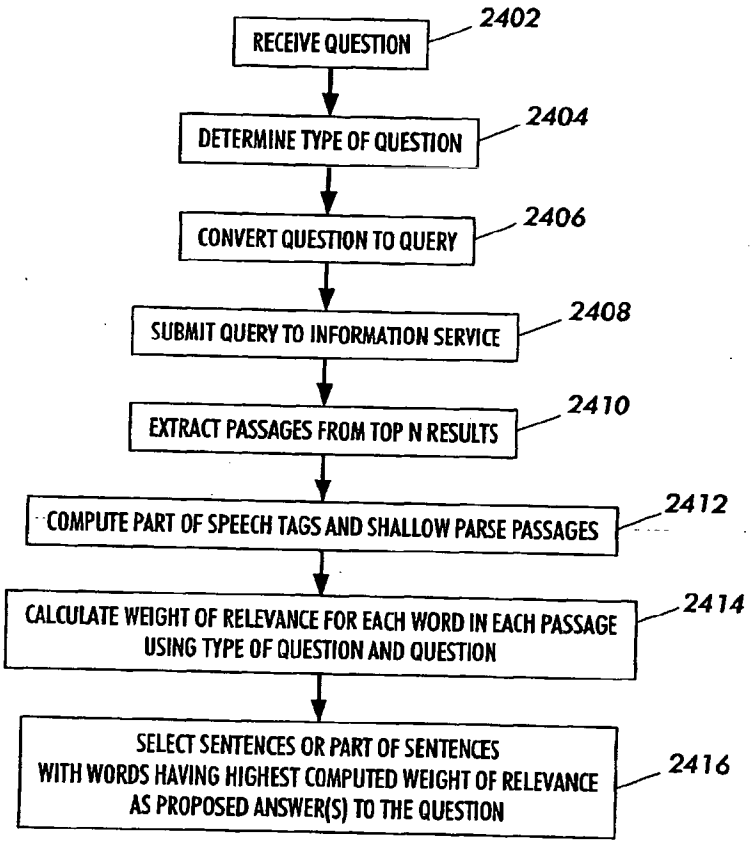


FIG. 24

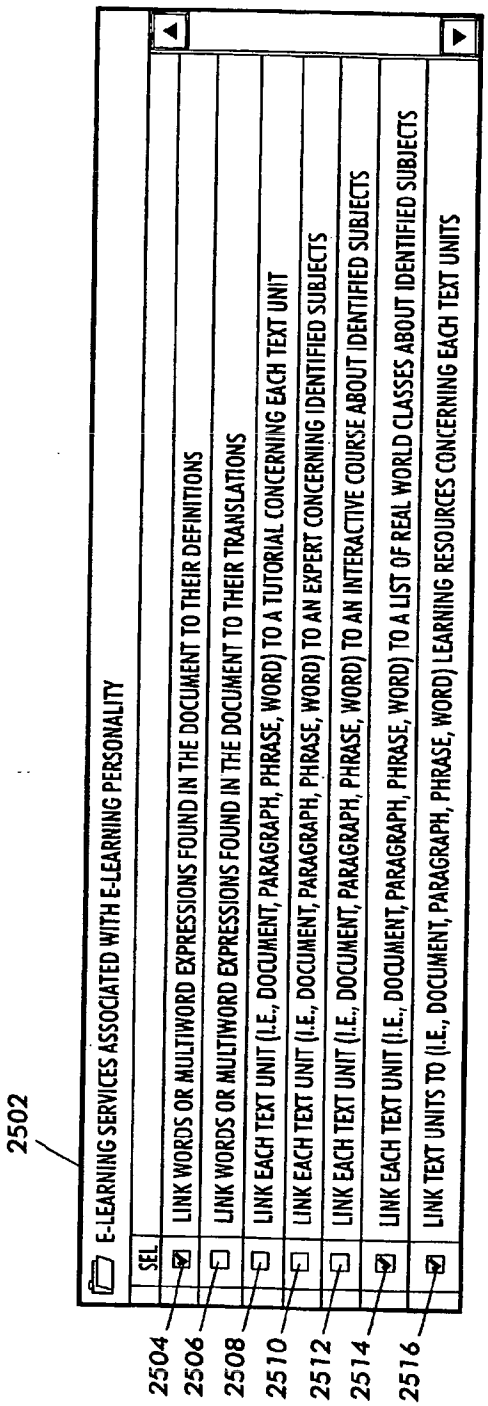


FIG. 25

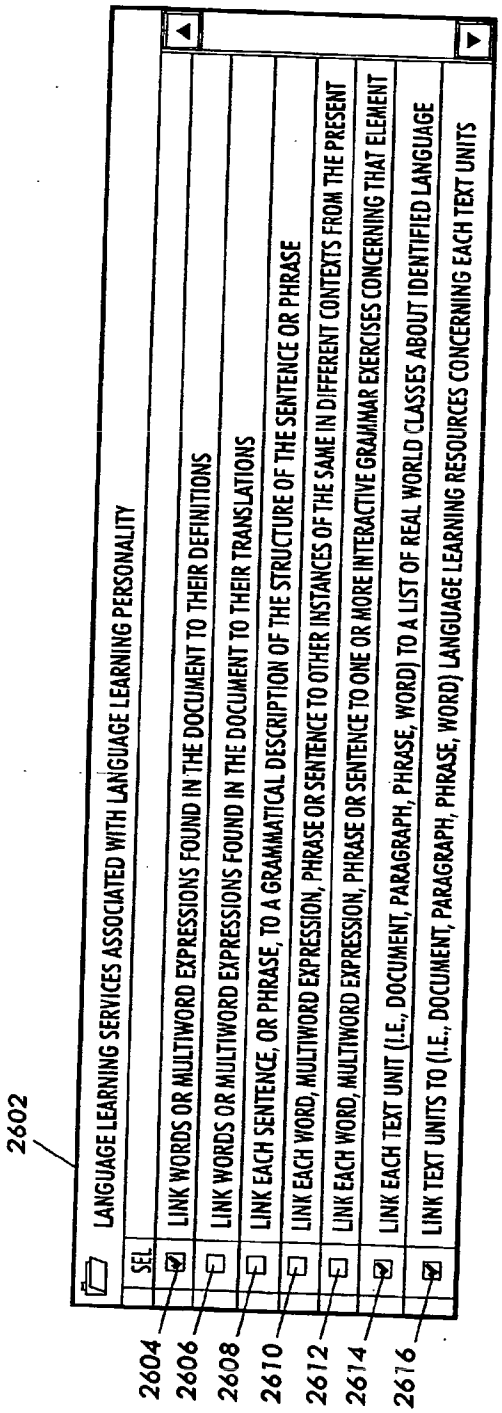


FIG. 26

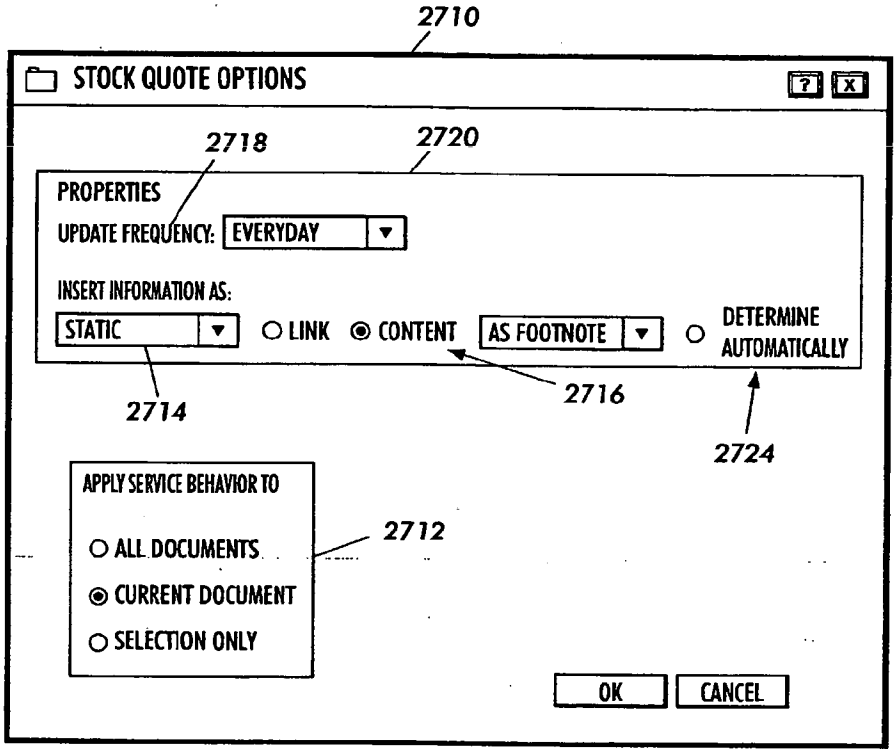


FIG. 27

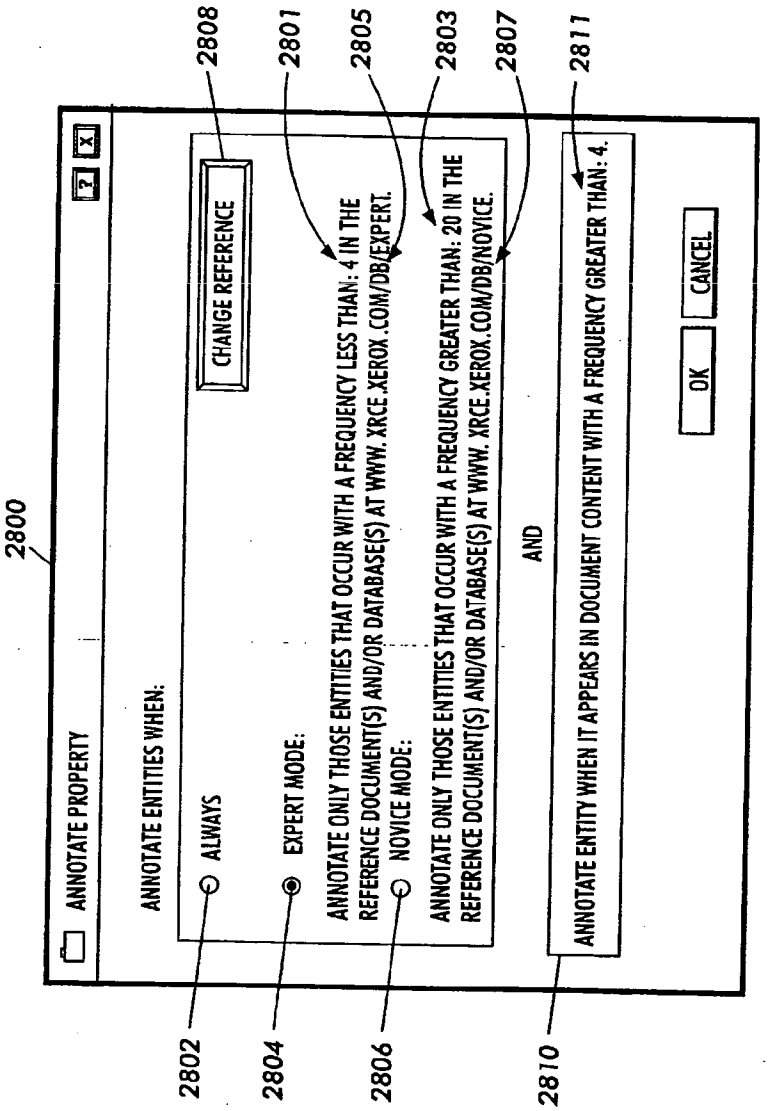


FIG. 28

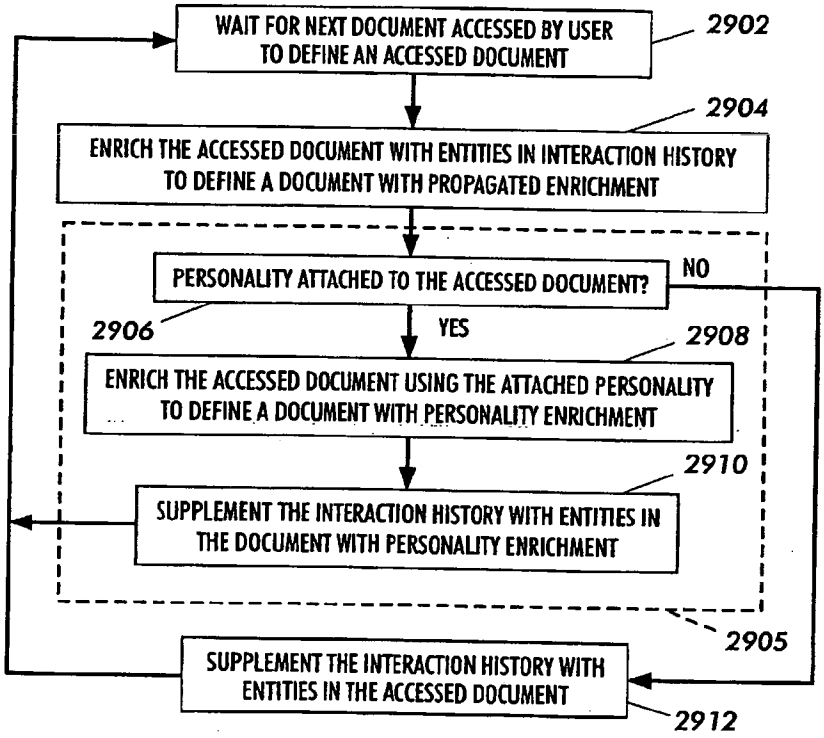


FIG. 29

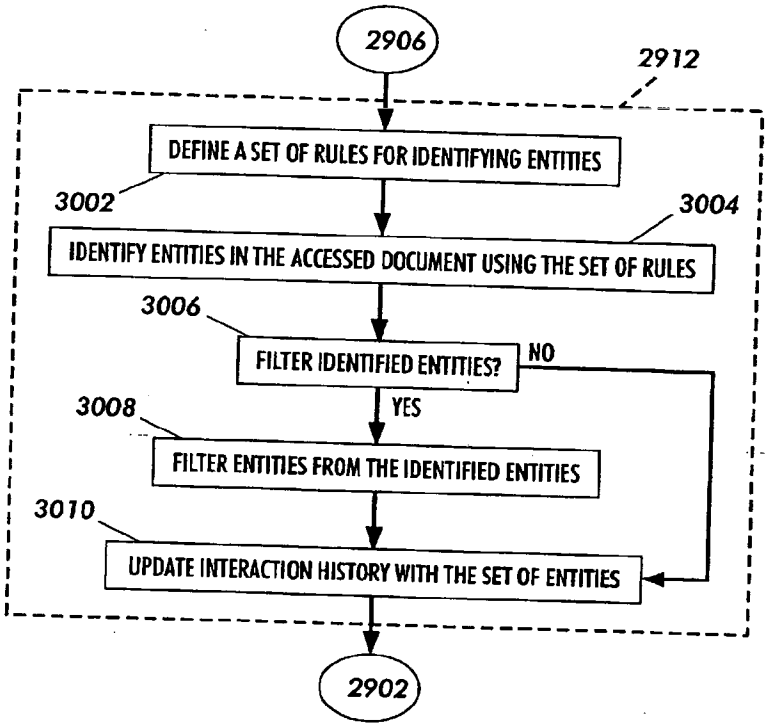


FIG. 30

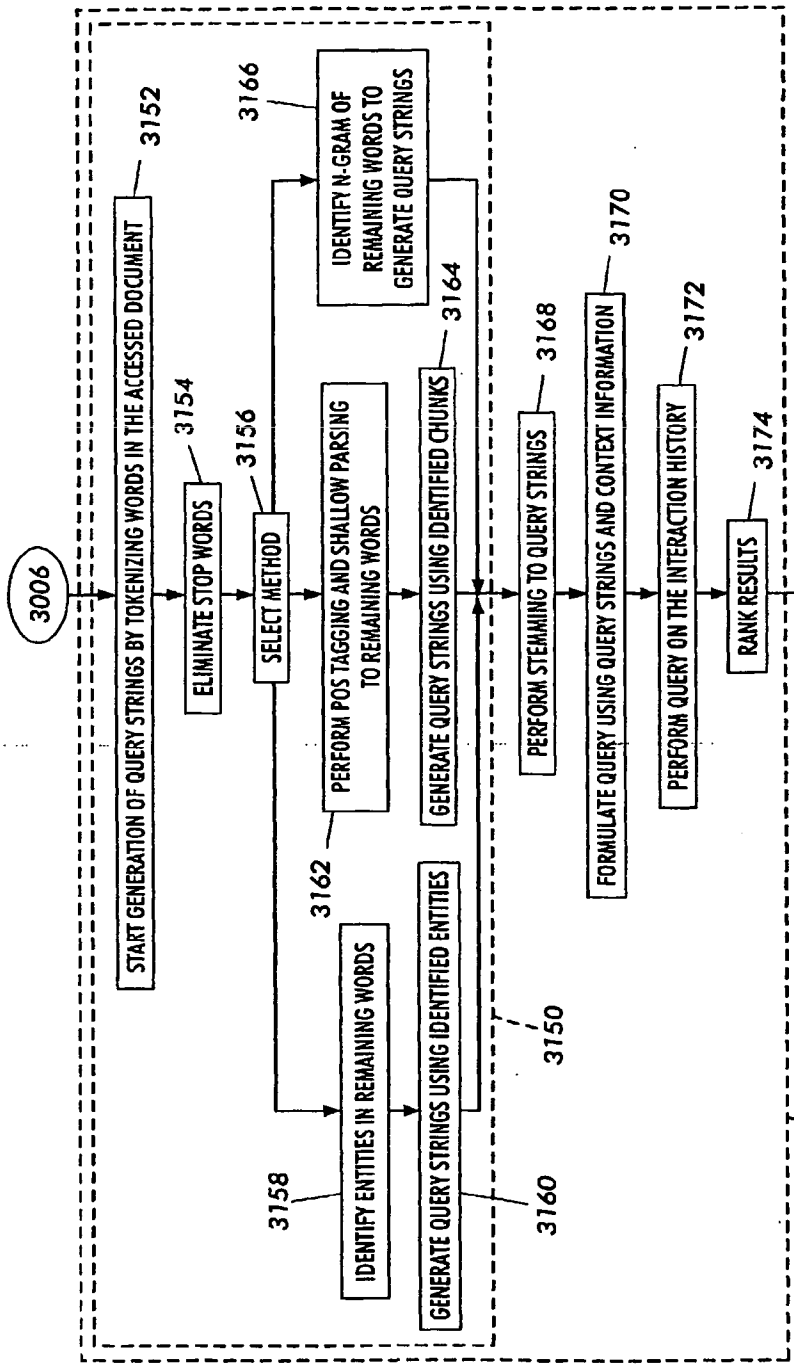


FIG. 31

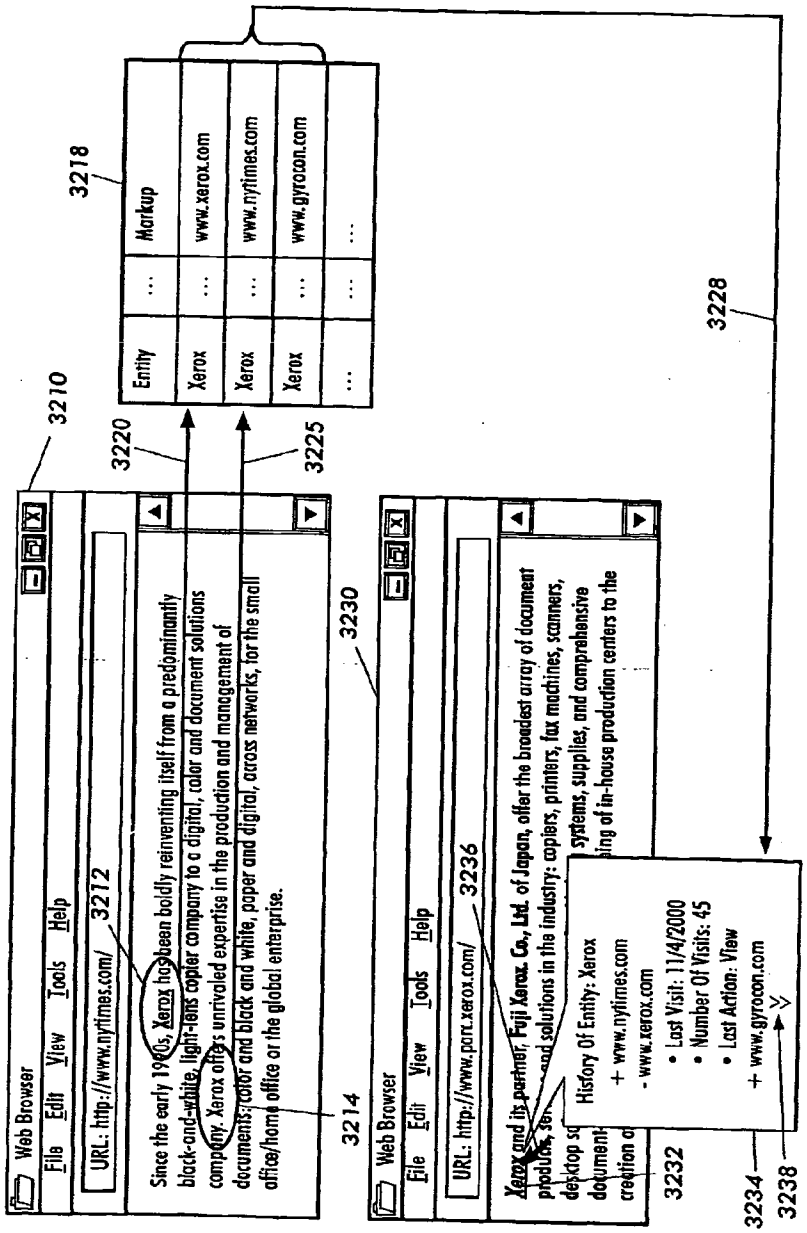


FIG. 32

3218

ENTITY	ENTITY TYPE	VISITS	LAST VISIT	ACTION LIST	DMOZ CATEGORY	LOCATION	POS	MARKUP
XEROX	BUSINESS	45	11/4/2000	VIEW (43), PRINT(2)	BUSINESS	40	NOUN	WWW.XEROX.COM
XEROX	BUSINESS	1	5/1/2001	VIEW(1)	FINANCE	689	NOUN	WWW.NYTIMES.COM
XEROX	BUSINESS	4	2/22/2001	EMAIL(1), VIEW(3)	COMPUTER	343	NOUN	WWW.GYROCON.COM
...	...	...	...	...	...	...	...	...

FIG. 33

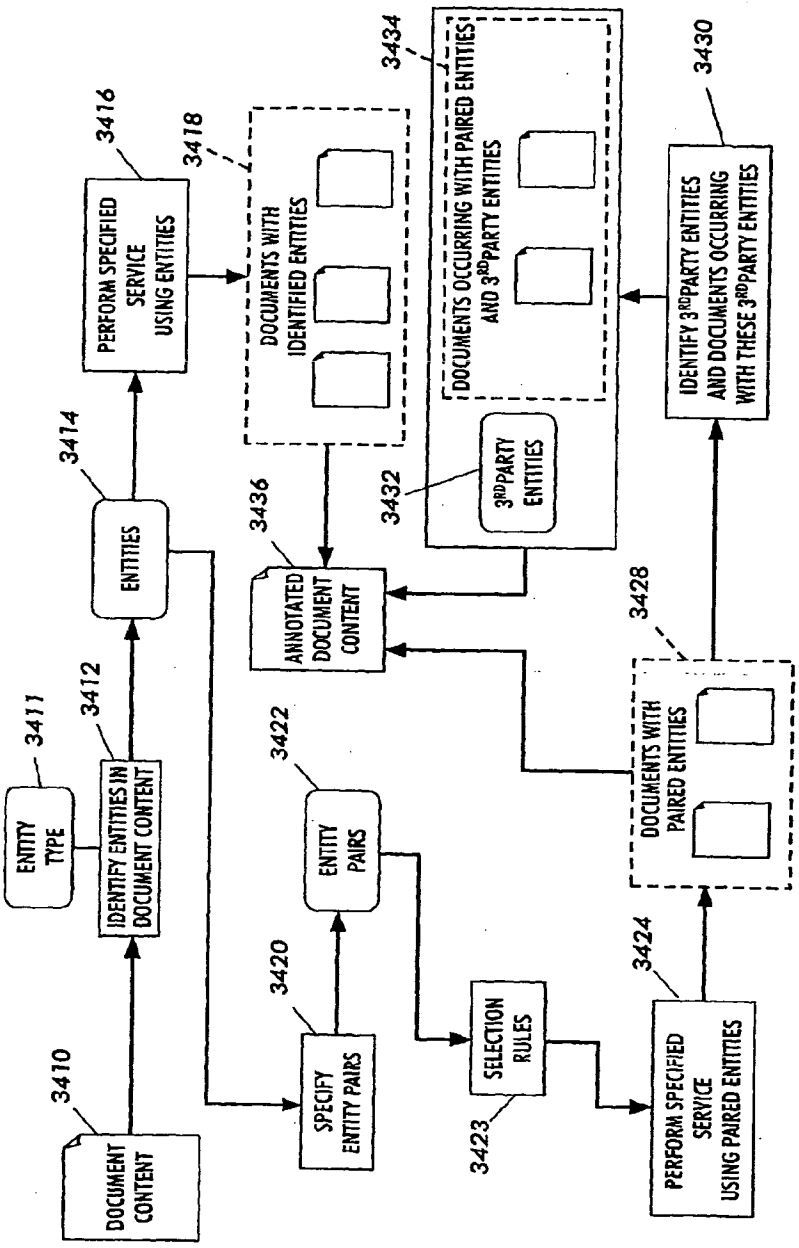


FIG. 34

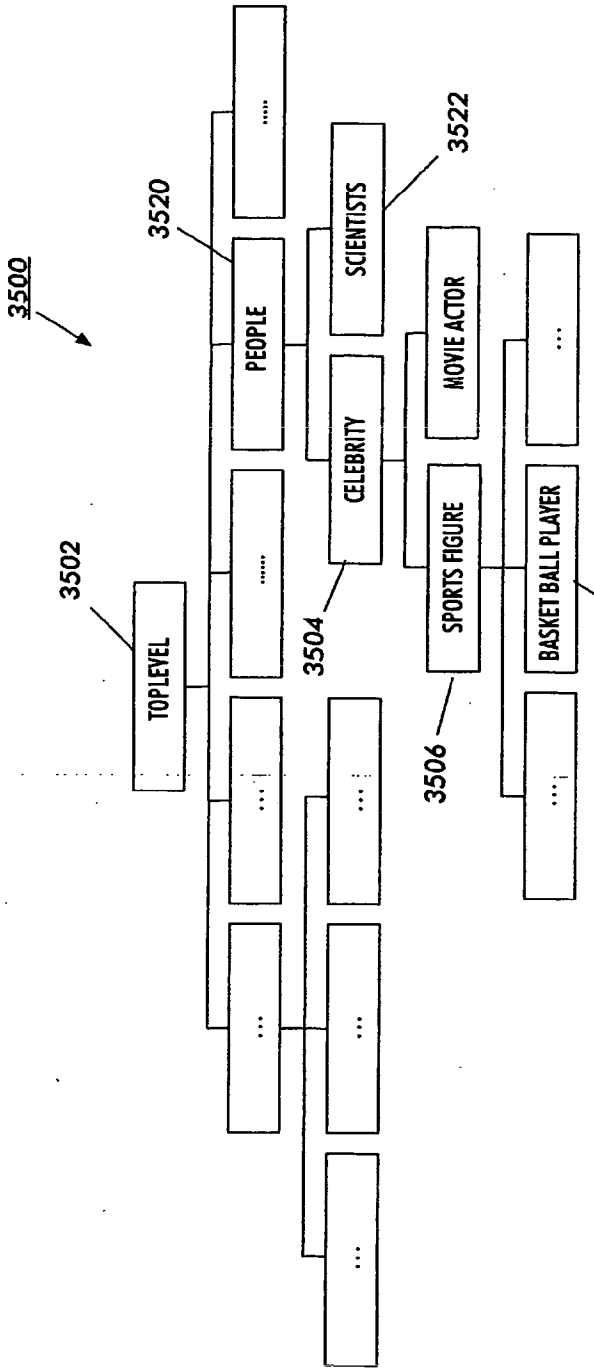


FIG. 35

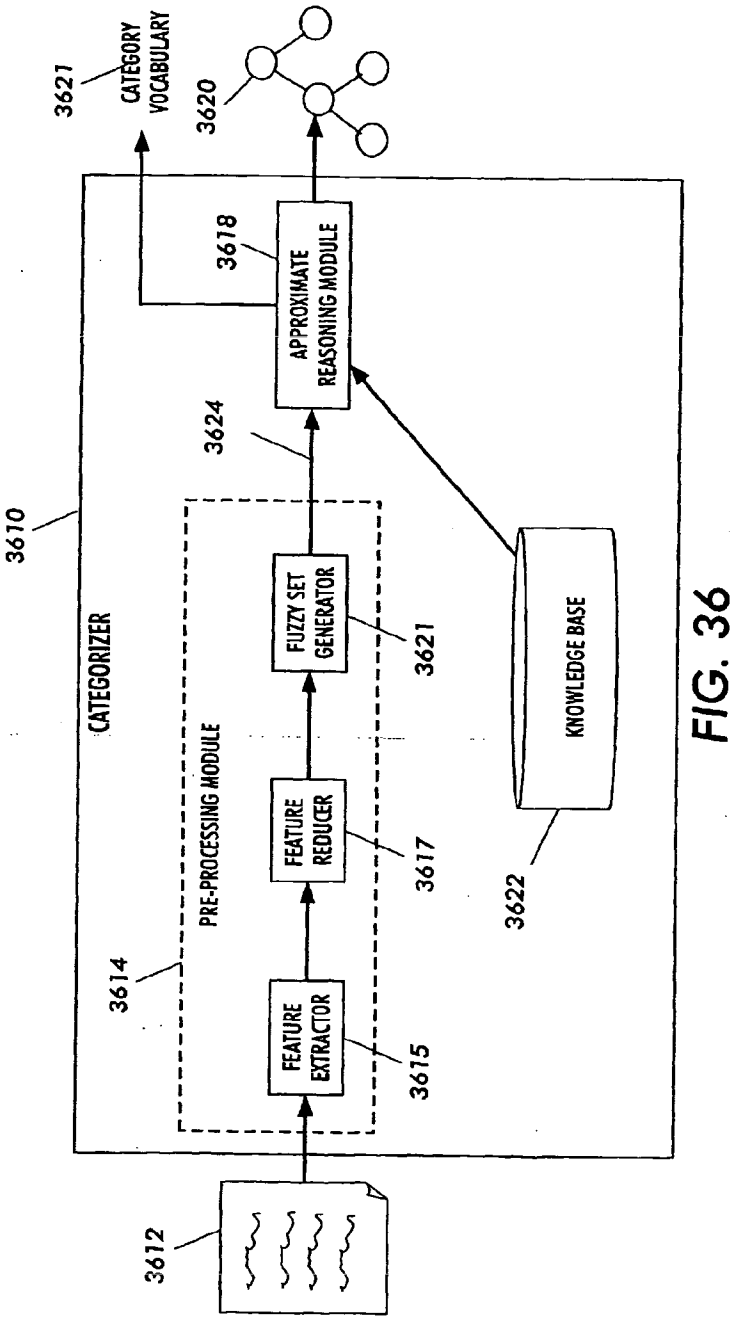

3621
CATEGORY VOCABULARY
3620
3618
APPROXIMATE REASONING MODULE
3624
3610
CATEGORIZER
FUZZY SET GENERATOR
3621
KNOWLEDGE BASE
PRE-PROCESSING MODULE
FEATURE REDUCER
3617
3622
3614
FEATURE EXTRACTOR
3615
3612

FIG. 36

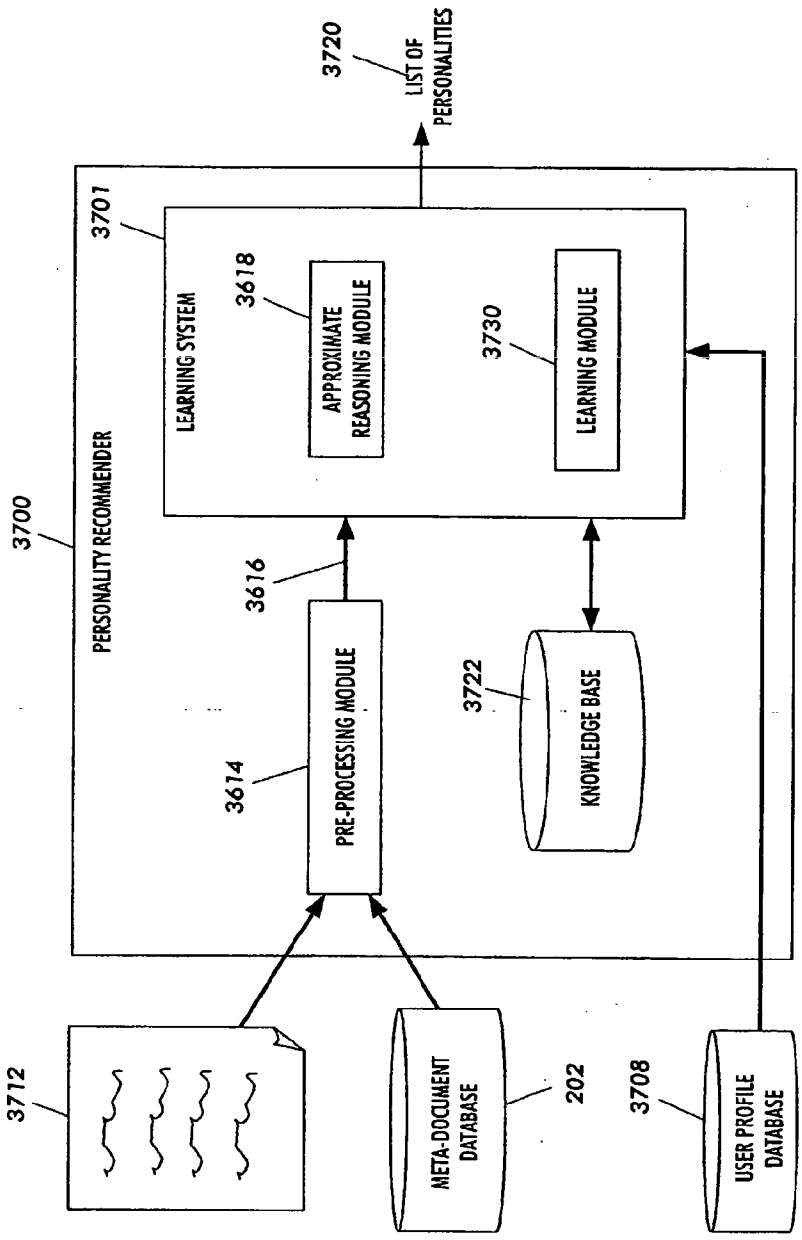


FIG. 37

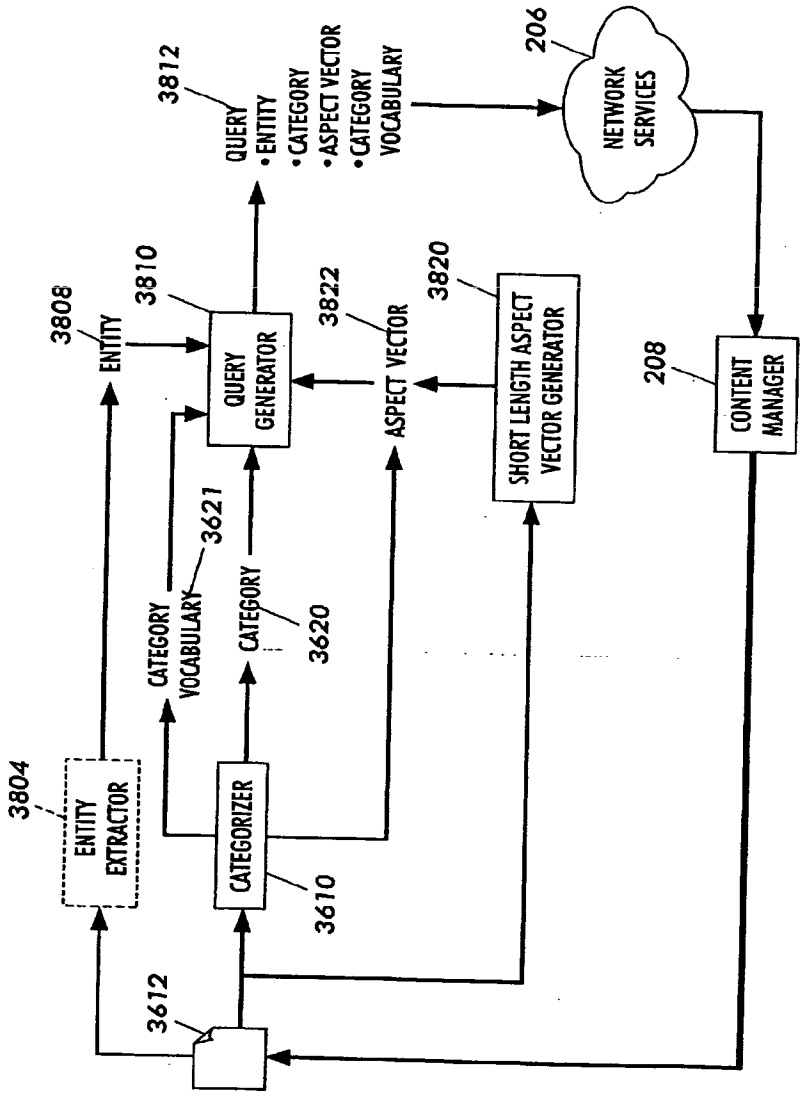


FIG. 38

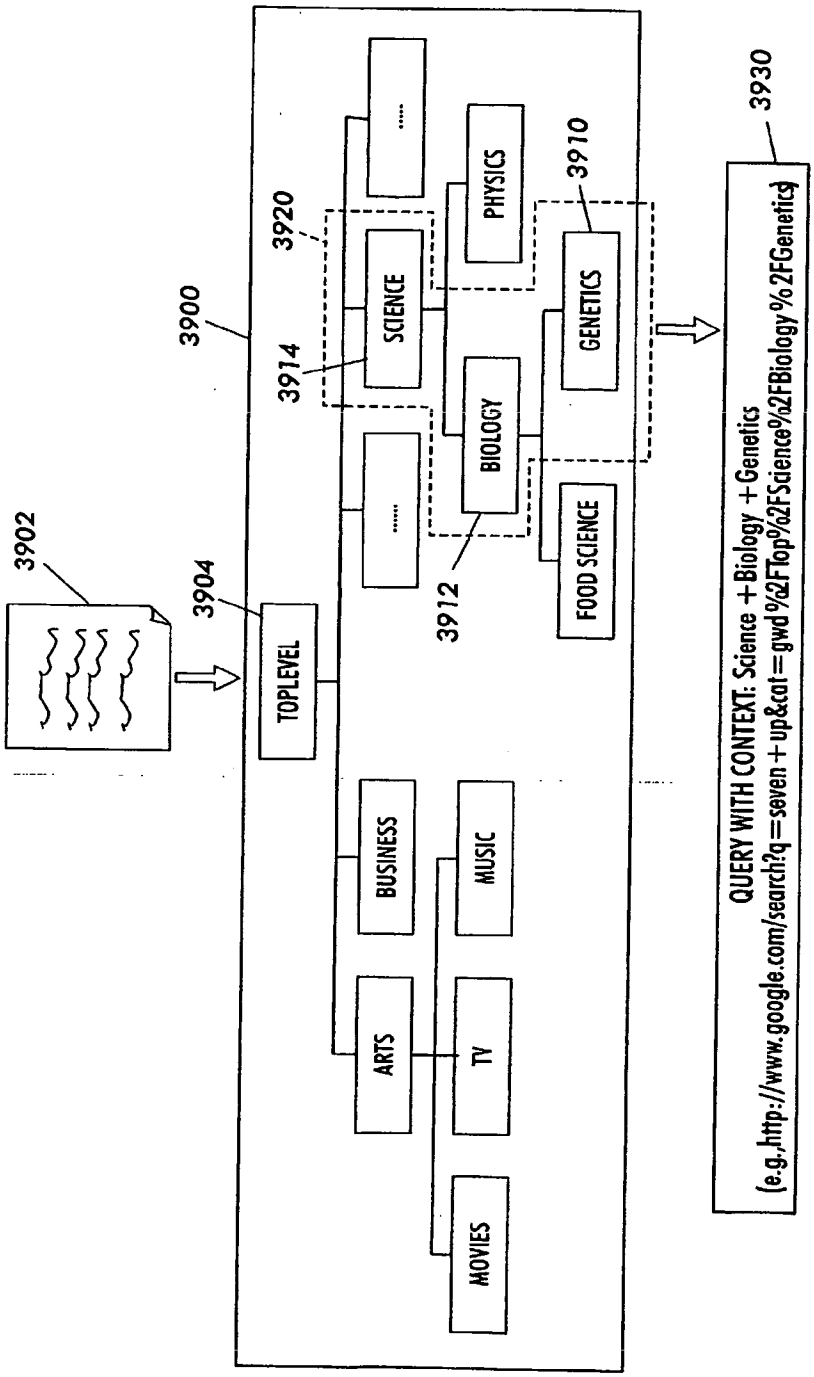


FIG. 39

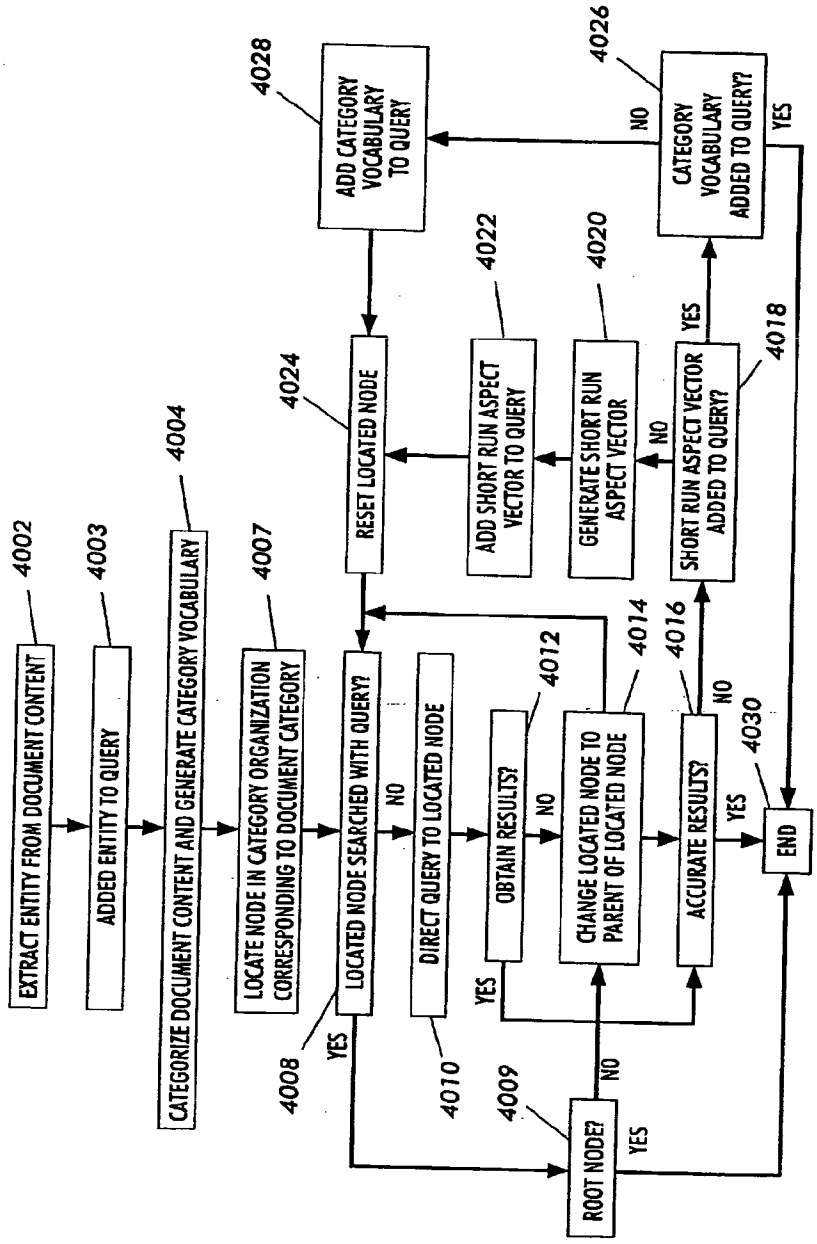


FIG. 40

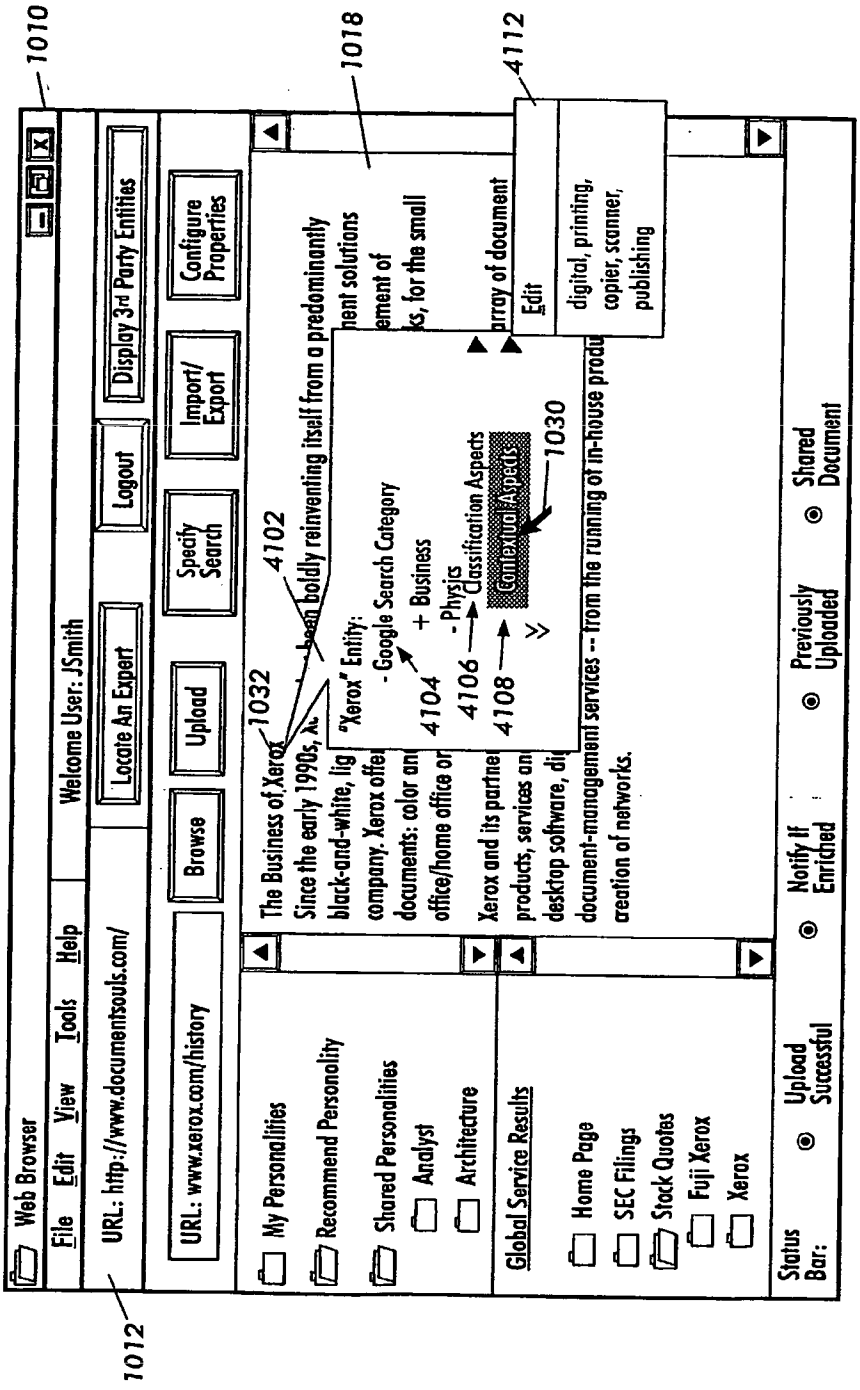


FIG. 41

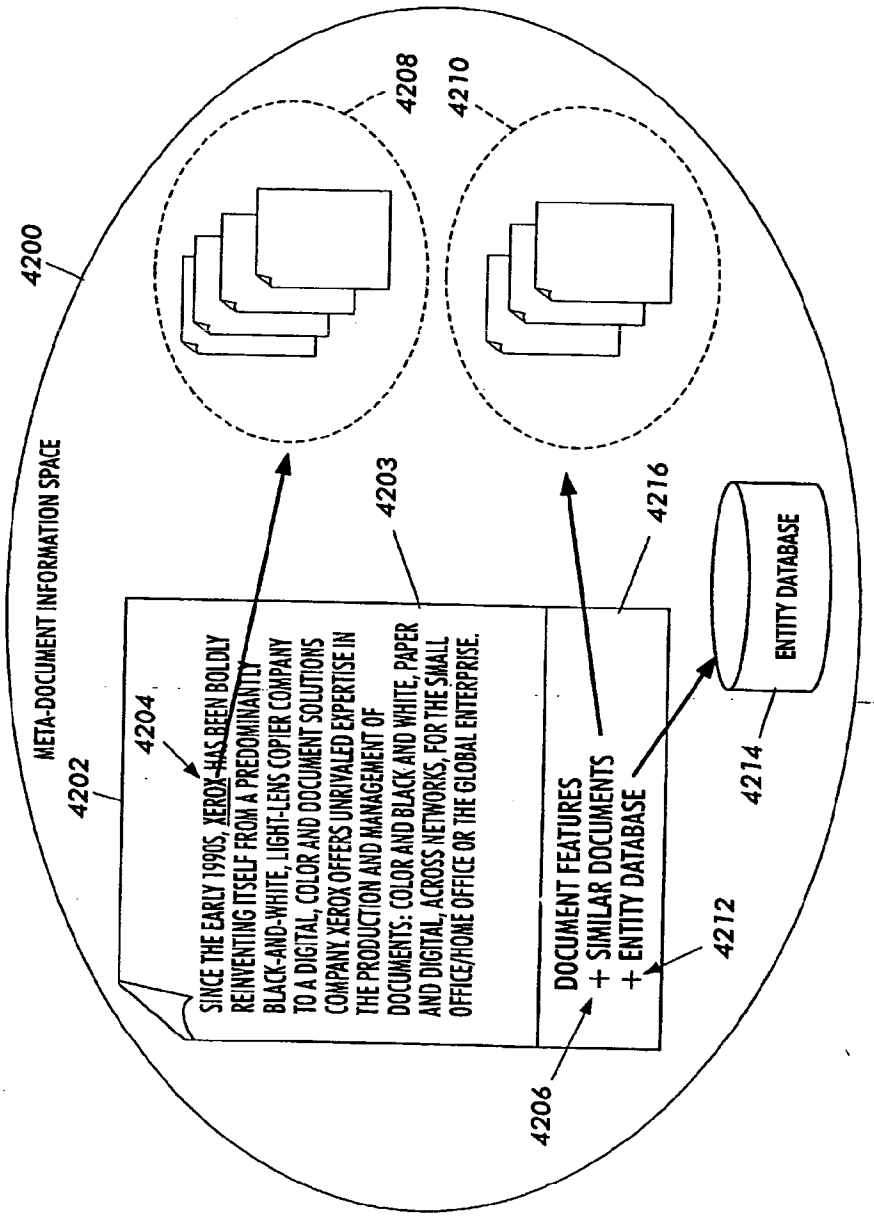


FIG. 42

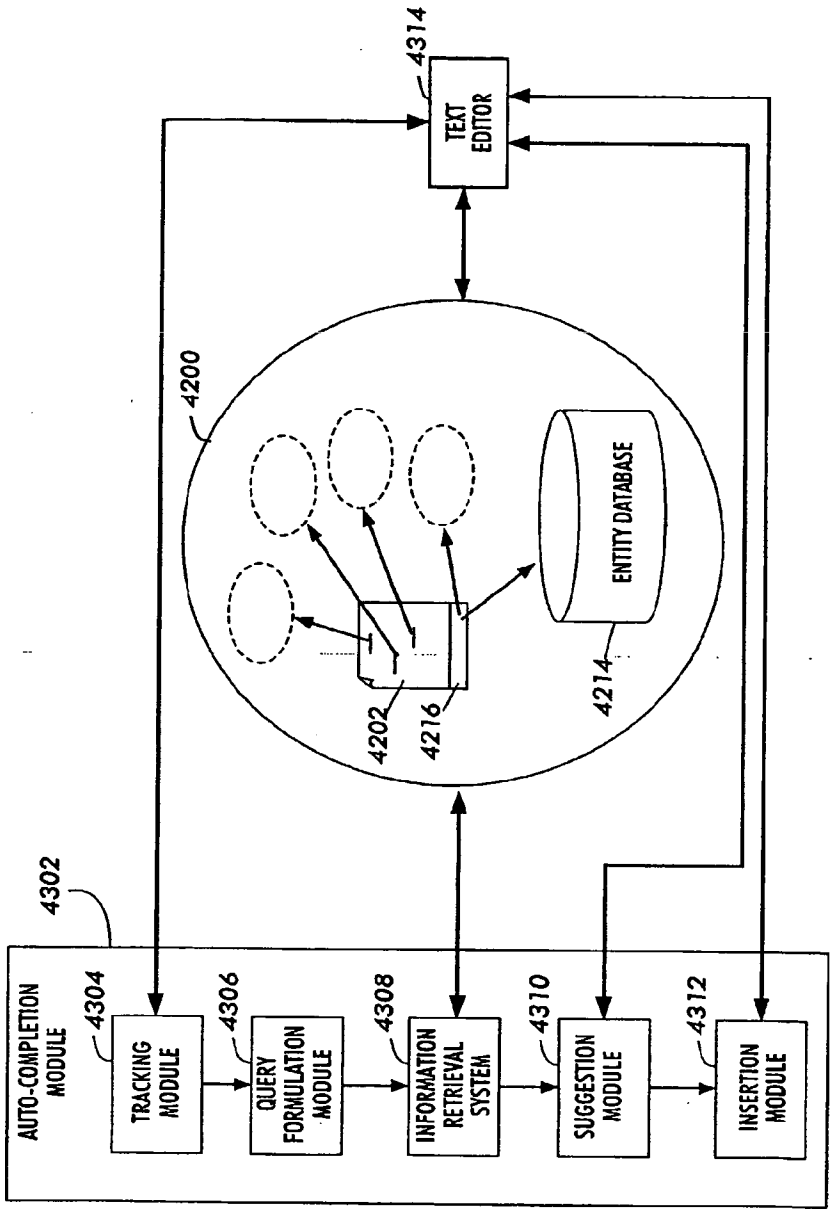

4314
TEXT EDITOR
4200
ENTITY DATABASE
4202
4216
4214
4302
AUTO-COMPLETION MODULE
4304
TRACKING MODULE
4306
QUERY FORMULATION MODULE
4308
INFORMATION RETRIEVAL SYSTEM
4310
SUGGESTION MODULE
4312
INSERTION MODULE

FIG. 43

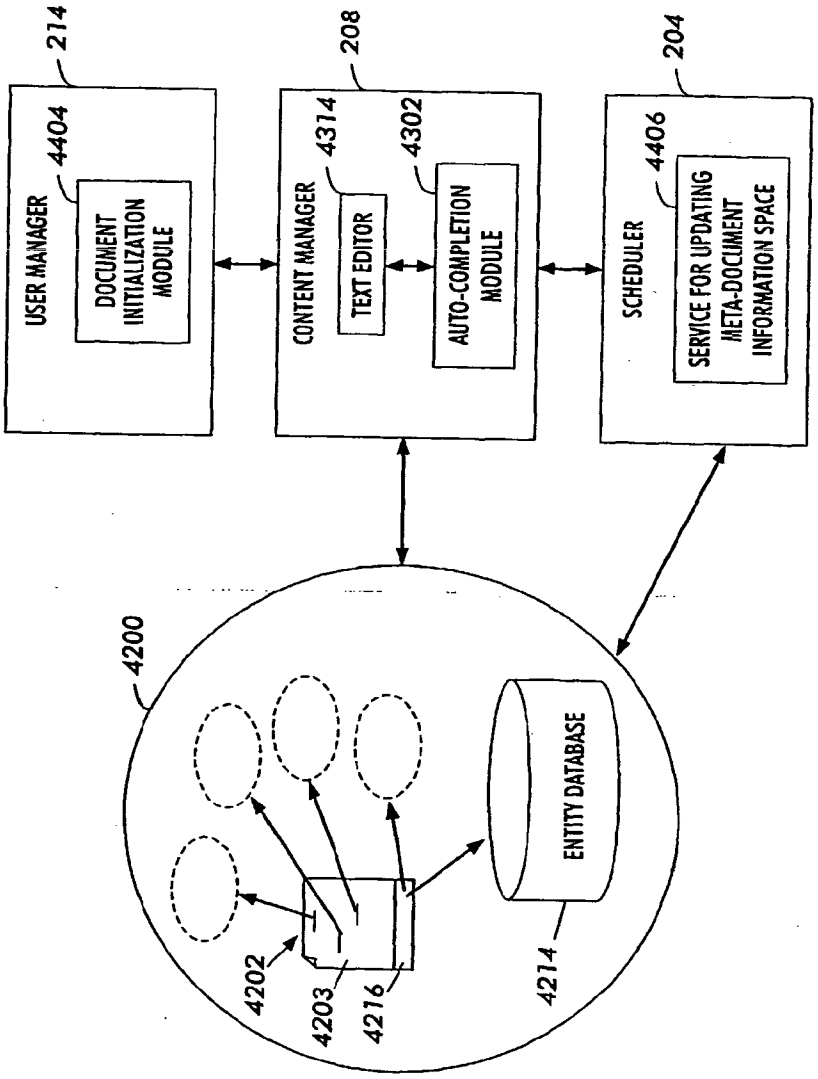


FIG. 44

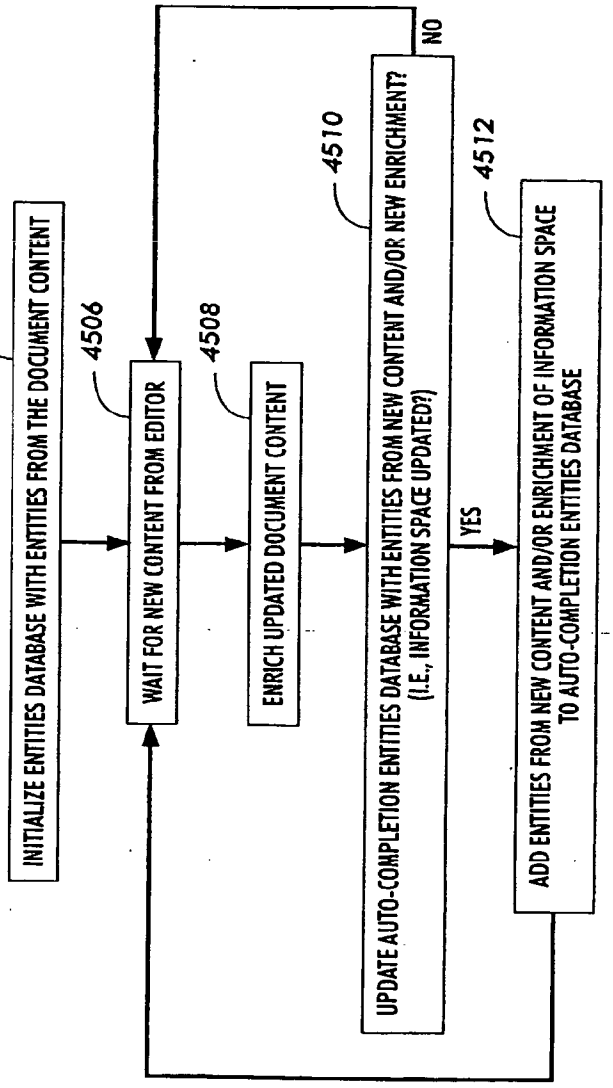


FIG. 45

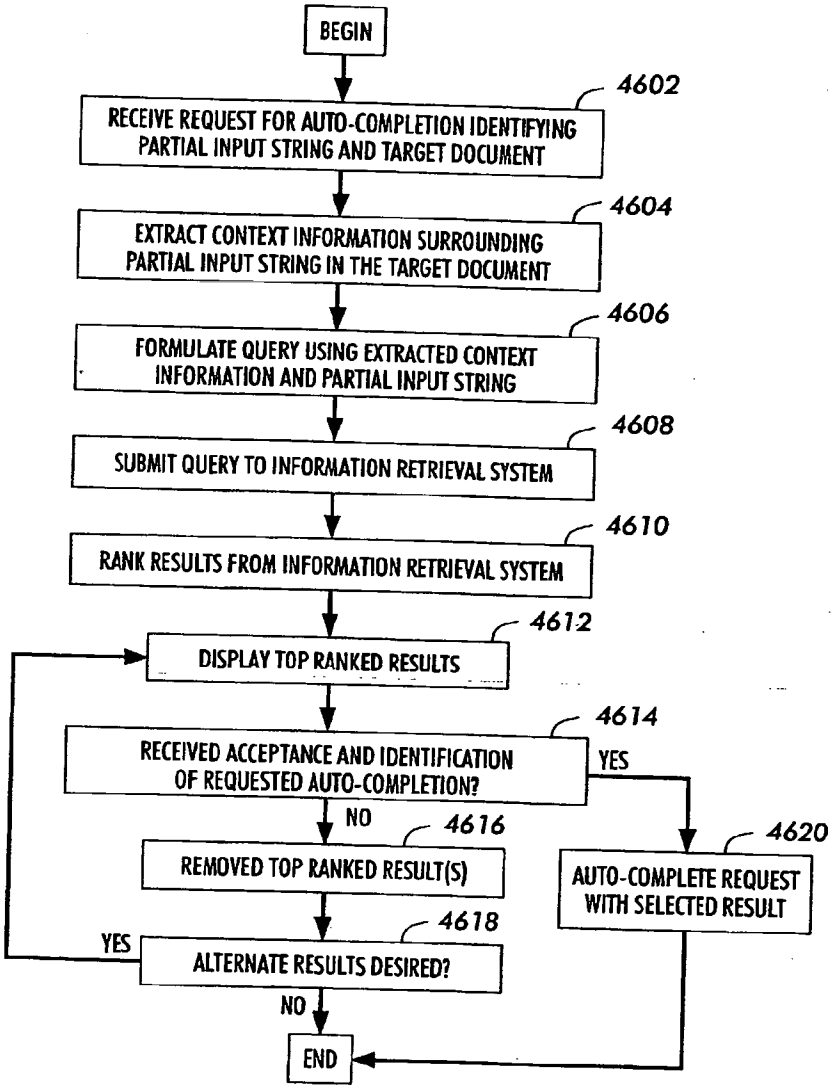


FIG. 46

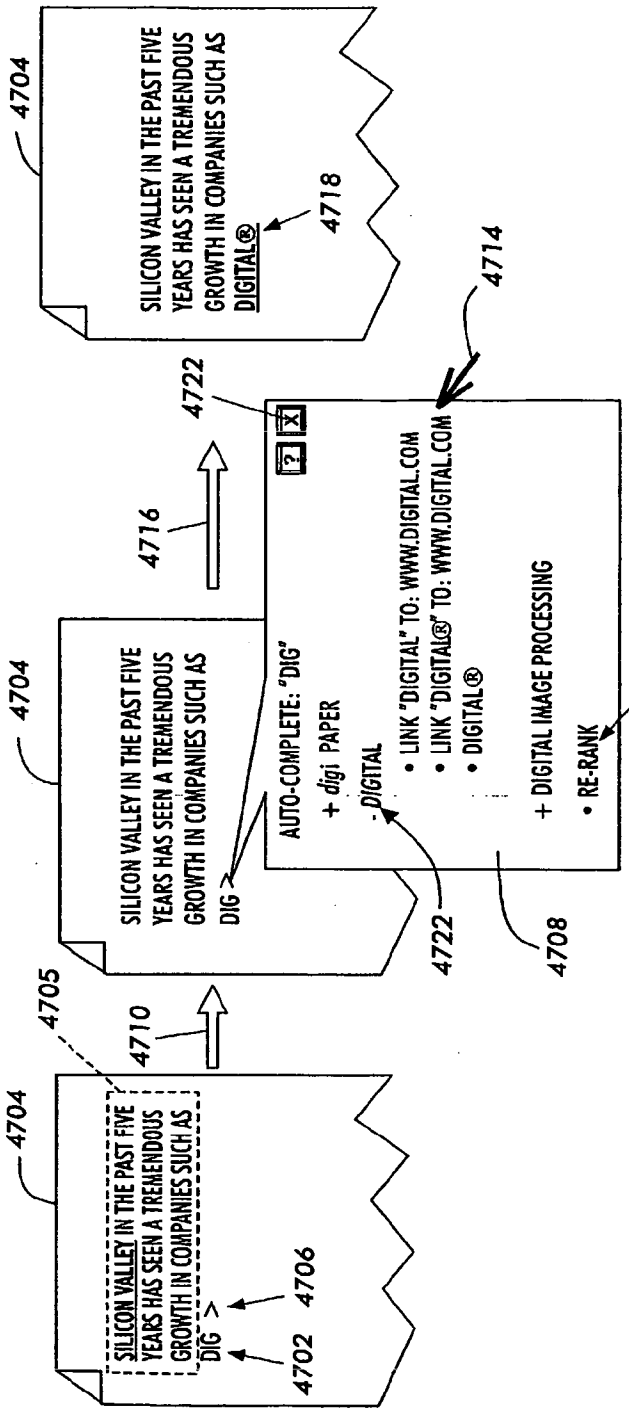


FIG. 47

(A) KEY	(B) ENTITY	(C) MARKUP	(D) FORMAT	(E) ORIGIN	(F) ORIGIN POSITION	(G) PART OF SPEECH	(H) CONTEXT
Sili	Silicon Valley	www.siliconvalley.com	Capitalize	TARGET DOCUMENT CONTENT	1	Noun	Science
Dig	Digital	www.digital.com	®	www.digital.com	434	Adjective	Science
Dig	DigiPaper	NONE	Bold	www.parc.xerox.com	23	Noun	Science
Dig	digress	NONE	NONE	www.nytimes.com	343	Verb	Politics
...	...	...	...	...	...	...	...

FIG. 48

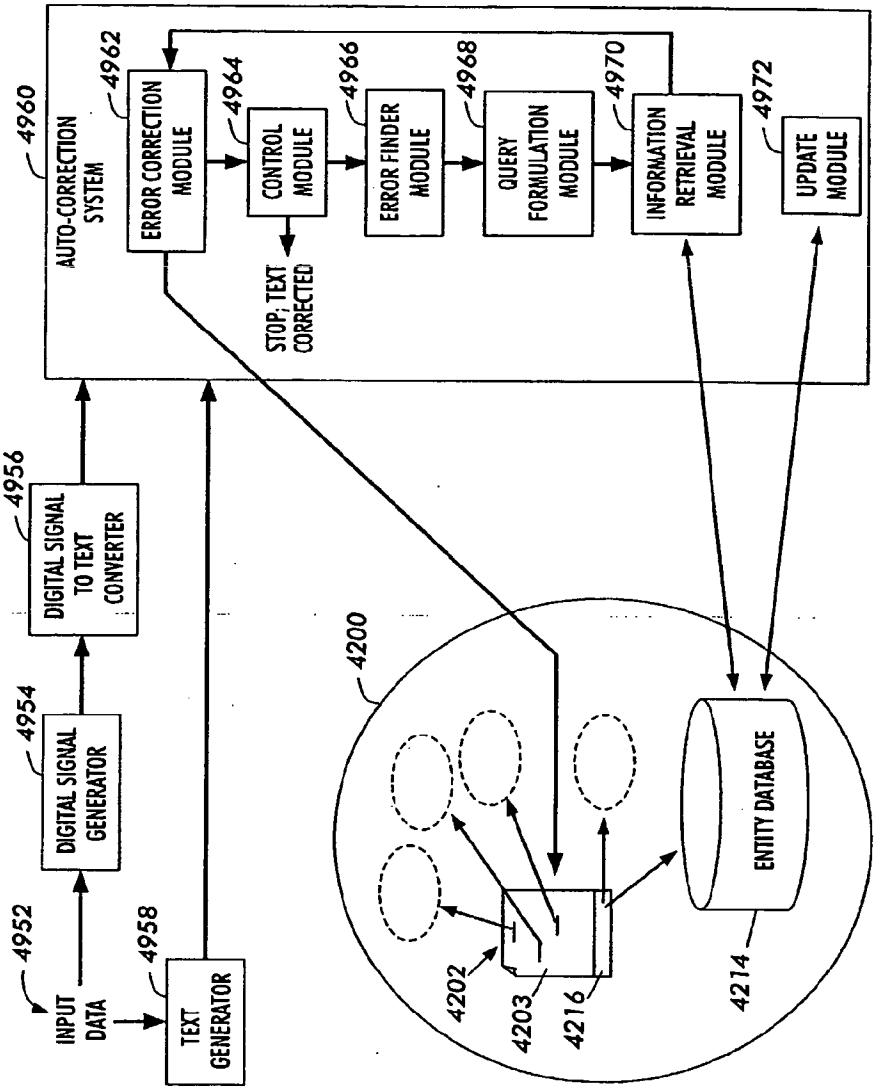


FIG. 49

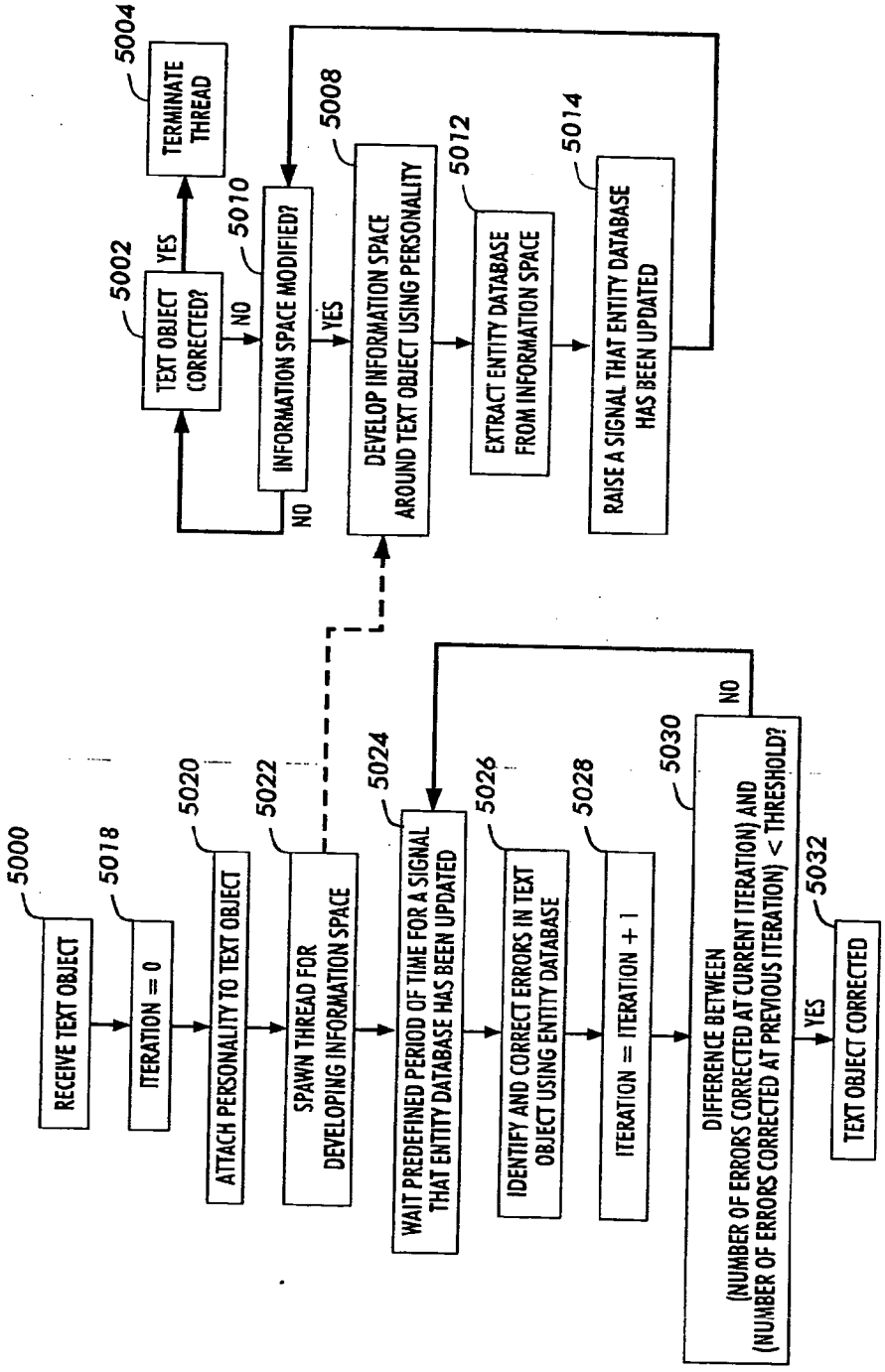


FIG. 50

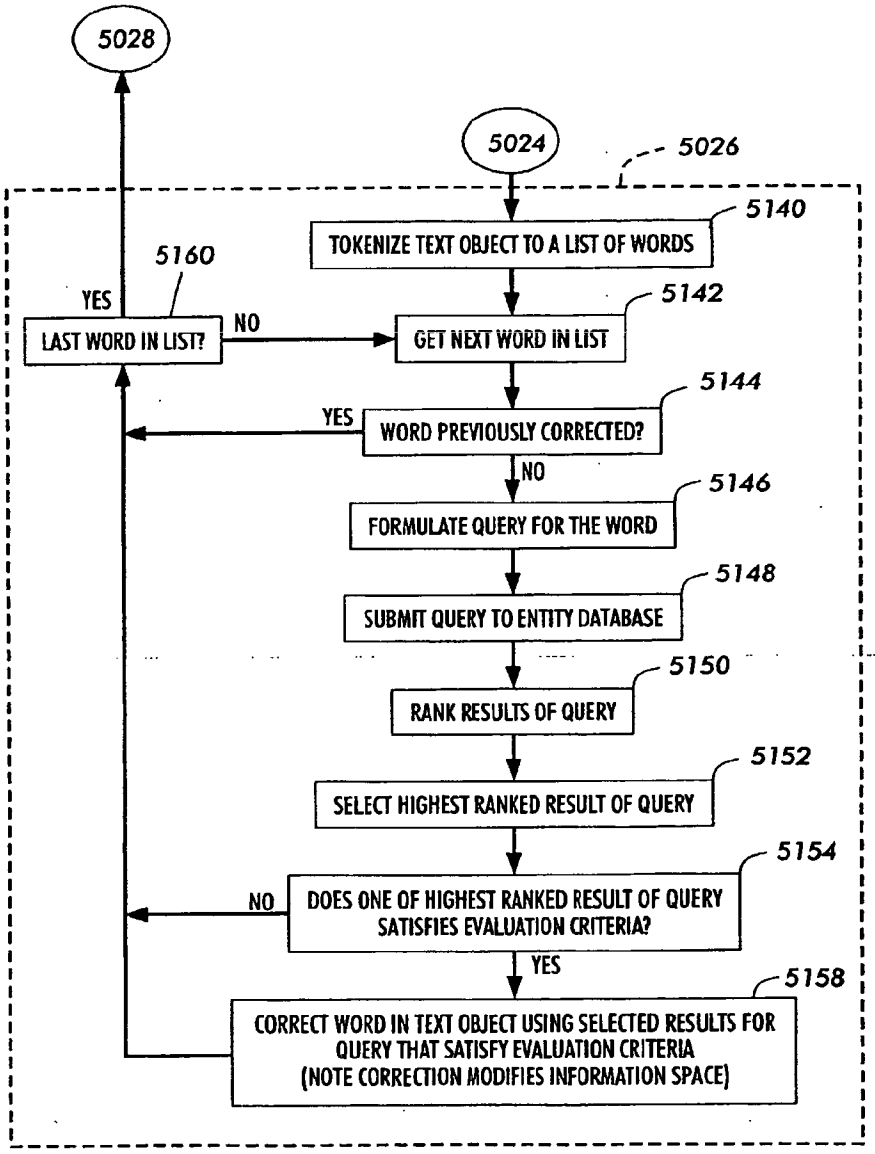


FIG. 51

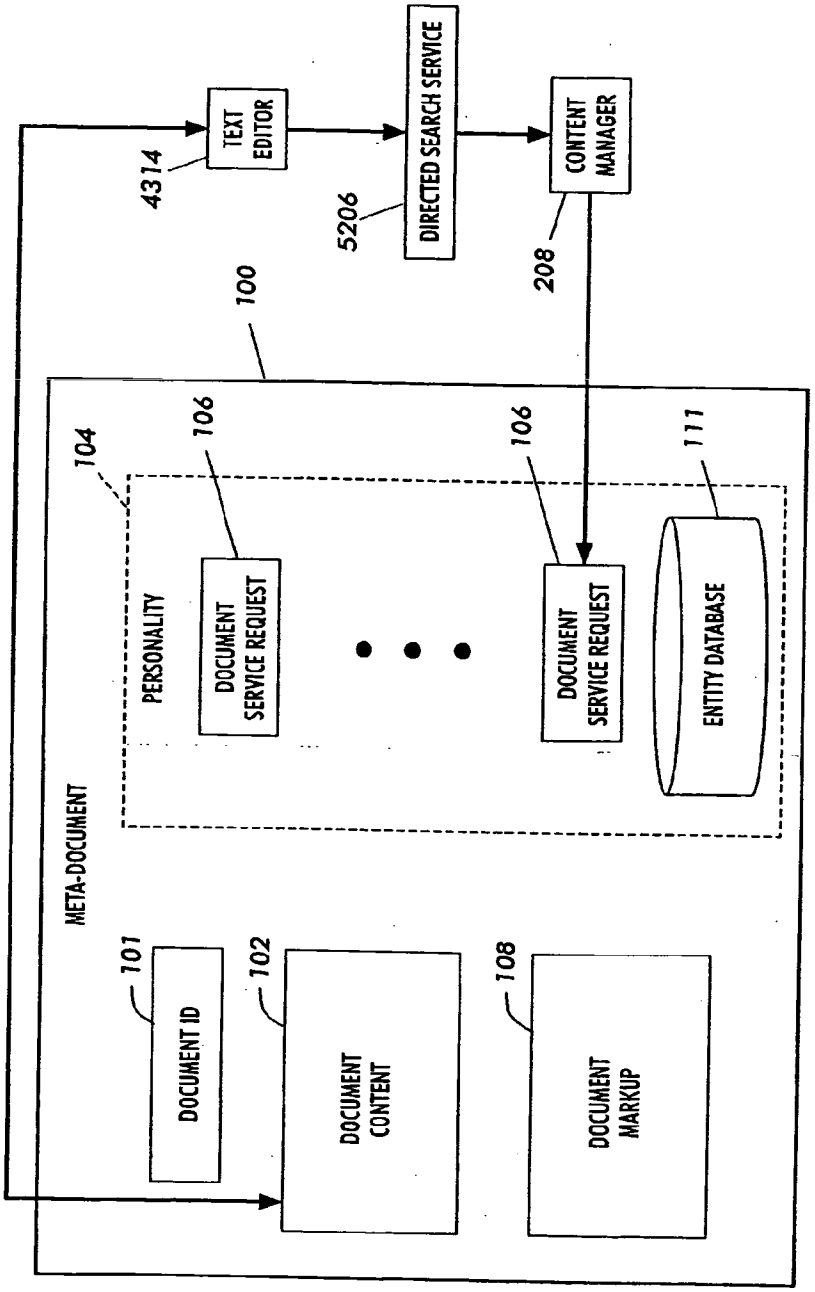


FIG. 52

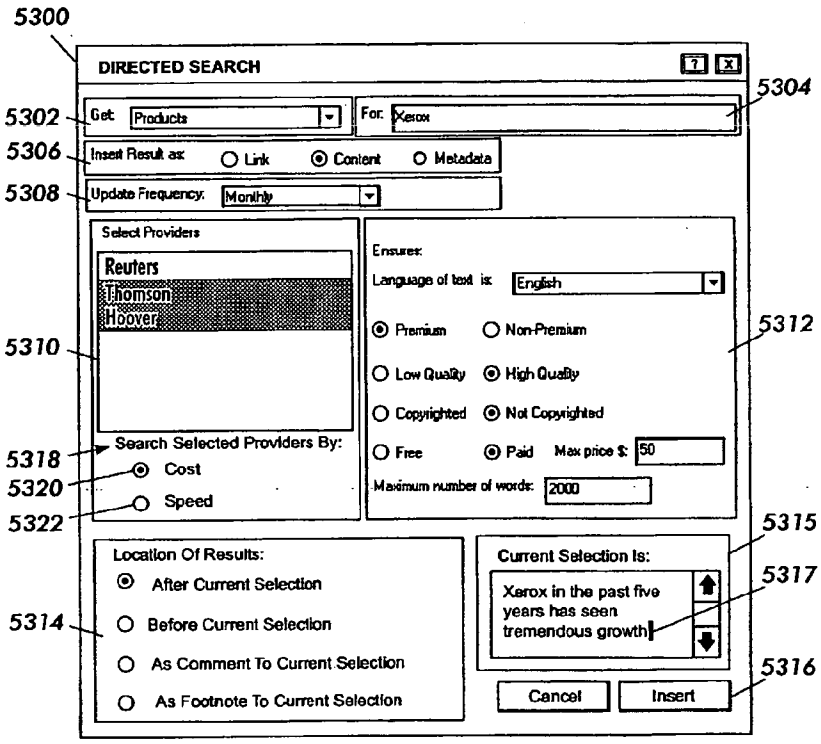

5300
DIRECTED SEARCH
5302
Get: Products
For: Xerox
5304
5306
Insert Result as: Link Content Metadata
5308
Update Frequency: Monthly
Select Providers
Reuters
Thomson
Hoover
5310
Ensures:
Language of text is: English
Premium Non-Premium
Low Quality High Quality
Copyrighted Not Copyrighted
Free Paid Max price $: 50
Maximum number of words: 2000
5312
Search Selected Providers By:
5318
Cost
5320
Speed
5322
Location Of Results:
After Current Selection
Before Current Selection
As Comment To Current Selection
As Footnote To Current Selection
5314
Current Selection Is:
5315
Xerox in the past five years has seen tremendous growth
5317
Cancel
Insert
5316

FIG. 53

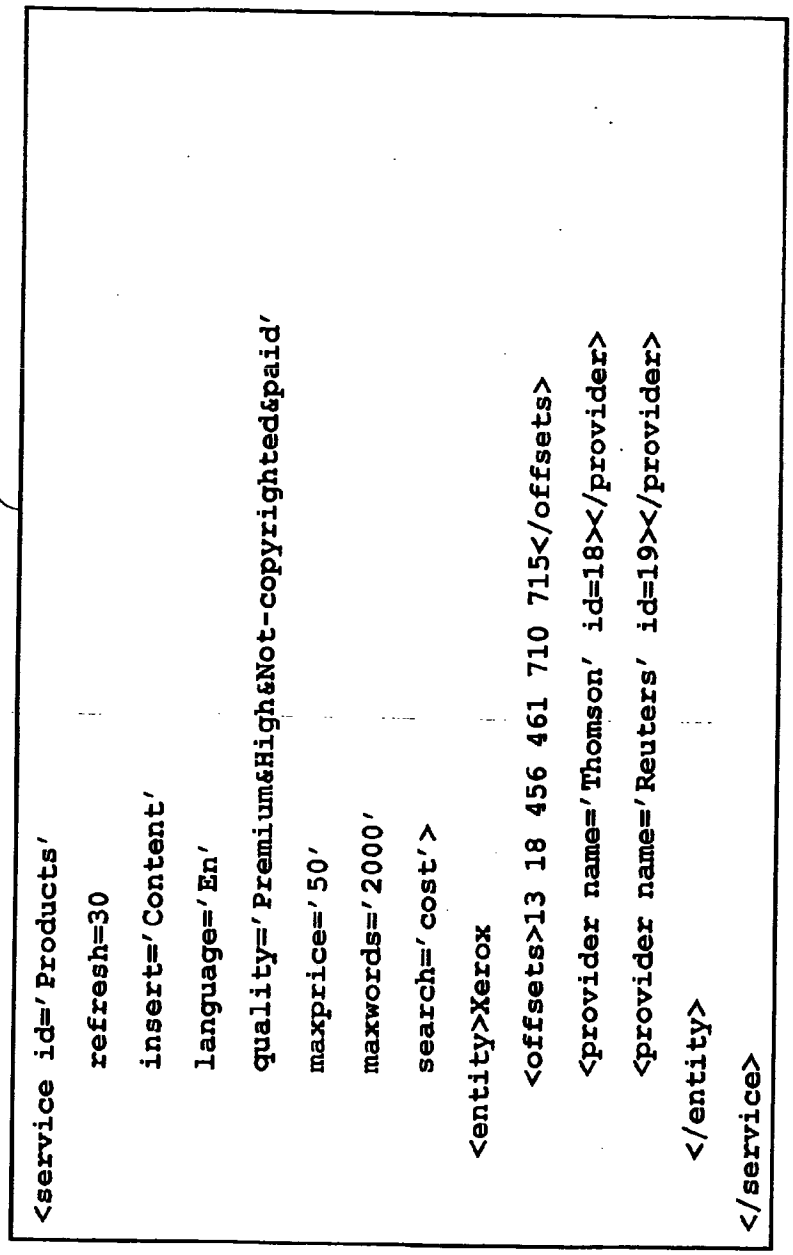


FIG. 54

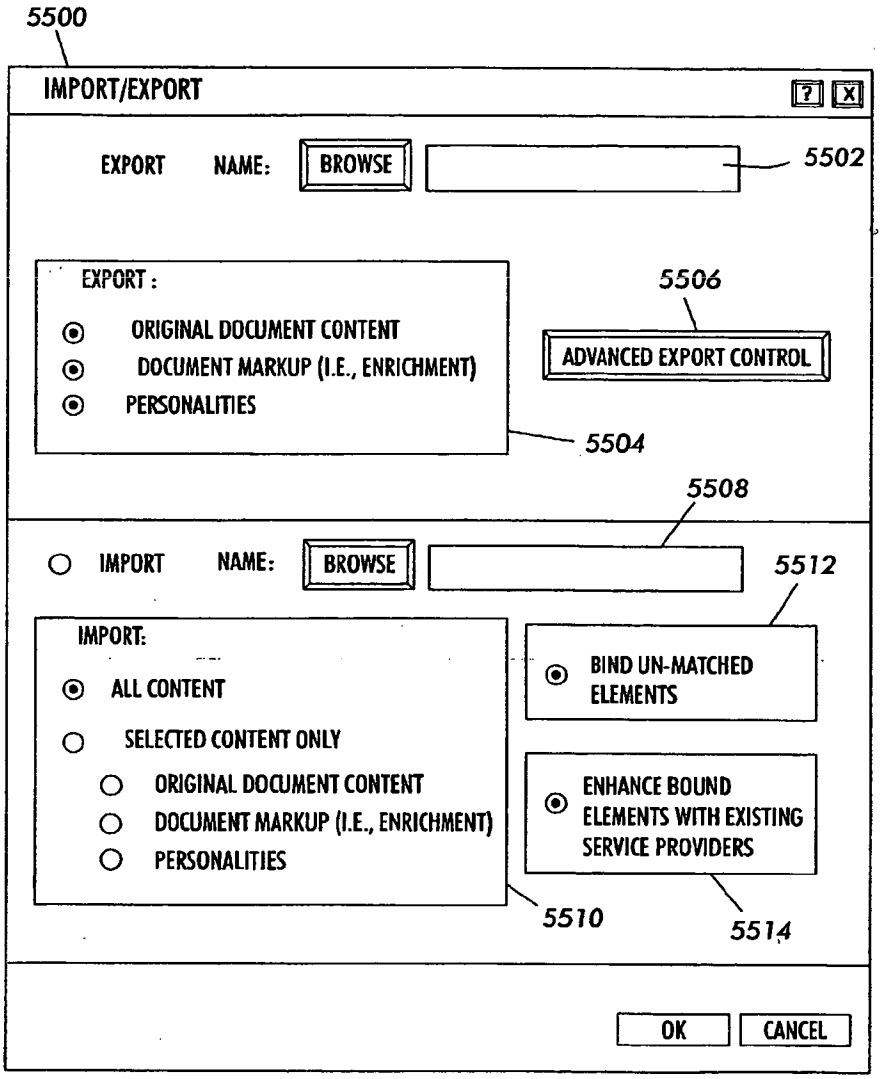


FIG. 55

```
<!- SERVER ID>
<docsouls:serverid> DS v1.0 server name </docsouls:serverid>
     <!- PROPERTIES OF DOCUMENT>
     <docsouls:docid> xxx </docsouls:docid>
     <docsouls:name> xxx </docsouls:name>
<docsouls:owner> xxx </docsouls:owner>
...
<!-- ORIGINAL DOCUMENTS DOCUMENT CONTENT>
<docsouls:originaldoc format="HTML">toto.html</docsouls:originaldoc>
<docsouls:browsingdoc format="HTML">browse.html</docsouls:browsingdoc>

<!- LIST OF PERSONALITIES AND DOCUMENT SERVICE REQUESTS>
<docsouls:endowmentdescription>
     <!-- LIST OF PERSONALITIES>
     <docsouls:personality id = " " name="" etc.>
          <docsouls:service id="" name="" etc.>
               <docsouls:provider name="" etc.>
               <docsouls:provider name="" etc.>
          </docsouls:service>
     <docsouls:personality>
     ...
     <!-- LIST OF SERVICE REQUESTS>
     <docsouls:service id="" name="" etc.>
          <docsouls:provider name="" etc.>
          <docsouls:provider name="" etc.>
     </docsouls:service>
     ...
</docsouls:endowmentdescription>
```

FIG. 56

108

```
<!- DESCRIPTIONS OF PERSONALITIES>
<docsouls:personalitydescription>
     <docsouls:id> xx </docsouls:id>
     <!-- all the fields AND service categories>
     <docsouls:servicecategory id="" name = "">
</docsouls: personalitydescription>

<!- NEXT PERSONALITY DESCRIPTION>
...

<!- DESCRIPTION OF DOCUMENT SERVICE REQUESTS>
<docsouls:servicedescription>
     <docsouls:id> xx </docsouls:id>
     <!-- list all fields (e.g., "refresh period", category, etc.)>
     <!-- list all providers>
     <docsouls:provider id="" name = "">
     <!-- list all strategies>
     <!-- list status of services (original, frozen, mapped, etc.)>
     <docsouls:servicestatus status="mapped" originalid="" originalserver="">
     <!-- list ServerId if the service was not available on current server>
     <docsouls:serverid>id</docsouls:serverid>
</docsouls:servicedescription>

<!- NEXT SERVICE REQUESTS DESCRIPTION>
...
```

FIG. 57

110

```
<!- DESCRIPTION OF PROVIDERS>
<docsouls:providerdescription>
     <docsouls:id> xx </docsouls:id>
     <!-- all the fields e.g. key, URL, etc.>
</docsouls: providerdescription >

<!- DESCRIPTION OF DICTIONARIES>
<docsouls:dictionarydescription>
     <docsouls:id> xx </docsouls:id>
     <!-- all the fields, e.g. category>
</docsouls:dictionarydescription>

<!- DESCRIPTION OF STRATEGY>
<docsouls:strategydescription>
     <docsouls:id> xx </docsouls:id>
     <!-all the fields>
     <!-all the dictionaries>
</docsouls:strategydescription>

<!- RESULTS OF GLOBAL SERVICES >
<docsouls:globalservice id="Summary">
     <docsouls:link> xxx </docsouls:link>
     <docsouls:content> xxx </docsouls:content>
</docsouls:globalservice>

<!- OTHER RESULTS (E.G. KEYWORDS, CATEGORIES, ETC.>
```

FIG. 58A

```
<!- (OPTIONAL) DESCRIPTION OF ENRICHMENT PERFORMED AT OTHER DOCUMENT ENRICHMENT
SYSTEMS>
<docsouls:serverid>olderServerId</docsouls:serverid>
<docsouls:personalitydescription>...</docsouls: personalitydescription>
<docsouls:servicedescription>...</docsouls:servicedescription>

<!- (OPTIONAL) LIST OF RESULTS OF THE LAST ENRICHMENT>
     <!- IDENTIFY DOCUMENT SECTION IN CASE OF MULTI-PART DOC>
     <docsouls:docpart>main_frame.html</docsouls:docpart>
     <!- OFFSET WHEN ENRICHMENT APPLIES TO A SECTION>
     <docsouls:offset> 0 200 </docsouls:offset>
     <!- STATUS OF SERVICE APPLIED ("ACTIVE" OR "FROZEN")>
     <docsouls:service id="StockQuote" status="active">
          <!-- ENTITY>
          <docsouls:entity>Xerox
               <!-- NORMALIZED FORM >
               <docsouls:normform>Xerox</docsouls:normform>
               <!-- LIST OF OFFSETS WHERE ENTITY IS FOUND IN TEXT RELATIVE TO SECTION>
               <docsouls:offsets start=35 end=40 />
               <docsouls:offsets start=232 end=237 />
               <!-- DICTIONARIES >
               <docsouls:dict> xxx </docsouls:dict>
               <!-- STRATEGIES >
               <docsouls:strat> xxx </docsouls:strat>
               <!-- LINK INFORMATION >
               <docsouls:link> xxx </docsouls:link>
               <!-- CONTENT INFORMATION >
               <docsouls:content> xxx </docsouls:content>
          </docsouls:entity>
          <!-- NEXT ENTITY>
          ...
     </docsouls:service>
```

FIG. 58B

EP 1 284 461 A1

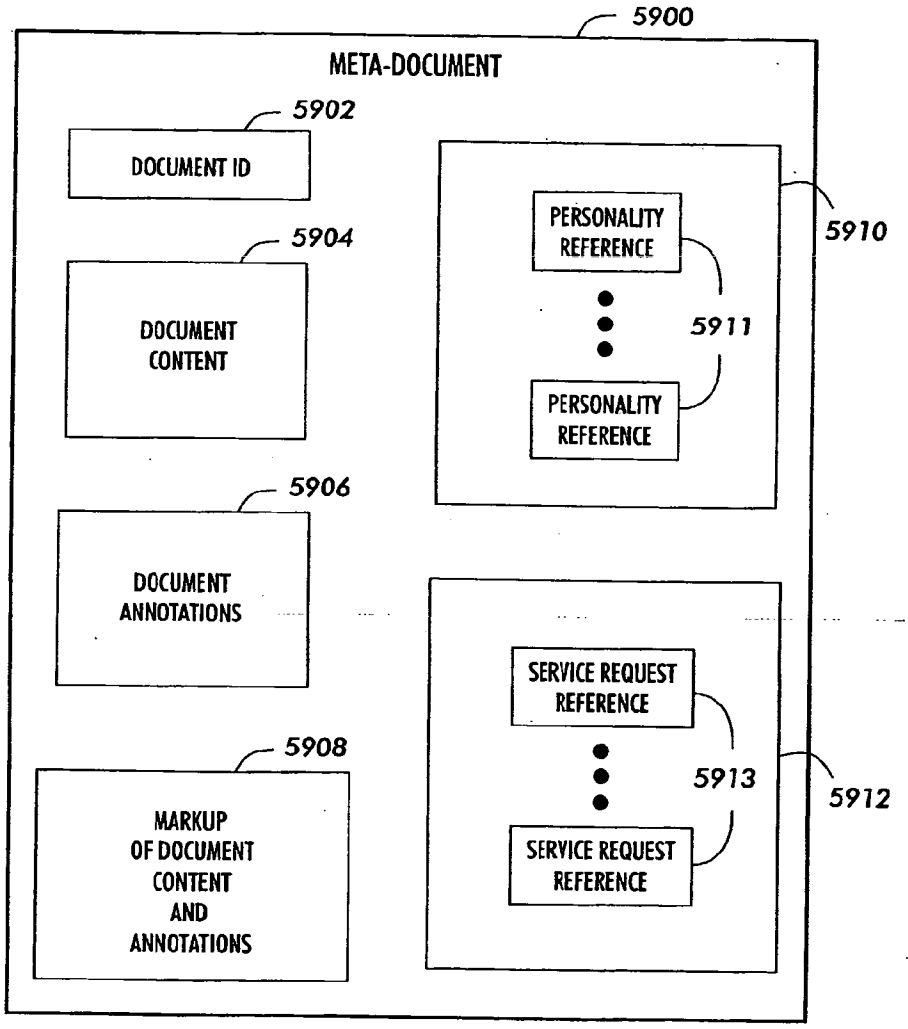


FIG. 59

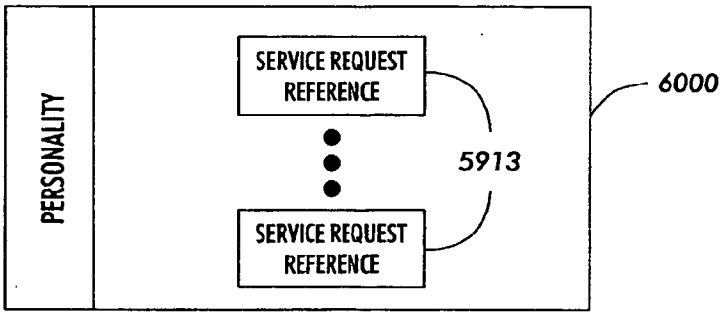


FIG. 60

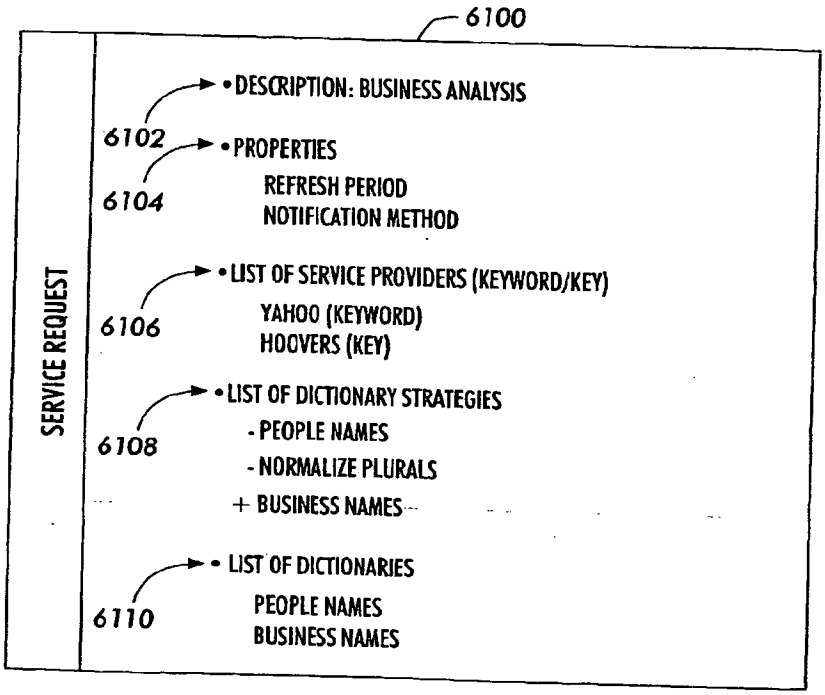


FIG. 61

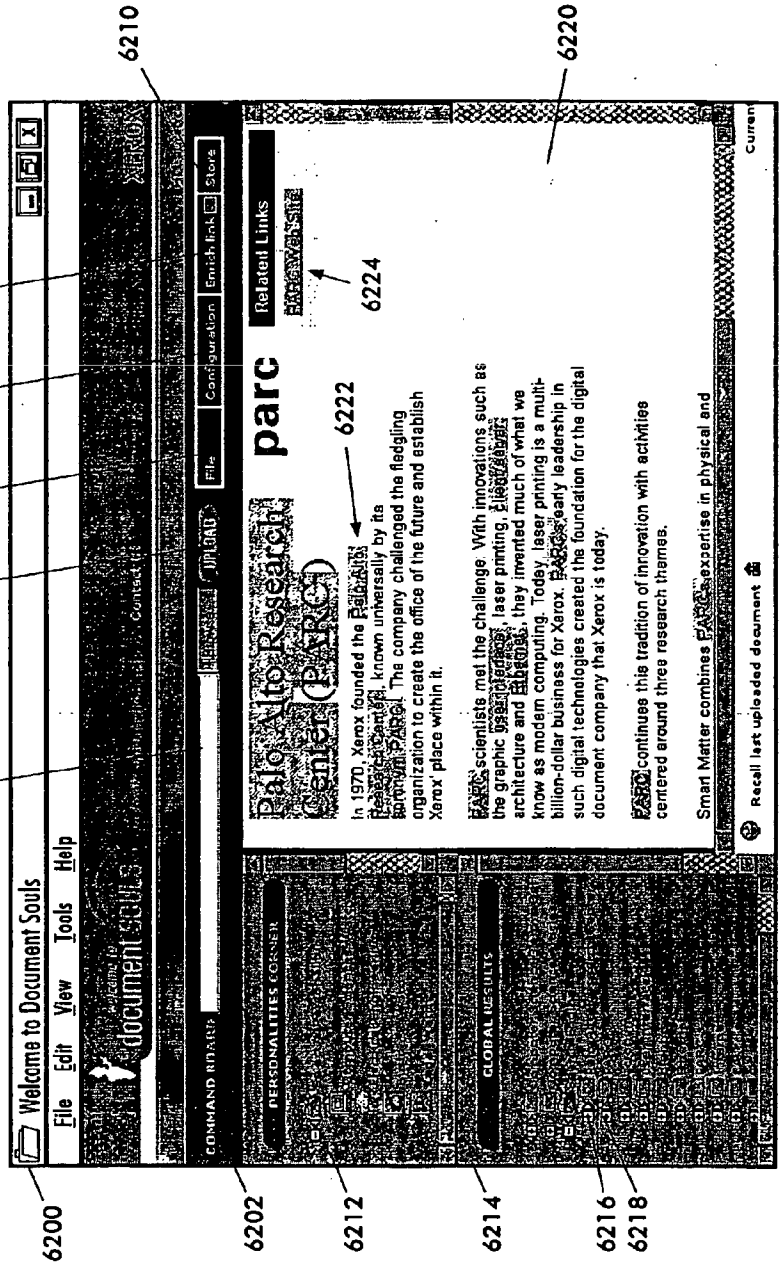


FIG. 62

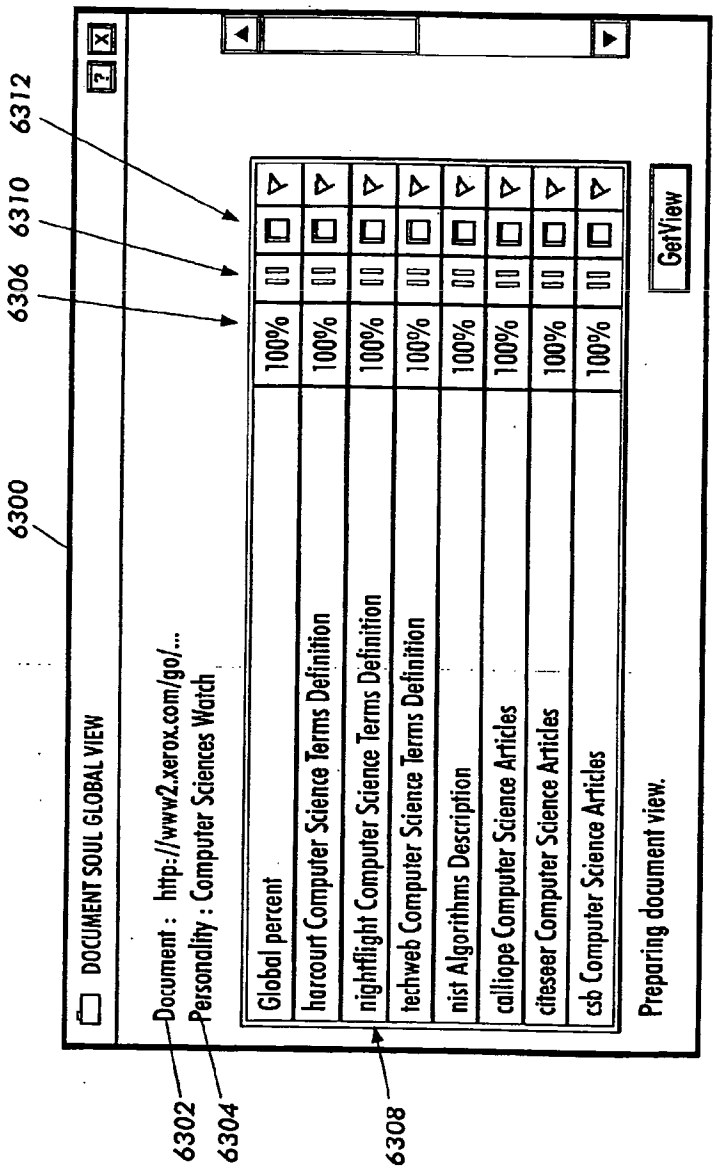


FIG. 63

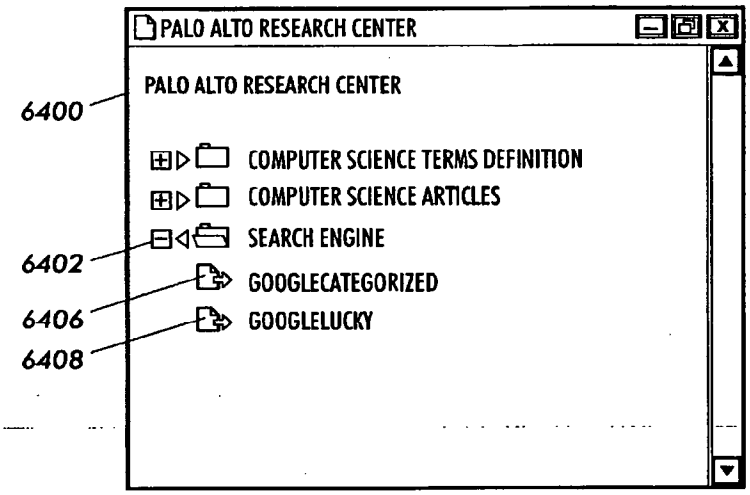


FIG. 64

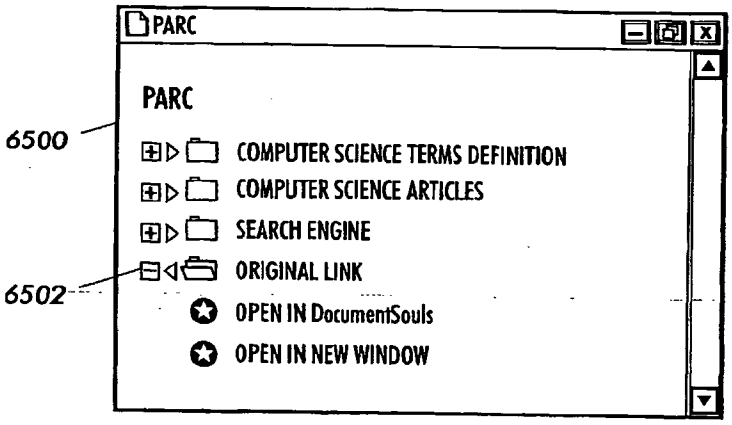


FIG. 65

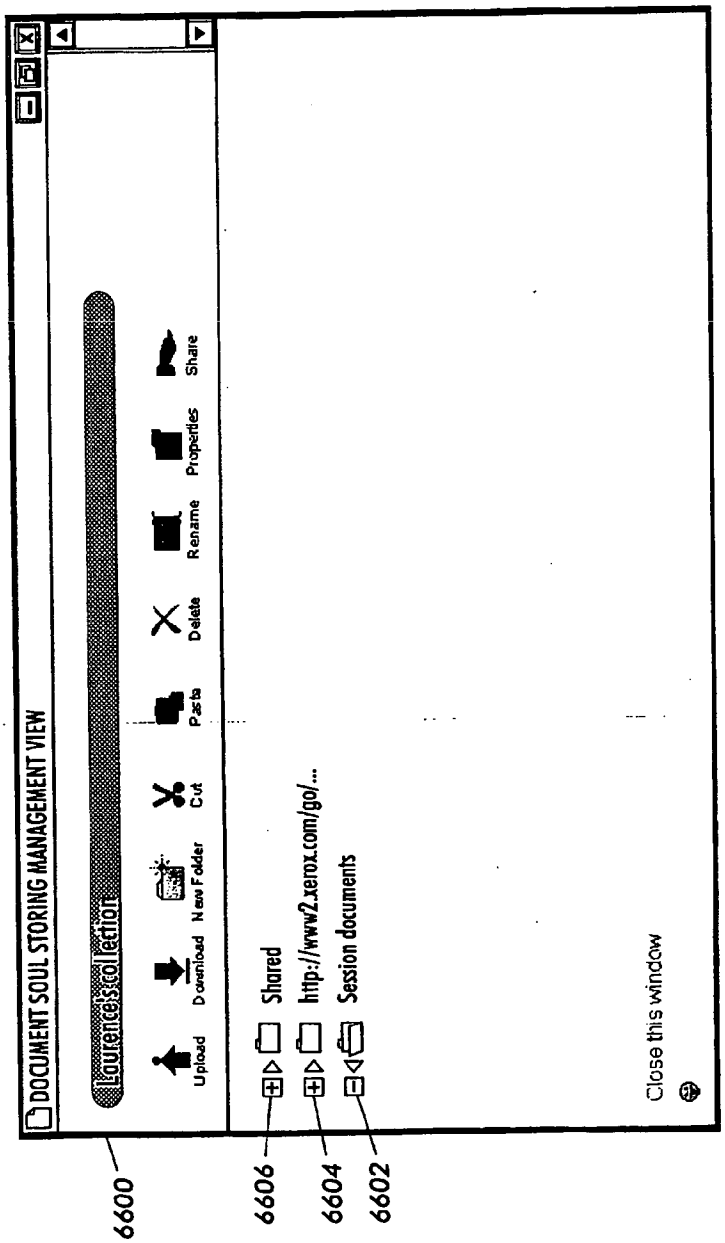


FIG. 66

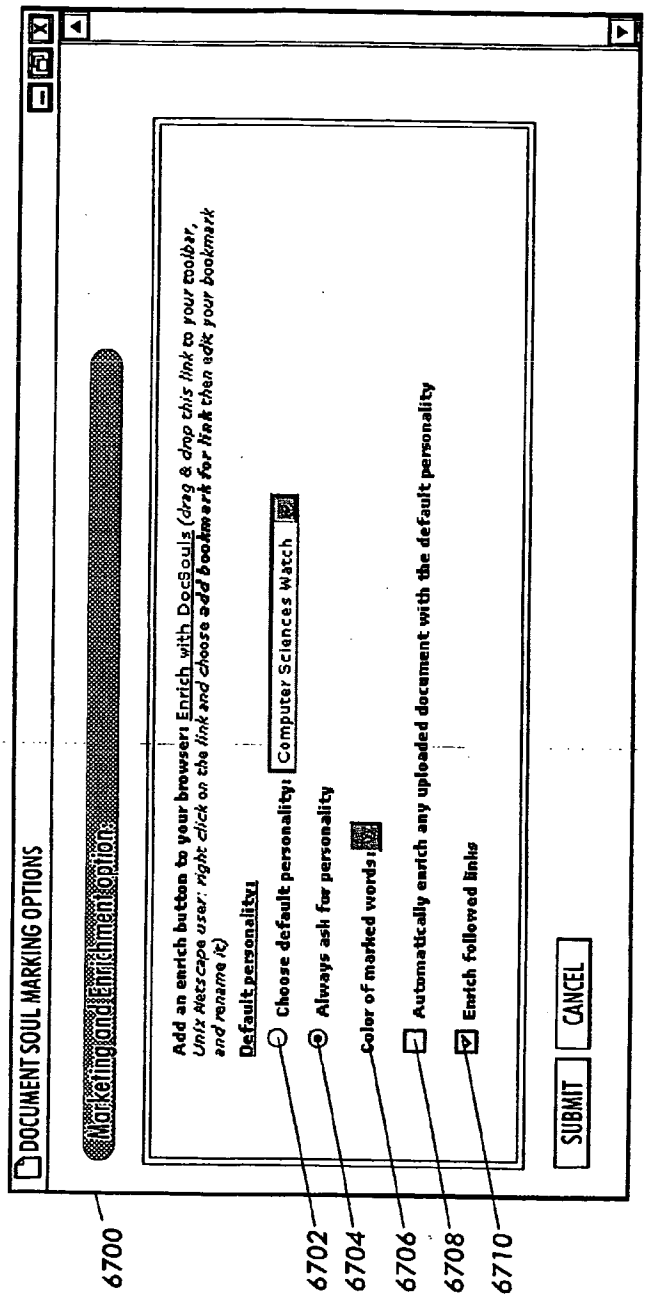


FIG. 67

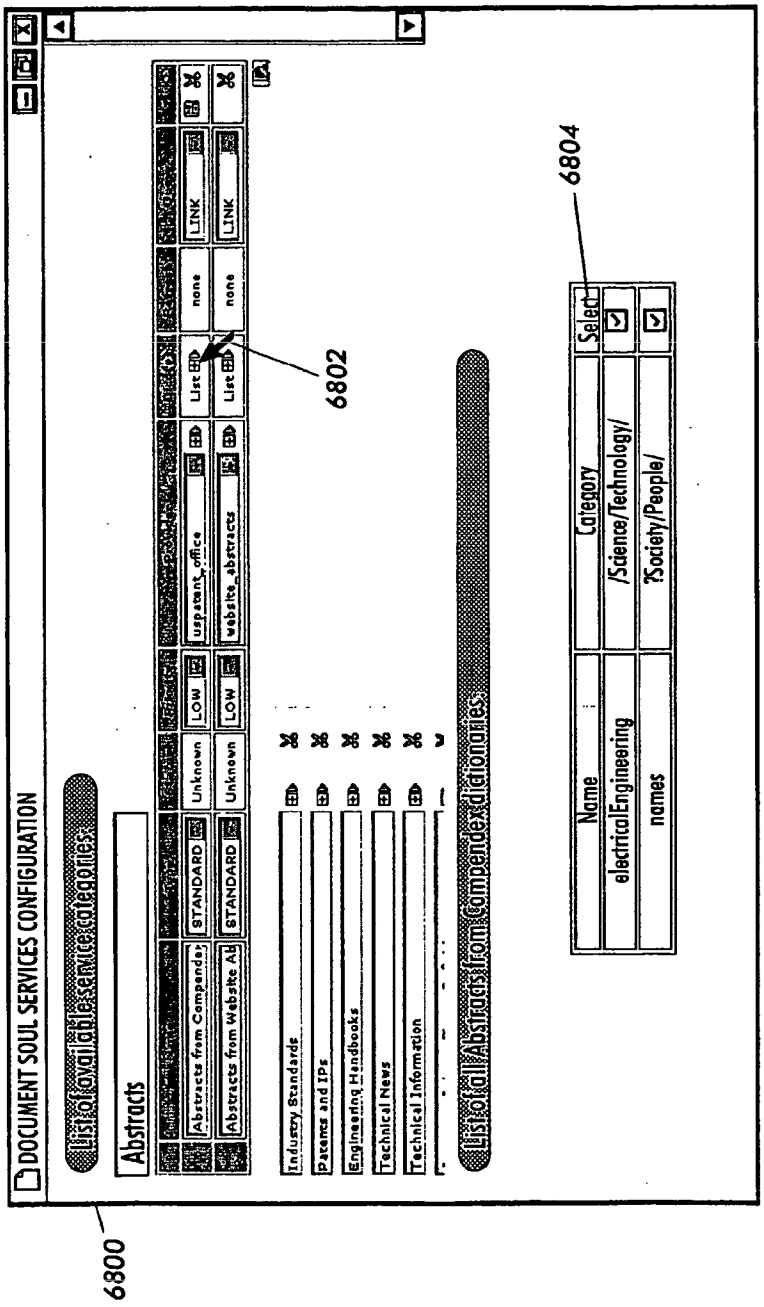


FIG. 68

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 01 8111

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
P,X P,A	EP 1 143 356 A (XEROX CORP) 10 October 2001 (2001-10-10) * page 2, paragraph 2 - paragraph 3 * * page 2, paragraph 9 - page 3, paragraph 14 * * page 4, paragraph 23 * * page 5, paragraph 27 * * page 5, paragraph 32 - paragraph 34 * * page 6, paragraph 37 - paragraph 39 * * page 6, paragraph 42 - paragraph 44 * * page 7, paragraph 46 * * page 11, paragraph 57 * * figure 3 * * page 12, paragraph 69 *	1-3,11, 16,17 4-10, 12-15,18	G06F17/30
D,A	JAY BUDZIK, KRISTIAN HAMMON: "Watson: Anticipating and Contextualizing Information Needs" PROCEEDINGS OF THE SIXTY-SECOND ANNUAL MEETING OF THE AMERICAN SOCIETY OF INFORMATION SCIENCE, 'Online! 1999, XP002224424 Retrieved from the Internet: <URL:http://citeseer.nj.nec.com/cache/papers/cs/10883/http:zSzzSzdent.infolab.nwu.edu zSzinfolabzSzdownloadszSzpaperszSzpaper10056.pdf/budzik99watson.pdf> 'retrieved on 2002-12-10! * abstract * * page 3, line 27 - line 47 * * figure 2 *	1-18	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G06F

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	18 December 2002	Dumitrescu, C

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 01 8111

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	FREDRIKSSON & M SVAHNBERG M: "Fargo: Technical Description" UNIVERSITY COLLEGE OF KARLSKRONA/RONNEBY, 1997, XP002151013 * page 2, line 15 - line 37 * * page 3, line 14 - line 35 *	1-18	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	18 December 2002	Dumitrescu, C

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 02 01 8111

This annex lists the patent family membersrelating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

18-12-2002

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1143356 A	10-10-2001	EP 1143356 A2	10-10-2001
		JP 2001331511 A	30-11-2001

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 288 795 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
05.03.2003 Bulletin 2003/10

(51) Int Cl.7: **G06F 17/30**

(21) Application number: **01307210.3**

(22) Date of filing: **24.08.2001**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **BRITISH TELECOMMUNICATIONS public limited company**
London EC1A 7AJ (GB)

(72) Inventor: **The designation of the inventor has not yet been filed**

(74) Representative: **Maggs, Michael Norman et al**
Kilburn & Strode
20 Red Lion Street
London WC1R 4PJ (GB)

(54) Query systems

(57) A computer query system and method enhances the results produced by standard queries, for example in search engines, by filtering and optionally ordering the results on the basis of affinity with individuals known to the search initiator, or individuals within the search initiator's Personal Social Network. The system allows users searching databases for information to evaluate the results of a query with reference to feedback that has previously been provided by others within the user's social network. This allows users to make informed decisions based on how much they value the opinions of others within their personal network.

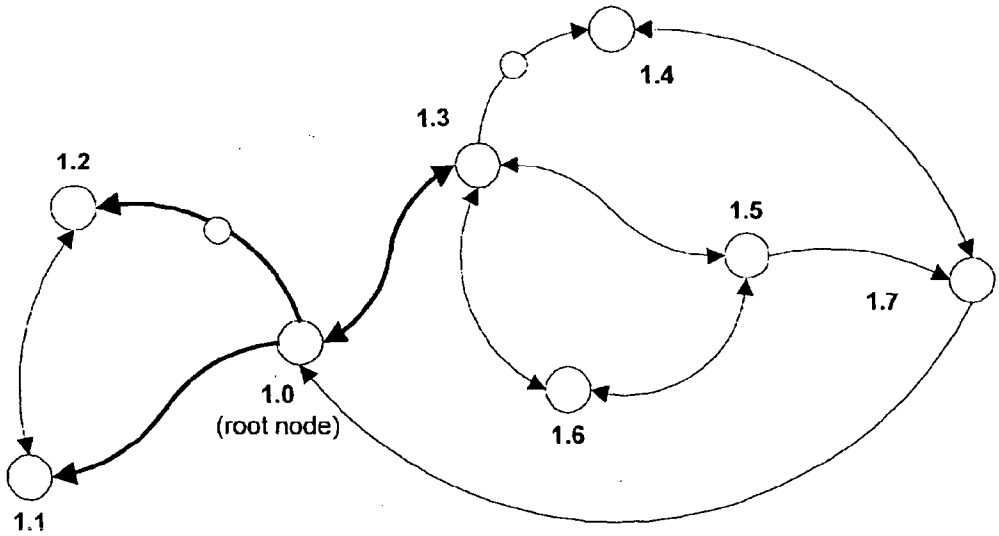


Figure 1

EP 1 288 795 A1

Printed by Jouve, 75001 PARIS (FR)

Description

[0001] The present invention relates to computer based queries systems and has applications in fields such as search engines, instant messaging services and knowledge management systems.

[0002] Keyword based query systems, such as for example those used by internet search engines, have been available for many years. Because of the large number of results or "hits" typically returned by such search engines, there has been increasing interest in finding ways of filtering and/or prioritising the results so that they can be presented to the user or "search initiator" in a meaningful form. Typically, weighting factors may be applied to the returned results in some way, so that the more interesting more relevant results appear towards the top of the returned list. One way of doing this, used by some internet search engines, is to assign "popularity scores" to URLs based upon the number of times the URL has been accessed by the search engine website.

[0003] Another approach to prioritisation is to use Collaborative Based Filtering (CBF) scheme. These systems find relevant documents by looking at ratings and recommendations (which may be explicit or implicit) made by others. Examples of Collaborative Based Filtering schemes include the following:

Information Forwarding Systems (IFS):

[0004] These systems allow people to forward information they find interesting to groups of people with a similar interest. These systems are sometimes called Active Collaborative Filtering systems because, here, users explicitly select the group of people to which they recommend an item. The system's role is simply to forward the material to a selected group.

Recommender Systems (RS):

[0005] These systems find items (documents, movies, products) relevant to a particular user by analysing the preferences of other users with similar profiles. These systems are sometimes called Passive Collaborative Filtering Systems since users do not recommend items to any one in particular. Instead, the system determines what items should be recommended to whom, based on similarities between user preferences. A licensable example of this technology can be found at NetPerceptions (www.netperceptions.com). This solution is used by e.g. online bookstores and provides a recommended reading list following a user's purchasing profile, weighed against other users who have purchased similar books.

"Top N" Systems (TNS):

[0006] These systems allow a user to navigate through a list of the most popular documents in a given collection. These systems differ from Recommender Systems insofar as their recommendation is not based on a priori knowledge about a user's particular interest, but only on the overall popularity of the documents themselves.

[0007] Other approaches allow authors to rate their own websites for others to use and to prioritise or filter based upon those ratings. Such system are used by service providers who wish to provide "safe" content, for example www.safesurf.com.

[0008] According to a first aspect of the present invention there is provided a computer query method comprising receiving a query from a search initiator, carrying out a search for information on the basis of the query and filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

[0009] According to a second aspect of the invention there is provided a computer query system comprising means for receiving a query from a search initiator, means for carrying out a search for information on the basis of the query, and means for filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

[0010] The invention further extends to a computer program for carrying out the above method and to a data carrier which carries such a computer program.

[0011] The method of the invention will be referred to within this patent application by the term "Affinity based queries" (ABQ). ABQ allows users searching databases, or the internet, an ability to evaluate the results returned by virtue of inter-personal relationships in the context of their social network. The user can, in that way, be presented preferentially with results which have been recommended by people known to him or her, or otherwise with in his or her social network.

[0012] In the preferred embodiment, results are categorised not only by the degree with which they fit the search criteria (this is currently done), but they are further categorised with metadata describing the relationship between the search initiator and an individual within the network who has refereed, rated or is otherwise acquainted with one of the results. This allows users to make informed decisions based on the extent to which they value the opinion, or the prior

actions, of other members of their social network.

[0013] The invention is preferably implemented through a combination of relational databases, instant messaging (or some other community tool holding information on the social network), along with other query software to implement the ABQ rules and to drive the process of traversing through the social network. This is preferably implemented through nested SQL queries and graphing techniques.

[0014] In a preferred embodiment, an instant messaging application or an internet search engine forms the key service entry point. A relational database backend holds information on the social network, and uses the internal relationships within the network to carry out filtering and/or sorting on the results of a standard search (a keyword-based search). Whenever information is publicly exchanged between individuals within the network (either explicitly or automatically/implicitly) the system preferably intercepts the exchange and records entries in the data base defining metadata links which indicate:

1. Where the information is,
2. Summaries of that information,
3. Keywords about the information,
4. Any ratings that may have been applied (e.g. to an article product or individual),
5. Details of the individual who referred it,
6. Details of the individual to whom it was referred,
7. Details of the time and date when it was referred.

[0015] Then, whenever a user initiates an ABQ, the system preferably uses that person's entry in the database as the root node in the search graph and begins to percolate out through the social network conducting nested queries and table joins between persons and content, to a predefined distance or order, to retrieve the information.

[0016] In order to encourage involvement of the individuals within the network, reward feedback may be provided to the individual who recommended to the user the search result that the user decides is "best". For example, if a user is looking for an article on a particular subject, and selects an article on the internet that has been recommended by an individual X, X receives a reward in the form of an enhancement to his or her "action value". Individuals with high action values will have made good recommendations many times and other users will know that they can be relied upon. Preferably, the system allows the user the opportunity to sort the search results by the action values of the referring individuals thereby insuring that the recommendations from trusted individuals appear at the top of the list.

[0017] The invention may be carried into practice in a number of ways and one specific embodiment will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 shows an exemplary Personal Social Network (PSN):
Figure 2 illustrates schematically the main apparatus elements of an embodiment of the present invention:
Figure 3 illustrates the linking of the PERSON tables:
Figure 4 illustrates the action tables:
Figure 5 shows the method of registering and logging in referees in the database:
Figure 6 shows the method of modifying PERSON records:
Figure 7 shows the method of deleting PERSON records:
Figure 8 shows the method of storing relations:
Figure 9 shows the method of deleting relations:
Figure 10 shows the method of summarising and storing actions:
Figure 11 shows a method of processing incoming query parameters:
Figure 12 shows the method of building an in-memory view:
Figure 13 shows a digraph method of building in-memory views:
Figure 14 shows a digraph and binary tree method of building in-memory views, avoiding duplicates:
Figure 15 shows the method of filtering key word matches:
Figure 16 shows the method of ordering result items by rating:
Figure 17 shows how edge weightings may vary throughout a Personal Social Network:
Figure 18 shows the ordering of the in-memory view by relation:
Figure 19 shows the ordering of the results by relation:
Figure 20 shows the method of encoding and presenting the result list; and
Figure 21 shows the method of feeding back result actions.

1 Introduction

[0018] The preferred embodiment of the invention uses a system called Affinity Based Queries (ABQs) to improve

the relevance of search results to the person requesting them (known as the "search initiator"). In very general terms, the system works as follows:

1. The search initiator first does a standard keyword search e.g. to search for a particular book (a bookshop website).

2. The search returns a list of results or "hits" which are then filtered on the basis of information derived from a Personal Social Network (PSN) of which the search initiator is a member. Personal Social Networks will be described in more detail in section 2. The filter selects from the results list hits which are associated in some predefined way with others in the PSN, for example individuals who have rated the book. The other hits are discarded. This produces a list of filtered result items.

3. The result items are then ordered in some user-defined order for presentation to the search initiator. they could be ordered, for example, by the rating which other individuals have provided, or by the relationship between the search initiator and the other individuals (e.g. ratings by close friends take precedence), or they could be ordered by result "action" (e.g. putting at the top of the list recommendations by an individual whose previous recommendations have been good).

4. Finally, the system presents the sorted and filtered list of results to the search initiator along with an indication, in each case, of which individual within the PSN has recommended the book, and how that individual is related within the PSN to the search initiator.

[0019] It will be appreciated of course that Affinity Based Queries can be used for all types of information retrieval. We will refer to the information to be retrieved as a "document", it being understood that that term includes not only traditional documents in electronic form, but also internet web pages and, more generally, all types of searchable information whether held with in a database or otherwise. When we say, for example, that the search initiator is looking for "books" what we really mean is that he or she is searching one or more databases, typically held on the internet, which contain not books themselves but information about books.

[0020] A more formal description of the preferred embodiment will now be set out. The reader is referred to the glossary of terms, at the end of the specific description, which provides an explanation of certain expressions.

[0021] When an Affinity Based Query (ABQ) is run, it returns a result list which consists of a number of result items. Each result item has some association with the "referee" who is somebody within the search initiator's PSN who is either directly associated with the result item, or who is indirectly associated with the item through being related to someone else within the search initiator's PSN. Along with the result, each result item may also indicate a "relation" (showing how the referee is related to the search initiator through the PSN), along with a "rating" (an award which has been applied by a referee to a result).

[0022] Each result item in an ABQ result list is presented in the form:
<referee><relation><result><rating>.

[0023] Order is not important, nor is the frequency of in which terms are used.

[0024] In the above expression, the following terms are used:

- **referee** - mandatory term, a person associated with a result;
- **relation** - optional term, relating a referee to the search initiator through the PSN;
- **result** - mandatory term, the location (e.g. URL) of something returned from a query;
- **rating** - optional term, an award applied by a referee to a result.

[0025] The result of an ABQ is a result list in which keyword matches are "annotated" or "embellished" by additional information specifying not only what other referees thought of the result but how those referees are associated with or linked to the search initiator.

[0026] Taking the example of an on line bookstore, a result item might read:
*e.g. "Kate **(referee)**, who knows Nick **(relation)**, purchased 'Wuthering Heights' by Emily Bronte **(result)**"*

[0027] Or, consider a search engine. Rather than just keyword-match a URL query, ABQ's return results referencing other persons known by the search initiator who have visited/rated the URL:
e.g. "Dan (referee), recommended http://www.domain.com/ (result), awarding it 4/5 (rating)"

[0028] Or, consider skills searching in an organisation. Rather than just return persons possibly unknown to the user, ABQ's return results in relation to someone known by the search initiator:
e.g. "Andy (referee), who knows Nick (relation), knows Oracle (result), and rates himself as EXPERT (rating)"

[0029] Or finally, consider an article rating:

e.g. "Jim (referee), who Andy knows (relation), who knows Nick (relation), rated article X (result), as FAIR (rating)"

[0030] Of course, in the above examples the text in brackets would not actually be seen by the user: it is included merely to show how the expressions are made up from the basic terms which include referee, relation, result and rating.

[0031] By decorating standard database query results with respect to someone known by the search initiator, ABQ's allow users to make more informed decisions about which result to follow first, and a reference point to qualify that order with (i.e. the referee).

[0032] As mentioned above, the result list may be ordered in a variety of different ways, according to the needs of the search initiator. That will be described in more detail below.

2. Personal Social Networks (PSN)

[0033] Personal Social Networks are, per se, a known concept. Such a network connects an individual to a number of other individuals with in a community (real or virtual). A PSN can be described by means of a directed graph (a "digraph"). An example is shown in figure 1. Each node in the digraph represents an individual, and the directed edges between the nodes represent relationships. Relationships are
"reflexive" in nature and do not necessarily imply reciprocity: thus, a may know b, but b may not know of a.

[0034] In figure 1, the first order PSN represents those nodes that can be reached in one "hop" from the root node (in this case the search initiator), the second order network those nodes that can be reached in two "hops", and so on.

[0035] In addition to the nodes, representing a community of individuals, the PSN of this embodiment also stores information on "actions". These will be discussed in more detail in section 3, below, but, essentially, represent things done by individual nodes, which are relevant to the particular document being sought. Examples include visit URL, read article, purchase products, volunteer skill and so on. Information on actions is stored against each node of the PSN.

[0036] A search initiator may have a very large available PSN, in theory, out to order N. However, not all of that PSN may be useful for the search being carried out, and the search initiator is provided with means (such as a data entry field or a series of tick boxes) to restrict the search to a PSN of a user-selectable order. Typically the user may wish to restrict the search to a first, second or third order PSN.

[0037] A further option may be made available allowing the search initiator to exclude particular individuals or classes of individuals from that part of the PSN which is going to be used for the search. For example, the user may wish to exclude work-based contacts for a search which is being carried out in respect of a personal hobby.

[0038] Once the search initiator has specified the appropriate parameters, the system works through that individual's PSN, link by link, to construct the subset of the PSN that is going to be used in the search. This subset is then stored in memory as an "in-memory view". Storing information in memory ensures that that part of the PSN that is going to be used for searching does not need to be recreated if the user wishes to carry out additional searches using the same criteria.

[0039] Further details of the way in which the digraph of figure 1 is built up from a relational database, and traversed, will be described later.

[0040] The grouping of individuals making up the PSN, and the type of relations between them, is not of importance to the operation of the invention as a whole, but will depend upon the implementation and application. Typical PSNs might include individuals and their work contacts, groups of friends, groups of individuals with common interests or individuals conveniently grouped together for any other reason.

[0041] Typically (although not necessarily) each individual may choose his or her direct "buddies" (that is those individuals with which there will be a direct link) , with the second and higher order PSNs being generated by the system without direct involvement from the individual representing the root node. Alternatively, the entire PSN may be generated automatically, for example, by analysing things done by the individuals, without any individual needing to confirm explicitly whether another individual should or should not be added to the PSN.

[0042] If affinity ordering is to be carried out on the result items (see sections 3.7 and 4.12), the links or edges in the digraph need to be weighted. The weighting could be fixed (e.g. family members have a higher weighting than friends,) or it could be dynamic (for example on the basis of past actions affecting the relationship between connected individuals). Weightings could include such things as the time the linked individual has been known, the number of messages exchanged, the number of times a good recommendation has been received, and so on. To ensure that out of date information is not maintained within the database, the weighting could decay with time. Edge weights may be unidirectional or bi-directional. For the message used to weight the edges in the preferred embodiment, see section 4.12.

3. Apparatus

[0043] A schematic view of the apparatus used in the present embodiment is shown in figure 2. A DBMS 10 stored on an ABQ server 12 is used to generate an in-memory view 14 of the PSN. The in-memory view is used to generate an affinity filter 20 which filters the raw key word matches 16, 18 to provide a list of result items 22. On the basis of

information taken from the in-memory view 14, the result items 22 are ordered in an affinity order 24 to create a result list 26. Affinity ordering may be by relation, rating or result action. Finally the result list is encoded and presented to the user at 28 along with, where appropriate, a copy of the filtered keyword list 30.

[0044] Further details of the preferred apparatus of the present invention will be described in the following subsections.

3.1 Persistent storage of relations, actions and metadata.

[0045] The system requires some form of repository- preferably in persistent form- for the storage of relations, actions and metadata (such as date information and so on). Either a Relational Database Management System (RDBMS) or Object Oriented Database Management System (OODBMS) can be employed. In the specific embodiment to be described, an RDBMS is used, and code samples are given in generic SQL (Structured Query Language).

[0046] In order for persistence of relations, actions and metadata to be realised in an RDBMS, normalisation of the data into database tables needs to be done. Simplistically speaking, the database must hold information on a number of PERSONS, the relations between PERSONS, and data on their actions (e.g. purchase products, rate products, store documents, write articles etc.).

[0047] To that end, figure 3 illustrates the key PERSON and PERSON_TO_BUDDY tables. Keys within the table are indicated by the hash symbol.

[0048] In addition to information on individuals, the database also stores information on actions. (The exact nature of an action, and the information to be stored, will depend upon the application, but typically an "action" represents something that has been done by an individual, or statistics on a plurality of such events which may be of interest to the search initiator.)

[0049] For the purpose of illustration, we will describe an online bookstore example. However, it should be understood that the same principles apply whether actions are for an article rating, a URL search, for a skills match, etc. In each case, appropriate action and look up tables need to be added to the database in a similar manner to those described below. Such a task is within the capabilities of any database designer.

[0050] In an RDBMS, actions are stored within specific database tables such as those illustrated in figure 4. The actions here specifically relate to things being done by a person in relation to a specific book, such as purchasing it, reviewing it, rating it and so on. Metadata will typically be stored as well, for later retrieval, such as the date that the book was purchased, reviewed or rated.

3.2 Add/Modify/Delete Relations, Actions and Metadata

[0051] Once the basic schema has been defined in a database specific format, a mechanism to insert (add), update (modify) and delete relations, actions and metadata needs to be provided.

[0052] Because the bookstore example uses a relational database, SQL provides a process to achieve this requirement through INSERT, UPDATE and DELETE keywords. e.g. to create a relation from one buddy to another, the following SQL could be used:

```
INSERT INTO   PERSON_TO_BUDDY
              (ROOT_ID_FROM, ROOT_ID_TO, FRIEND_SINCE, NUM_MSGS, ...)
VALUES        (id, id2, DATE, 0, ...);
```

[0053] In the above, the elements id and id2 are provided as parameters.

3.3 Marshal In Memory Views

[0054] Bringing results to queries back from the database (i.e. as result items) requires an in-memory view of the PSN to be constructed in computer memory. This method is described in section 4.9. Suffice to say however that a working storage apparatus is needed for this process.

3.4 Publicly Volunteer Actions

[0055] Actions are classified by type and encoded in a client message header (or similar) from the client (search initiator) to the ABQ server. For example, ACTION_BUDDY_IM refers to a standard message from one buddy to another. ACTION_REFERRAL and ACTION_PURCHASE refer to a specific class of action.

[0056] Application specific action classifications can be derived according to the requirements of the implementation. In the bookstore example however, ACTION_PURCHASE_BOOK is used to distinguish this type of action from others,

e.g. ACTION_REFERRAL_URL for a URL referral, ACTION_REFERRAL_DOC for a document referral or ACTION_REFERRAL_ARTICLE for an article referral etc. Action types are embedded in a message header and handled in an ABQ server process. Thus an appropriate apparatus for encoding messages needs to be in place.

[0057] Preferably a community of users needs to consent to publicly volunteering actions. This has obvious privacy issues and, as such, implementers of the system may define their own policies and implementation rules. However, assuming an 'open' approach to information sharing is adopted (without exposing sensitive details such as account details etc.), the ABQ framework needs to capture and store publicly volunteered information on community actions.

[0058] Actions are volunteered either 'directly' or 'indirectly'. For example, a 'referral' from one PERSON to another is an example of a direct action. In a referral (e.g. book, article, product, service, URL etc.), a PERSON sends a specific type of action message to a PERSON ('buddy'), which then gets encoded in the message header (e.g. ACTION_REFERRAL_BOOK).

[0059] An example of an indirect action is a rating. A rating is either applied to an artefact or document (e.g. a book, product, article etc.) or to a PERSON (e.g. rating proficiency in a skill etc.). That is to say, indirect actions are applied to information or a PERSON without being passed to another PERSON. These are also encoded in the action field of the header as ACTION_RATING_BOOK,
ACTION_RATING_ARTICLE or ACTION_RATING_SKILL etc.

[0060] The derivation of action types is application specific, but the apparatus should typically classify actions by type and allow the community to volunteer actions publicly. Of course actions can be performed privately; however, the more that is shared, the more productive the system becomes on the whole.

3.5 An Apparatus to Create ABQ's

[0061] In compiling the result list, all actions need to pass through a central process (the ABQ server 12 in figure 1) in order for the system to act on them. Actions not involving any data storage or specific processing can simply be routed onto the PERSON recipient(s), e.g. ACTION_BUDDY_IM. However, other direct and indirect actions may require the system to respond. One example of this in the online bookstore system is the action of purchasing of a book (identified by the ACTION_PURCHASE_BOOK action type). This requires the system to make an association between the BOOK and PERSON. For example, following the server receiving and decoding this type of action, the following SQL would be executed, with book_id and root_id being provided as parameters:

```
INSERT INTO   BOOK_TO_PERSON
              (BOOK_ID, ROOT_ID, PURCHASE_DATE, ...)
VALUES        (book_id, root_id, DATE, ...);
```

[0062] Once an apparatus for sending, processing and storing data on actions is in place, further apparatus relating actions back to result items needs to be considered.

[0063] An apparatus to capture the initial search string (e.g. from a web form) is required. This can be a simple edit box on a web page with a server script to process the input. Information from client-side cookies (or similar) can extrapolate the search initiator's ROOT_ID and prepare the QUERY_STRING for the server CGI script. The server script will then parse the QUERY_STRING and construct a standard expression, transposed into SQL (in the RDBMS case).

[0064] Within the server scripts lie the rules for ABQ's as laid out in section 4, where details of the algorithms are contained. Suffice it to say at this stage that there should be:

- a server-based apparatus (database) to store information on persons, relations and actions;
- a server-based apparatus to process input keyword queries (e.g. CGI scripts) and;
- a process to present the result list of result items back to the initiator (e.g. CGI + DHTML).

3.6 Affinity Filtering of Result Items

[0065] Stored actions become result items once they are affinity 'filtered' and 'ordered'. Until then, they exist only in the database as actions. Once a keyword search has been composed by a search initiator, a standard database SELECT query (or similar) extracts a number of keyword matches from the action tables. For example, in our book system scenario, the search initiator could enter the word 'Bronte' as a search string. A simple keyword lookup in the database (using the BOOK, BOOK_TO_KEYWORD and KEYWORD action tables from our example) may then produce the following keyword matches:

- *Charlotte and Emily Bronte : The Complete Novels : Jane Eyre, Wuthering Heights, Shirley, Villette, the Professor -- Charlotte Bronte (Contributor), Emily Bronte; Hardcover*

- *Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback*
- *Wuthering Heights -- Emily Bronte; Paperback*

[0066] While providing valid keyword matches, this example has not has not yet taken into account any filtering with respect to 'affinities' in the PSN. Therefore, before the result items are ordered and presented back to the search initiator, keyword matches need to be filtered and then affinity ordered.

[0067] To achieve affinity filtering, the ROOT_ID of the search initiator needs to be provided (e.g. through cookies again) as does the 'distance' to search in the PSN (i.e. the ROOT_ORDER). If for example ROOT_ORDER = 3, the search initiator is prepared to extend the ABQ to a 'buddy of a buddy of a buddy' in their PSN.

[0068] Once the in memory view of the PSN has been constructed (see section 4.9), each of the keyword matches from the above list can be compared to each node in the in memory view. By extracting the ROOT_ID from each node in the PSN and repeatedly checking the return code from a SELECT query, a filtering of results matching PERSON's in the PSN to elements in the keyword match list can be made to produce filtered result items, e.g. (for each PSN in memory view node). In the following, book_id and root_id are provided as parameters:

```
SELECT    BOOK_ID
FROM      BOOK, BOOK_TO_PERSON, BOOK_REFERRAL
WHERE     BOOK.BOOK_ID=book_id
AND       BOOK.BOOK_ID=BOOK_TO_PERSON.BOOK_ID        // purchase action
AND       BOOK_TO_PERSON.ROOT_ID=root_id;
```

[0069] If the return code from this SELECT query indicates success, keyword matches can be categorised as being a filtered result item or; if the SELECT returns false, a filtered keyword match.

[0070] Filtered keyword matches can be ordered according to the matching rules applied by the default keyword ordering system, or whichever other system is employed.

[0071] From the book example, let us assume that the two 'Wuthering Height' books have been purchased by members in the search initiators PSN where ROOT_ORDER <= 3 from the search initiator. Then, following the affinity filtering process, the following lists are produced:

Filtered result items

[0072]

- *Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback*
- *Wuthering Heights -- Emily Bronte; Paperback*

Filtered keyword matches

[0073]

- *Charlotte and Emily Bronte : The Complete Novels : Jane Eyre, Wuthering Heights, Shirley, Villette, the Professor -- Charlotte Bronte (Contributor), Emily Bronte; Hardcover*

[0074] The next phase for the Filtered result items list is to affinity order them and form the result list.

3.7 Affinity Ordering of Result Items

[0075] Affinity ordering the result items depends on a user preference. Results can either be ordered by rating, relation or result action (see section 3.8). Rating ordering however, looks at the filtered result items and orders them according to the highest ratings applied by referees (see section 4.11), e.g.

Filtered and ordered result items (ordered by rating)

[0076]

- *Katie, who knows Nick purchased Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback, rating it 4.5/5*
- *Nick purchased Wuthering Heights -- Emily Bronte; Paperback, rating it 3/5*

[0077] Relation ordering orders the filtered result items by the strength of a relation, i.e. the 'edge weight' between referee nodes in the PSN (see section 4.12).

Filtered and ordered result items (ordered by relation)

[0078]

- *Nick purchased Wuthering Heights -- Emily Bronte; Paperback, rating it 3/5*
- *Katie, who knows Nick, purchased Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback, rating it 4.5/5.*

[0079] Result action ordering sorts the filtered result items according to how many result actions have been followed by the referee's past results (see sections 3.8 and 4.13 for more on result action ordering).

[0080] The now affinity 'filtered and ordered' result items are now referred to as the result list and can be encoded and presented back to the search initiator (see section 4.14).

3.8 Feeding Back Result Actions into the database.

[0081] Not all implementations of an ABQ system will necessarily include feedback. However, its usefulness is seen to be: (i) as another term by which to affinity order result items (see section 3.7) and; (ii) as a means of both rewarding and motivating a community to publicly volunteer actions (e.g. 'this week's most popular referee was **XXX**' etc.). Once the result list has been filtered, ordered, encoded and presented, and following a search initiator selecting a result action (i.e. following a result item), feedback can take place.

[0082] An apparatus for feeding back result actions into the database involves extracting the result identifier and referee identifier from the result item chosen by the search initiator. Once result and referee identifiers have been extracted, an SQL SELECT statement (or similar) can be formed to reward the referee by incrementing a reserved field in the database relating to the referee's node.

[0083] From figure 4, the BOOK_TO_PERSON table contains a RESULT_ACTIONS column: this is the reserved field used to count the number of result actions taken by search initiators of ABQ's.

[0084] For example, if the result action followed a result path in response to a buddy's action of purchasing a book, then the BOOK_TO_PERSON.RESULT_ACTIONS field would be updated following the result action, e.g.

```
UPDATE   BOOK_TO_PERSON
SET      BOOK_TO_PERSON.RESULT_ACTIONS=BOOK_TO_PERSON.RESULT ACTIONS+1
WHERE    BOOK_TO_PERSON.BOOK_ID=result_id
AND      BOOK_TO_PERSON.ROOT_ID=referee_id;
```

[0085] Where result_id and referee_id are provided as parameters.

[0086] Therefore, the ABQ apparatus should encode referee and result identifiers in the result list in order to reward a referee. Considering the bookstore system example, the result identifier is the BOOK_ID and the referee identifier is a ROOT_ID. The presentation of the results would contain elements (e.g. XML tags) identifying the result identifier (in this case, the BOOK_ID) and the referee identifier (e.g. the BOOK_TO_PERSON.ROOT_ID field) for each of the result items held in the result list, e.g.

```
<?xml version='1.0'?>

<!DOCTYPE ABQ SYSTEM "abq.dtd">
<results>
 <title>ABQ Results</title>
 <resultlist>
    <resultitem path="http://some.domain.com/books/bronte.html#wh">
       <resultaction>../cgi-bin/resultAction.pl</resultaction>
       <referee id="567321">Katie</referee> who knows
       <relation id="567274">Nick</relation> purchased
       <result id="984527">Wuthering Heights</result> rating it
       <rating>4.5</rating> out of 5
    </resultitem>
    ...
 </resultlist>

    </results>
```

[0087] A result item path (in the book example) is an XML tagged URL pointing to a web page about the book, and a result action would be the search initiator clicking on one of the result paths. After which, SQL such as that in section 3.8 can be run once the required tags have been extracted from the result item.

[0088] Apparatus to encode referee and result identifiers in the result should be provided (XML tags are one solution). Additionally, apparatus to capture result paths taken by search initiators (e.g. parameter passing back a CGI QUERY STRING on the server from a web browser). Finally, apparatus to process the result path (from the result action) needs to update the NUM_REFERRALS fields for the referee and present the result endpoint to the search initiator (e.g. a CGI server script). The result endpoint is considered (in the book example) a web page describing the book. Other examples of result endpoints are: (i) product pages; (ii) articles; (iii) email addresses; (iv) newsgroup articles; or (v) URL's and documents etc.

4. Methods

[0089] The following sub-sections describe the methods required to build the example book ABQ system. Details such as user interface have been omitted, as have references to specific Internet browsers or other software etc.

[0090] While some of these specific details of the methods described relate particularly to the example book ABQ system, the underlying principles are generic and it is well within the capabilities of a skilled person in the field to modify the methods shown for use in other applications. In effect, all that is required to implement another system is to capture information about referees, results, actions and metadata in any convenient way, and to modify the methods described below to handle that information in an appropriate manner.

4.1 A method to create the ABQ database schema

[0091] The first step is to create a database. Standard ANSI SQL can achieve this (assuming an RDBMS is being used). Once this is done, the ABQ schema needs to be defined. This is best illustrated through code. The following SQL illustrates schema creation for the online bookstore system (see figure 4)

```
USE ABQ;                                                    // assumes DB = 'ABQ'
CREATE TABLE PERSON                                         // Person table
(
     ROOT_ID           INT UNSIGNED PRIMARY KEY,            // primary key
     FIRST_NAME        VARCHAR,
     LAST_NAME         VARCHAR,
     EMAIL             VARCHAR,
     ...                                                    // other fields
);
CREATE TABLE PERSON_TO_BUDDY                                // Buddy table
(
     ROOT_ID_FROM      INT UNSIGNED,                        // foreign key
     ROOT_ID_TO        INT UNSIGNED,                        // foreign key
     FRIEND_SINCE      DATE,
     NUM_MSGS          INT UNSIGNED,                        // used for ordering
     ...                                                    // other fields
);
CREATE TABLE BOOK                                           // Book table
(
     BOOK_ID           INT UNSIGNED PRIMARY KEY,            // primary key
     TITLE             VARCHAR,
     AUTHOR            VARCHAR,
     PUBLISHER         VARCHAR,
     SUMMARY           LONG VARCHAR,                        // summarised précis
     ENTRY_DATE        DATE,
     ...                                                    // other fields
);
CREATE TABLE BOOK_TO_KEYWORD                                // Book keyword table
(
     BOOK_ID           INT UNSIGNED,                        // foreign key
     KEYWORD_ID        INT UNSIGNED,                        // foreign key (hash)
     ...                                                    // other fields
);
CREATE TABLE KEYWORD                                        // Keyword table
(
     KEYWORD_ID        INT UNSIGNED PRIMARY KEY,            // primary key (hash)
     KEYWORD           VARCHAR,
     ...                                                    // other fields
);
CREATE TABLE BOOK_TO_PERSON                                 // An 'action' table
(
     BOOK_ID           INT UNSIGNED,                        // foreign key
     ROOT_ID           INT UNSIGNED,                        // foreign key
     PURCHASE_DATE     DATE,
     REVIEW            LONG VARCHAR,                        // used for reviewing
     RATING            INT UNSIGNED,                        // used for filtering
     RESULT_ACTIONS    INT UNSIGNED,                        // used to reward
results
     ...                                                    // other fields
);
```

4.2 A method to store referees (PERSON's) in a database

[0092] Once the database and schema is constructed, the system waits for PERSON 'registrations' and 'logins',(see figure 5) This may be realised through an Instant Messaging client interface or extrapolated from information known to hold information on a Personal Social Network - e.g. through scanning closed community email address books - or through a more open online community (e.g. a newsgroup) where actions between PERSON's can be intercepted and processed by the ABQ server.

[0093] In the bookstore system example, an Instant Messaging client could be used to accompany the ABQ service and provide the raw relation data for the PSN.

[0094] The IM system works in conjunction with an Internet browser to monitor actions and act on them. Therefore, the following process outlines joint REGISTRATION and LOGIN methods, adding user details into the PERSON table in the database in the case of new registrations. For clarity, exception, primary key and handling processes etc. are omitted from figure 5.

4.3 A method to modify PERSON records

[0095] Should a PERSON wish to alter their details (e.g. their email address), a method allowing them to do so must be in place. This can be realised for example through the following SQL.

```
UPDATE    PERSON
SET       PERSON.EMAIL=email
WHERE     PERSON.ROOT_ID=root_id;
```

[0096] Figure 6 shows the process in flow chart form.

[0097] This assumes that the user has already registered with the ABQ service and has been allocated a unique ROOT_id. There after, the user is logged in and is known to the system (see section 4.2).

4.4 A method to delete PERSON records

[0098] Should a PERSON wish to delete their details completely, a method allowing them to do so must be in place. This can be realised for example through the following SQL.

```
DELETE FROM    PERSON
WHERE          PERSON.ROOT_ID=root_id;
```

[0099] Figure 7 shows the process in flow chart form. As with figure 6, this assumes that the user has already registered.

4.5 A method to store relations between PERSON's

[0100] Following a process of 'find buddy' (i.e. a method allowing users to input values to search the database for buddies), a selected buddy is inserted to the PERSON_TO_BUDDY table using SQL (e.g. see listing under section 3.2).

[0101] The process is shown in flow chart form in figure 8. At box 80, a number of different terms may be used to find buddies. In its simplest form the PERSON may know the ROOT_id of the buddy they wish to add. In a more complex example, the PERSON may know only the surname, or some other partial details.

[0102] For clarity, no reference has been made here to other tables which might be present in a practical example. These could include a PERSON_TO_IGNORE table which has similar properties to the PERSON_TO_BUDDY table and which provides a mean of ignoring actions from unwanted individuals within a community.

4.6 A method to delete relations

[0103] As well as adding buddies into a buddy list (or similar), a PERSON may want to remove a buddy from a buddy list. This could be done with the following SQL:

```
DELETE FROM    PERSON_TO_BUDDY
    WHERE          PERSON_TO_BUDDY.ROOT_ID_FROM=root_id_from
    AND            PERSON_TO_BUDDY.ROOT_ID_TO=root_id_to;
```

[0104] Figure 9 shows the process in a flow chart form.

4.7 A method to summarise and store actions

[0105] Actions from PERSON's are preferably summarised. This is so that keywords from the actions can be extracted and stored in a keyword table (for keyword matching). Summaries are stored along with action records, as are other data/metadata pertaining to the action.

[0106] The specifics of action summarisation are not part of the invention and various options are available. Summaries and keywords can be verbally agreed, provided (e.g. by a publisher in the example bookstore system) or extracted automatically (e.g. from a content or document summariser).

[0107] Following an action, a summarised version of the action is stored along with the data, metadata and keywords about it.

[0108] Figure 10 illustrates the method in flow chart form.

[0109] At box 100, an action key is generated or obtained. Keys can be generated in a number of ways (for example

by the use of hash values): any convenient approach can be used. At box 110, there is a check made of the last action summary date. This process allows for already-summarised actions to be re-summarised if they are out of date. This may be particularly necessary for volatile data, for example on the web. Box 120 illustrates obtaining the action summary and key words: this may either be provided manually or, preferably, through an automated content summariser that reads documents or texts and that provides summaries and key words. With particular reference to the bookstore system, at box 130 the action would be stored in the BOOK_TO_PERSON or the BOOK_REFERRAL tables.

4.8 A method to process incoming ABQ parameters

[0110] A generic approach to the processing of incoming ABQ parameters is shown in figure 11. At box 140, the ROOT_ID is obtained; this may be through the use of cookies (if a web browser is in use), or obtained from a registry setting, or from elsewhere. The encoding step shown at box 150 may use a bespoke TCP/UDP application header, or may encode the values in a QUERY_STRING. Further details of the items shown within the dotted line 160 may be found at section 4.10.

4.9 A method to build in memory views

[0111] In memory views require building the PSN digraph in computer memory. To do this, relations from the database need to be built with a number of abstract data types. There are numerous ways to do this and the following is but one example using a 'list of lists' (modelling a digraph).

[0112] However, the principle and methods of the flowchart (figure 12) remains the same irrespective of whichever implementation method of holding the in memory view is chosen.

[0113] At box 200 of figure 12, the next ROOT_ID is extracted. This is the ROOT_ID of the next node in the PSN to look up in the database. In the first loop it is the ROOT_ID of the service initiator.

[0114] Figure 13 below shows how by iteratively extracting PERSON 'buddies' from the database leads to the production of an in memory view of the PSN digraph (binary tree description follows later).

[0115] Nodes are internally represented by 'objects' and edges encapsulate pointers or object references into other objects. By implementing the digraph using a doubly linked list (for example), a path back to the root node from any other node in the digraph can be traced. This is used to determine the PSN relations (e.g. 'Kate, who knows Nick...' etc.)

[0116] Nodes are placed in the graph in sequential order as read from tuples returned from a SELECT query. Each level of indirection in the graph relates to another ROOT_ORDER from the root node (i.e. 'ME'). This structure is built in computer memory from repeated SELECT queries iterating out from the root node. Nodes appear in a number of places due to the nature of the PSN.

[0117] Of course, tracing directly into the digraph (and avoiding duplicates) requires another process. This can be implemented through the use of a binary tree (using ROOT_ID's of each PERSON as the tree node value). A fast method of getting into the digraph can be achieved by holding pointers to digraph objects inside a binary tree, as shown in figure 14.

[0118] In figure 14, the numbers before the names in the binary tree indicate ROOT_ID values from the database.

[0119] The use of the binary tree allows objects to enter the digraph once only. That is to say, relations retrieved from the database are first checked if they appear in the binary tree. If so, they are discarded (as a 'shortest path' to them has already been computed — apart from reference to the root node). If not, they are added to the digraph and their pointer is added into the corresponding ROOT_ID node location in the binary tree for fast access into the digraph.

4.10 A method to affinity filter ABQ keyword matches

[0120] Once the in memory view is built, the next phase is to filter the keyword matches into filtered result items and filtered keyword matches (see section 3.6). The process is shown in figure 15.

[0121] At box 210, we can see obtain keyword matches. This of course assumes that we have managed to find keyword matches to the search initiator's ABQ keyword list-see section 4.8.

[0122] Following this process, keyword matches have been processed into two lists: (i) filtered keyword matches (ordered by keyword) and (ii) filtered result items (also ordered by keyword). The next phase is to affinity order the filtered result items.

4.11 A method to affinity order result items by rating

[0123] Now that we have filtered the results into filtered keyword matches (no longer considered until encoding and presentation) and filtered result items, we need to order them.

[0124] Ordering by rating involves sorting the filtered result items according to ratings applied by PERSON's in the

PSN. Although filtering and ordering can be seen as two separate tasks, to reduce the number of database accesses, the in memory view can be prepared according to the method of ordering selected by the search initiator during filtering. For example, if the search initiator chose to order by rating, then whilst filtering the results (section 4.10), the database query could extract information on ratings applied at the same time as filtering, as in the following SQL:

```
SELECT   BOOK_ID, BOOK_TO_PERSON.RATING
FROM     BOOK, BOOK_TO_PERSON
WHERE    BOOK.BOOK_ID=book_id
AND      BOOK.BOOK_ID=BOOK_TO_PERSON.BOOK_ID
AND      BOOK_TO_PERSON.ROOT_ID=root_id;
```

[0125] Combining the filtering method with collecting information useful for ordering the result list can drastically speed up performance during the ordering phase. This is because one database read per PSN referee for filtering places 'ordering' information into the in memory view allowing traversal of memory to take place rather than repeated calls (and locks) to the database.

[0126] Considering that we already have the rating information in the in memory view, we can proceed rating order as shown in figure 16.

[0127] The ordered linked list in figure 16 is now in effect a linear version of the in memory view, ordered by PERSON's who have rated a result item. Thus, looping through each result item and repeatedly sorting the ordered linked list (according to each result item) and copying the results into another list produces a filtered and ordered list of result items. This list may contain duplicate references to a result item (where more than one person has rated a result item).

[0128] It will be understood of course that the sorting may be carried out using any convenient method, including quicksort, bubblesort, mergesort, insertsort, and so on.

4.12 A method to affinity order result items by relation

[0129] As mentioned in section 4.9, each node of the in memory view holds the ROOT_ID of a PERSON so that filtering and ordering of keyword matches can take place.

[0130] The other attribute held in the digraph - for relation ordering only - is the reciprocal edge weighting (Ew) value between PERSON's.

[0131] Ew can be computed in a number of ways and derived according to the application. However, the following method is sufficient for many needs and uses derivatives of action and time in its calculation. The first part of computing Ew between two PERSON's (A→B) is to calculate the total number of actions sent from A to all buddies (where ROOT_ORDER = 1) and is given as:

$$A = \sum_{i=1}^{j} a_i \qquad \text{(eq. 4.1)}$$

where: a = the number of actions sent from A to the i^{th} buddy, where A's root order=1

[0132] The next calculation is the number of actions received by A from all buddies and is given as:

$$B = \sum_{i=1}^{j} b_i \qquad \text{(eq. 4.2)}$$

where: b = the number of actions received by A from the i^{th} buddy, where A's root order=1

[0133] Next comes calculating Mt. Mt represents the maximum time anyone has been a buddy of A and is used in calculating Aw and Bw (below). Aw represents the edge weight from A to B and is given as:

$$Aw = \frac{a/A}{Bi/Mt} \qquad \text{(eq. 4.3)}$$

where:

a = the number of actions sent from A to B

A = the sum of all actions sent from A to all buddies, given in (eq. 4.1)
Bt = the time in which B has been a buddy of A
and
Mt = the maximum time anyone has been a buddy of A

[0134] Bw represents the edge weight to A from B and is given by

$$Bw = \frac{b/B}{Bt/Mt} \qquad \text{(eq. 44)}$$

where:

b = the number of actions sent to A from B
B = the sum of all actions sent to A from all buddies, given in (eq. 4.2)
Bt = the time in which B has been a buddy of A
and
Mt = the maximum time anyone has been a buddy of A

[0135] Finally, sum all Aw and Bw values for all buddies and compute Ew between A→B as follows:

$$Ew_{AB} = \frac{Aw + Bw}{\sum_{i=0}^{j} Aw_i + \sum_{i=0}^{j} Bw_i} \qquad \text{(eq. 4.5)}$$

[0136] Completing equation 4.5 for all buddies will produce a normalised array such that:

$$\sum_{m=1}^{n} Ew_m = 1 \qquad \text{(eq. 4.6)}$$

[0137] As an example, consider the following table showing the Ew values from A→B_1 .. B_5:

	A	B	C	D	E	F	G
1	From A to	*Bt*	*a*	*Aw#*	*b*	*Bw#*	*Ew#*
2	B_1	10	10	0.438	5	0.259	0.131
3	B_2	5	10	0.875	10	1.037	0.360
4	B_3	2	5	1.094	3	0.778	0.352
5	B_4	8	5	0.273	6	0.389	0.125
6	B_5	14	2	0.063	3	0.111	0.033
7							1.000
8							
9	*note. '#' indicates a calculated property*					*A#*	32
10	*...the rest are stored in the DBMS...*					*B#*	27
11						*Mt#*	14

[0138] Thus, by traversing the in memory view and calculating Ew values for each node, we can begin to build a picture of the PSN that (in our example) could resemble figure 17.

[0139] Once all Ew values have been calculated for the in memory view a process of sorting the relationships by root order and Ew value can produce the arrangement of figure 17. Once the in memory view is ordered, it is simply a matter for the algorithm to inspect each filtered result item and traverse the (ordered) in memory view as shown in figure 19.

[0140] In that figure, box 250 requires us to obtain the next ROOT_ID: this is the ROOT_ID of the next node in the memory view according to the current root order ("X" value). So, where X equals 1 the ROOT_ID order will be 1.1, 1.2, 1.3 (from figure 1),...etc.

4.13 A method to affinity order result items by result action

[0141] This method works in a similar manner to that presented in section 4.11. The in memory view does not require ordering (as in section 4.12) as ordering is done according to how 'valuable' PERSON's in the PSN are seen. Result actions (from section 3.8) are an incremental value and get updated after result actions have been taken by search initiators (section 4.15).

[0142] However, assuming the RESULT_ACTIONS field is modelled in the action tables of the database (see listing at section 4.1), affinity ordering the filtered result items by result action takes place in an identical fashion to that in section 4.11 with the exception that the final list is ordered on result action and not rating.

4.14 A method to encode and present result items

[0143] Once the filtered and ordered result items have been compiled, the penultimate phase involves encoding and presenting the result list back to the search initiator. An apparatus for this was presented in section 3.8. In this section, an example using XML was presented. However, the specifics of encoding are left open to the ABQ application developer. Suffice to say, the format of the (pre-encoded) result list will hold information on referee, result, relation and (where appropriate) rating. These may be fields within a software class held in an abstract data type such as a linked list. The result list is then used in conjunction with the binary tree and the in memory view to look up affinities with result items. These are then pushed through an encoding and presentation process and presented back to the search initiator. This is shown in figure 20. Note that the result list holds the result identifier, the binary tree the root identifier and the in-memory view other data (for example edge weights) and relational data with respect the root node.

4.15 A method to feedback result actions

[0144] As mentioned in section 3.8, a process to feedback results action made by search initiators is the final method to explain. This method assumes an apparatus tracking which result item the search initiator has selected can feedback values to the ABQ server (figure 2).

[0145] In the bookstore example, this apparatus is a hyperlink calling an ABQ 'resultAction' server script with the following three parameters:

- RESULT_PATH — the URL of a web page describing the product (book)
- REFEREE_ID — the referee identifier (see section 3.8 and listing at section 3.6)
- RESULT_ID — the result identifier the product (book).

[0146] This is shown in figure 21.

[0147] Once the result action have been sent back into the database, the ABQ process is complete, and the system then simply waits for the search initiator, or indeed any other members of the on line community, to start another search.

[0148] In a variant of the preferred embodiment (not shown) the system could allow the output to be ordered by more than one parameter. For instance, the output could first be ordered by relation, and then, within that, by rating. Options may be provided to allow the search initiator various possibilities.

Glossary of Terms

[0149]

ABQ	Affinity Based Query
Action	An activity performed by a member of a virtual or online community (e.g. visit URL, purchase book etc.). Actions can be direct, indirect, public and/or private
Affinity Filtering	The process of qualifying keyword matches against members of a Personal Social Network (PSN).
Affinity Ordering	The process of ordering affinity filtered result items with respect to rating, relation or result action

Digraph	Directed Graph. A way of presenting the Personal Social Network using nodes and directed edges.
Edge Weight	A relationship weighting between members of a Personal Social Network
In Memory	A computer memory model of the Personal Social Network held as a
View	collection of objects in an Abstract Data Type, examples are hash tables, linked lists and arrays etc.
Keyword	A matching between an entry in the ABQ database and a search
Match	initiators keywords. Keyword matches are pre affinity filtering and ordering
Metadata	Data about data. Stored data about actions not directly concerned with the action itself (e.g. the date the action was made etc.)
PSN	Personal Social Network. A way of describing relationships with respect to a root node
Rating	Part of a result item. A grade given either to information or to oneself
Referee	Part of a result item. A person within a Personal Social Network who is associated with information being sought by a search initiator.
Relation	Part of a result item. The method of describing the relationships between a search initiator and a result.
Result	Part of a result item. A matching of an action with a referee in an ABQ
Result Action	The activity taken by the search initiator following the presentation of the result list, e.g. following a hyperlink or emailing a relation etc.
Result Item	Part of a result list. An affinity filtered keyword matched action from the ABQ database
Result List	The product of keyword matches after they have been affinity filtered and ordered
Result Path	The location of a result, e.g. URL, directory path etc.
Root Node	The nucleus of a Personal Social Network.
Root Order	A way of describing 'distances' from a root node in a Personal Social Network to other members, i.e. 1st order relationships = 1 root order from the root node etc.
Search Initiator	An ABQ community user initiating a query.

Claims

1. A computer query method comprising receiving a query from a search initiator, carrying out a search for information on the basis of the query and filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

2. A computer query system as claimed in claim 1 in which the linked network of individuals is a personal social network.

3. A computer query system as claimed in claim 1 or claim 2 including the step of presenting the filtered results as a list.

4. A computer query system as claimed in claim 3 in which the hit is sorted according to the affinity of each result with individuals within the network.

5. A computer query system as claimed in any of the preceding claims including displaying each filtered search result in conjunction with information identifying the individual within the network that has an affinity with the result.

6. A computer query system as claimed in claim 5 further including displaying in conjunction with each filtered search result information identifying the type of affinity between the individual and the search result.

7. A computer query system as claimed in claim 6 in which the affinity consists of a rating applied by the individual to the result.

8. A computer query system as claimed in claim 5 further displaying in conjunction with each filtered search result information representative of a relation between the search initiator and the individual.

9. A computer query system as claimed in claim 4 in which the list is sorted according to ratings applied to the results by individuals within the network.

10. A computer query system as claimed in claim 4 in which the list is sorted according to a type or closeness of relation between the search initiator and the individuals which are associated with the filtered results.

11. A computer query system as claimed in claim 10 in which the list is sorted so that the first order relations appear before second order relations.

12. A computer query system as claimed in claim 10 in which the network includes weights associated with person to person links, the list being sorted in dependence upon the weights.

13. A computer query system as claimed in claim 12 in which the weights are time dependent.

14. A computer query system as claimed in claim 13 in which the weights are determined based on prior interactions between individuals within the network.

15. A computer query system as claimed in any one of the preceding claims including maintaining a database representative of the network of individuals.

16. A computer query system as claimed in claim 15 including constructing an in-memory view of a relevant part of the network, and filtering the search results according to information contained within the in-memory view.

17. A computer query system as claimed in claim 15 or claim 16 including maintaining within the database information on actions associated with or done by the individuals within the network.

18. A computer query system as claimed in claim 17 including maintaining information on a referral from one individual within the network to another.

19. A computer query system as claimed in claim 17 or claim 18 including maintaining information or messages between one individual within the network and another.

20. A computer query system as claimed in any one of claims 17 to 19 including maintaining information on the purchase of an item by an individual within the network.

21. A computer query system as claimed in any one of claims 17 to 20 including maintaining information on a rating by an individual of a document, a link or a product.

22. A computer query system as claimed in any one of claims 17 to 21 including maintaining information on a rating by an individual of another, or the same, individual within the network.

23. A computer query system as claimed in any one of the preceding claims in which the search is a keyword search.

24. A computer query system as claimed in any one of the preceding claims in which the search initiator selects the extent of filtering required by varying a minimum level of affinity which will return a result.

25. A computer query system as claimed in claim 15 including the step of selecting and carrying out a result action on one of the filtered results, and storing in the database information representative of the result action against the individual who was associated with the selected result.

26. A computer query system as claimed in claim 25 in which the said information stored against the individual comprises an action value.

27. A computer query system as claimed in claim 26 when dependent upon claim 3 in which the list is sorted according to the action values of the individuals associated with the filtered results.

28. A computer query system as claimed in claim 3 in which the search is a keyword search and in which the list is further sorted by keyword.

29. A computer program for carrying out a method as claimed in any one of the preceding claims.

30. A data-carrier carrying a computer program as claimed in claim 29.

31. A computer query system comprising means for receiving a query from a search initiator, means for carrying out a search for information on the basis of the query, and means for filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

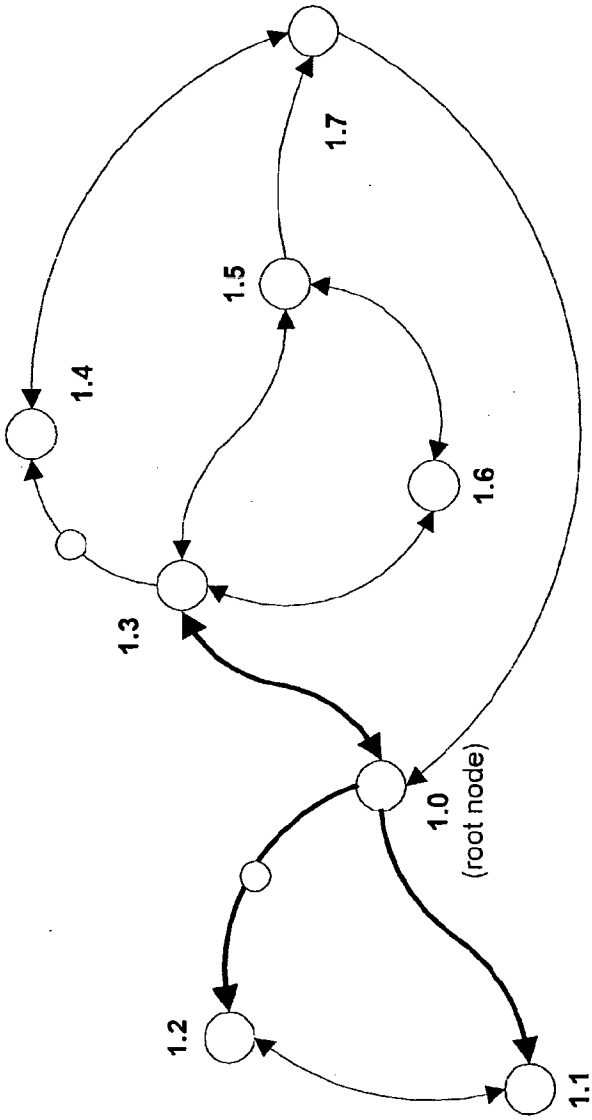


Figure 1

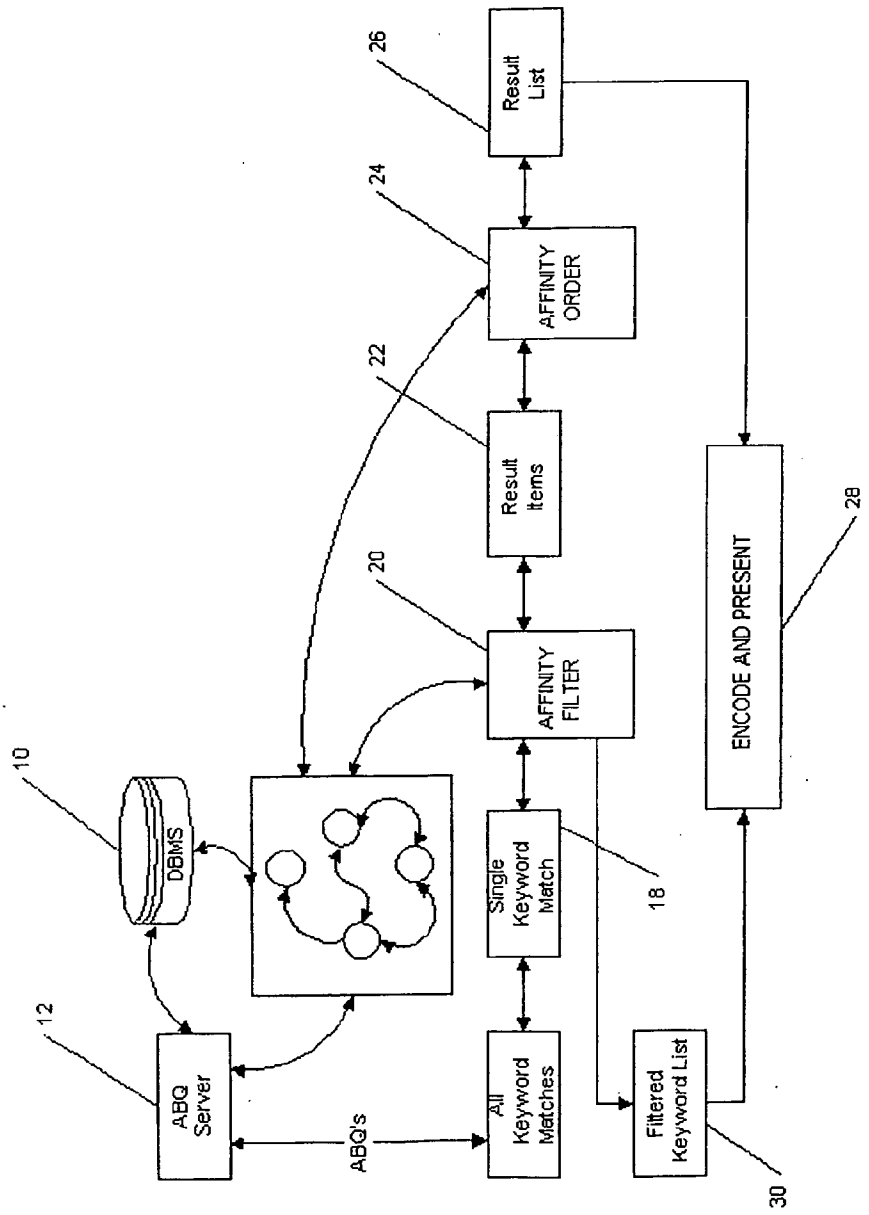


Figure 2

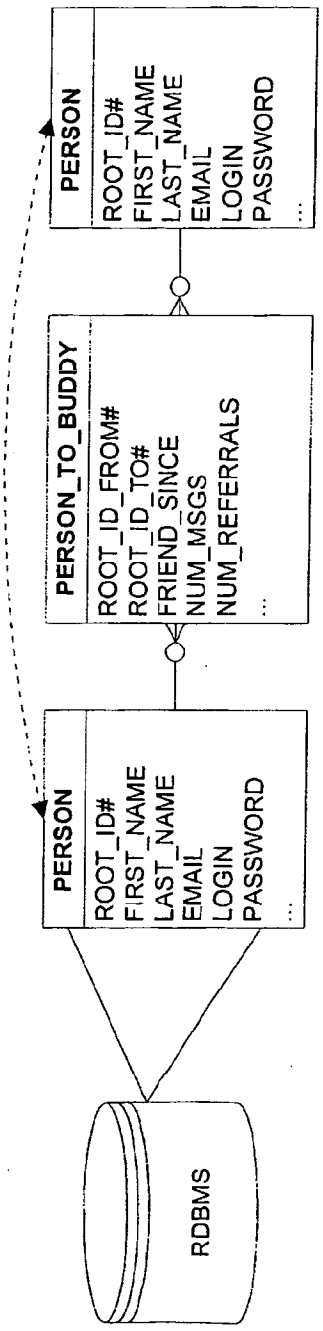


Figure 3

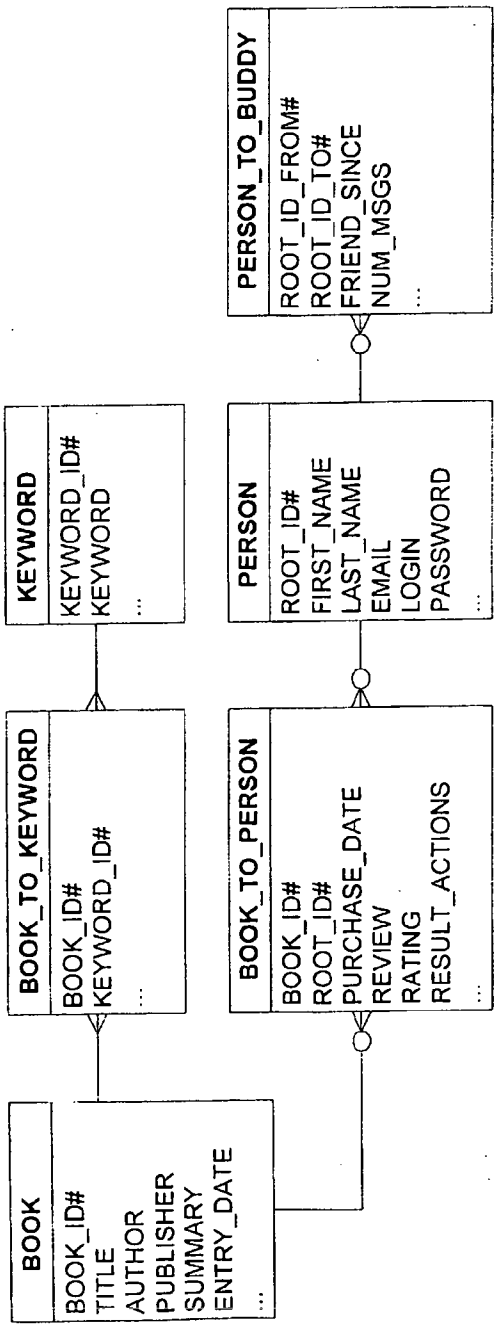


Figure 4

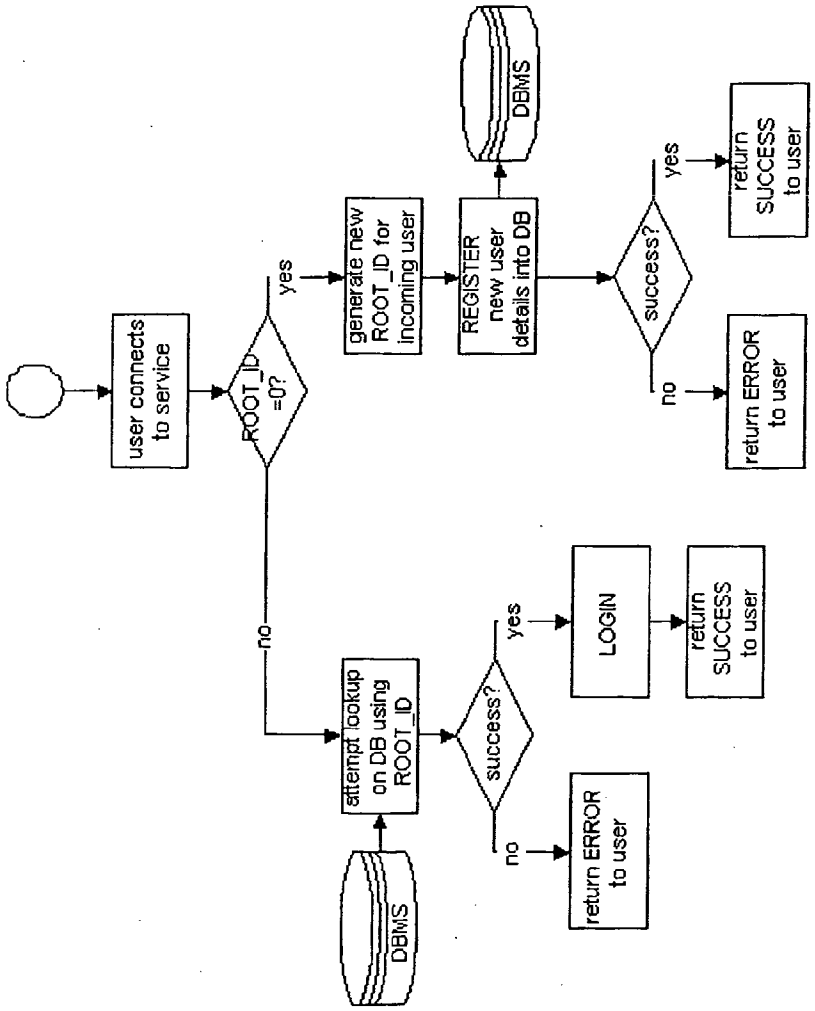


Figure 5

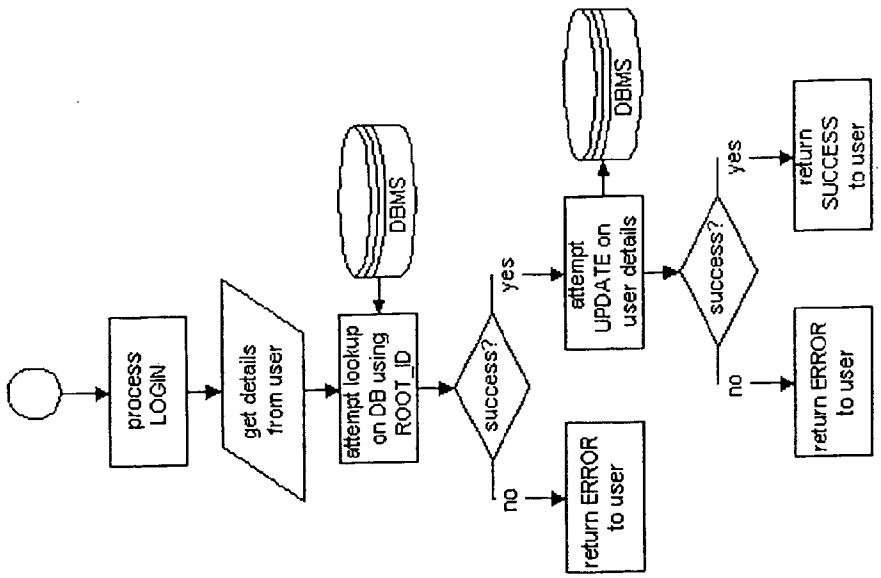


Figure 6

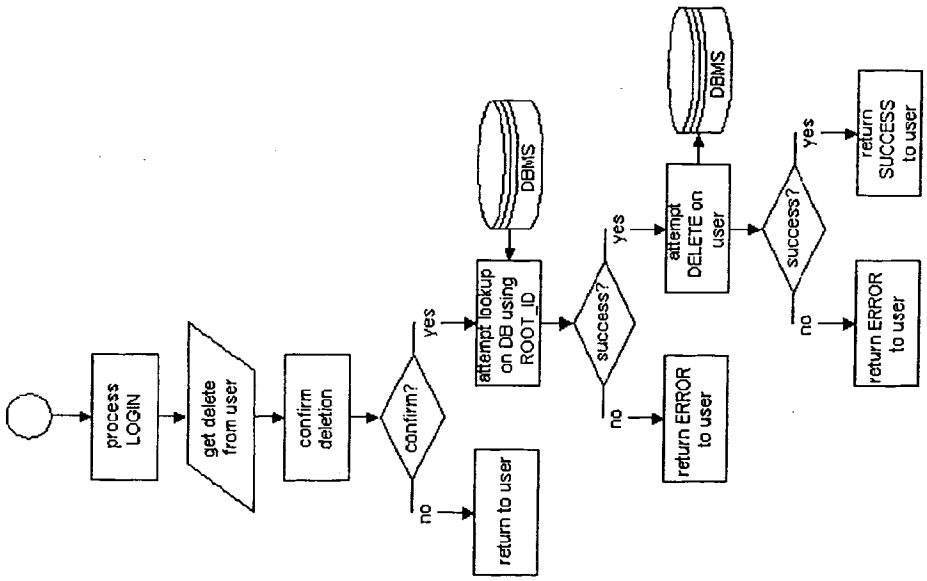


Figure 7

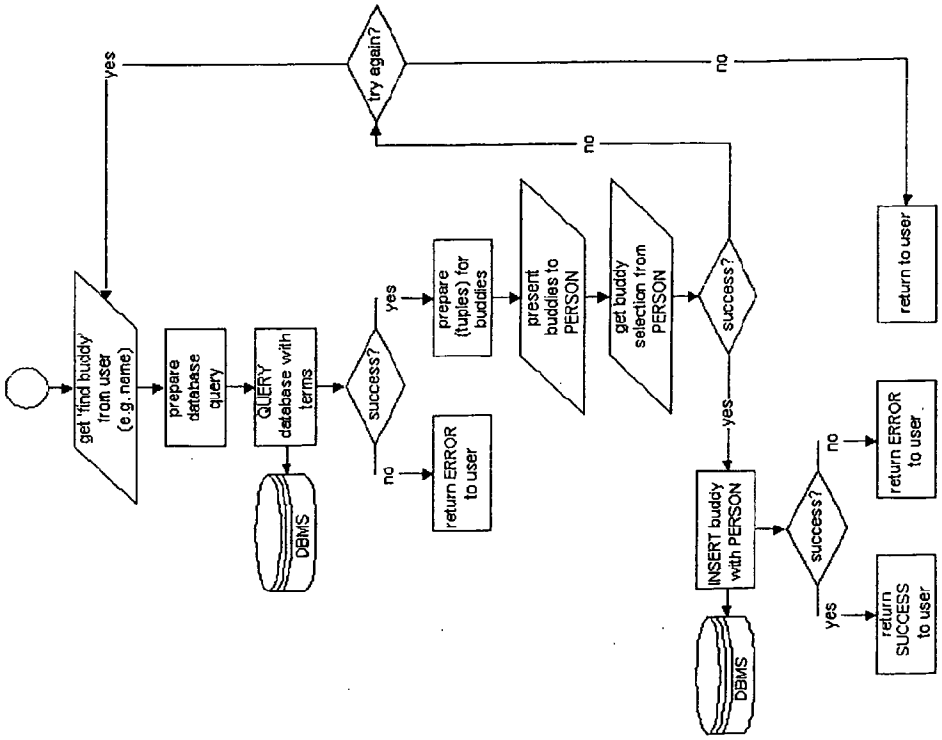


Figure 8

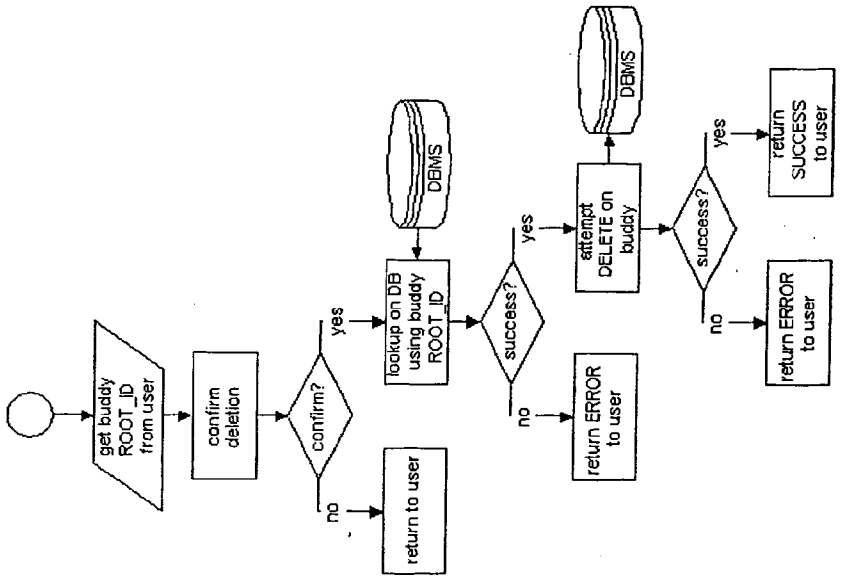


Figure 9

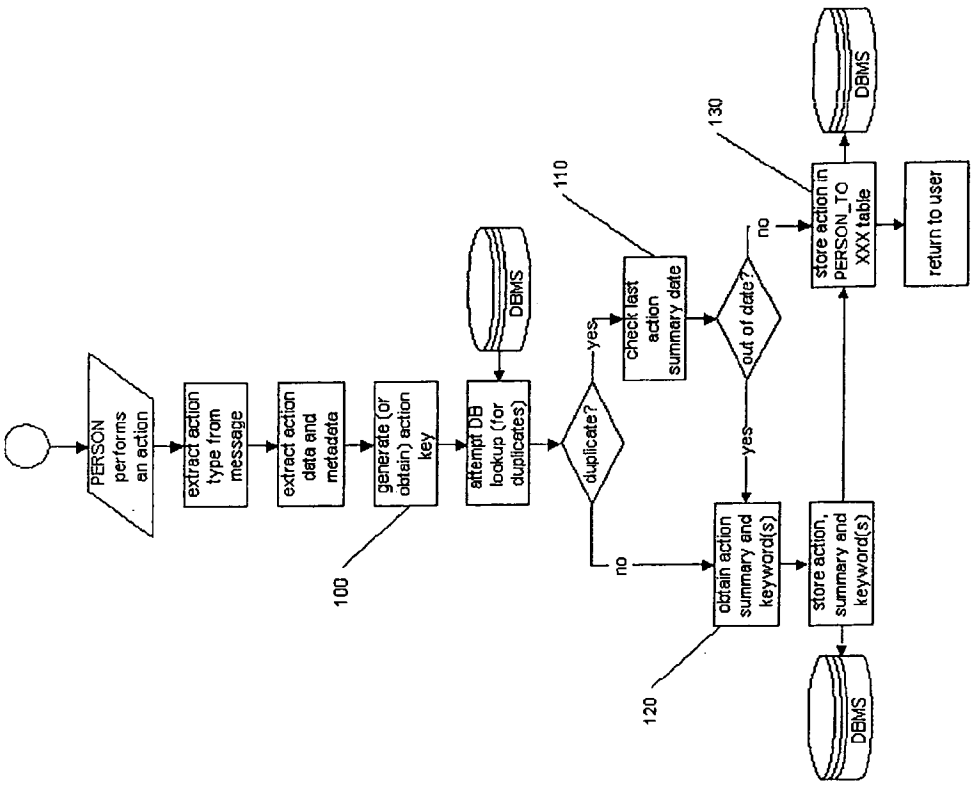


Figure 10

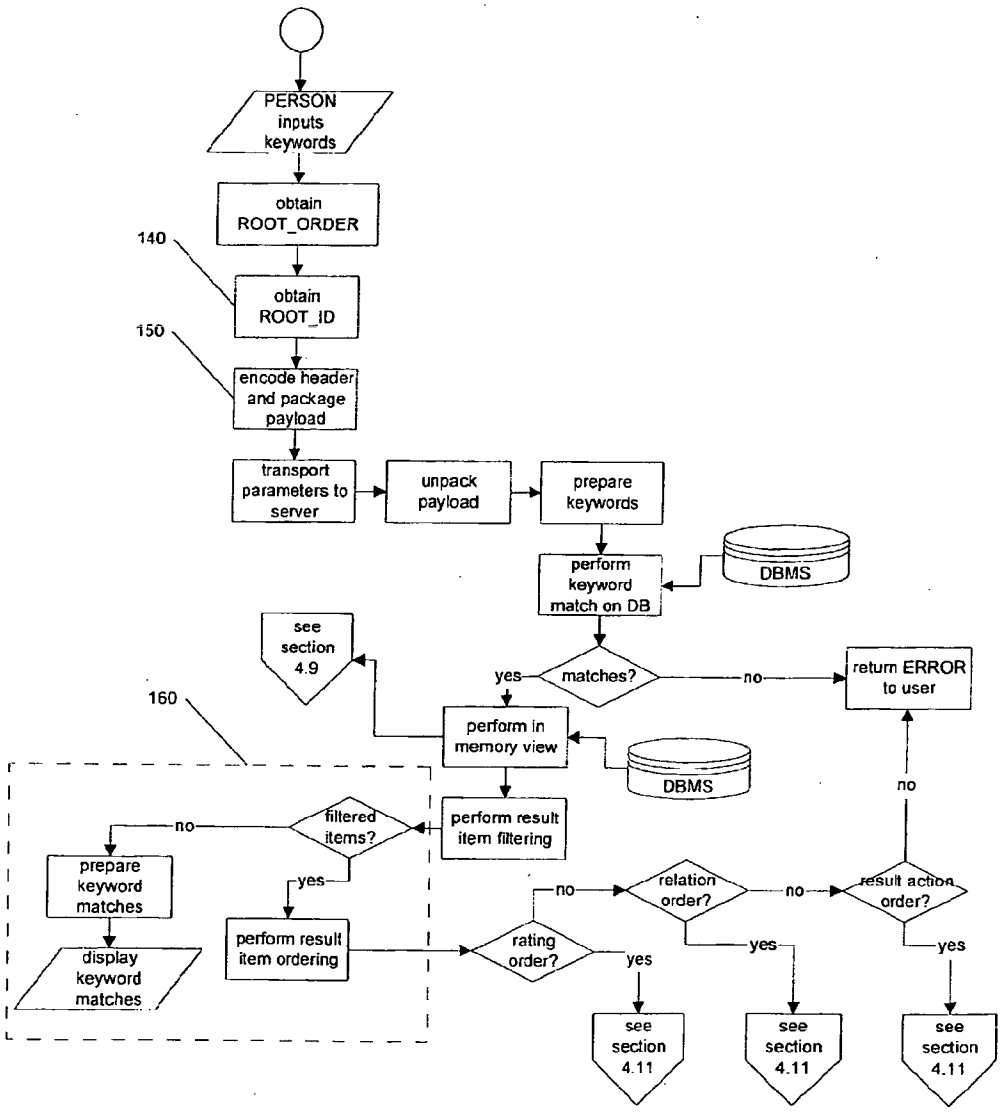


Figure 11

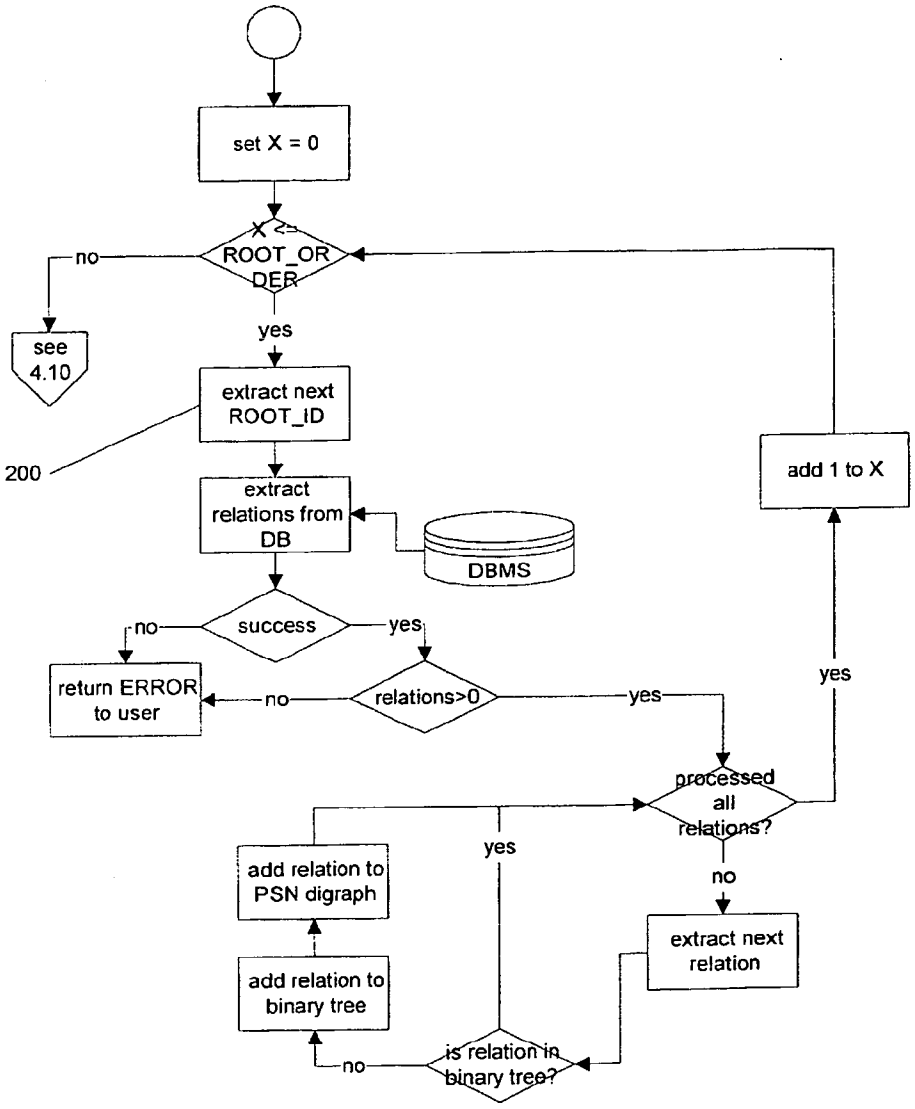

set X = 0
X <= ROOT_ORDER
no
yes
see 4.10
extract next ROOT_ID
200
extract relations from DB
DBMS
success
no
yes
return ERROR to user
relations>0
no
yes
add 1 to X
yes
processed all relations?
yes
no
add relation to PSN digraph
extract next relation
add relation to binary tree
is relation in binary tree?
no

Figure 12

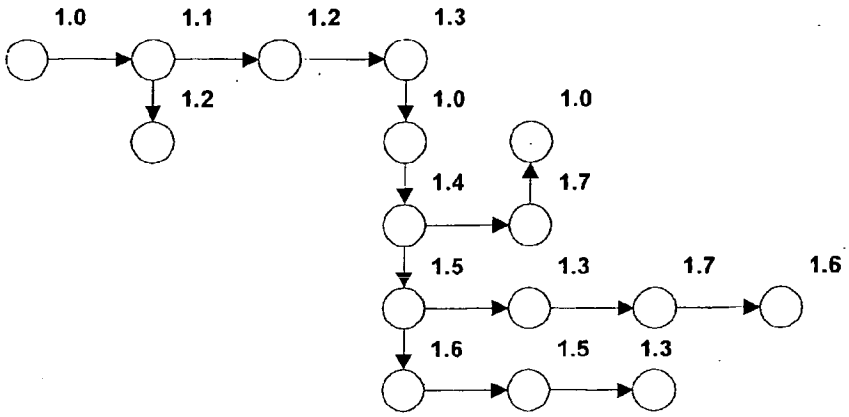


Figure 13

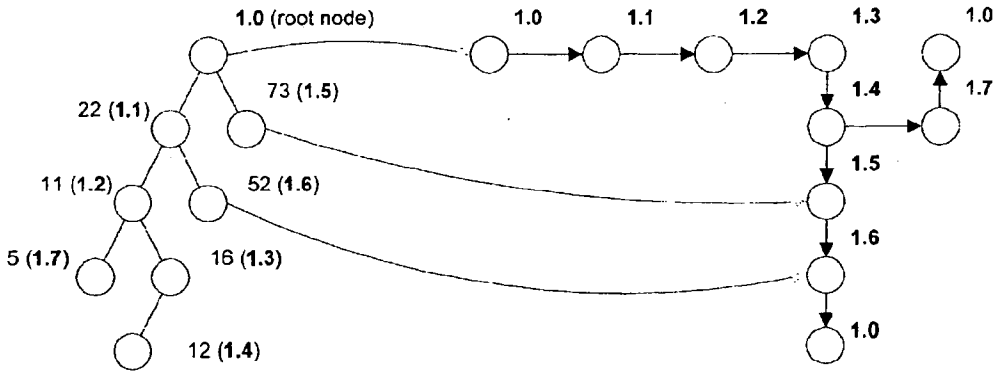


Figure 14

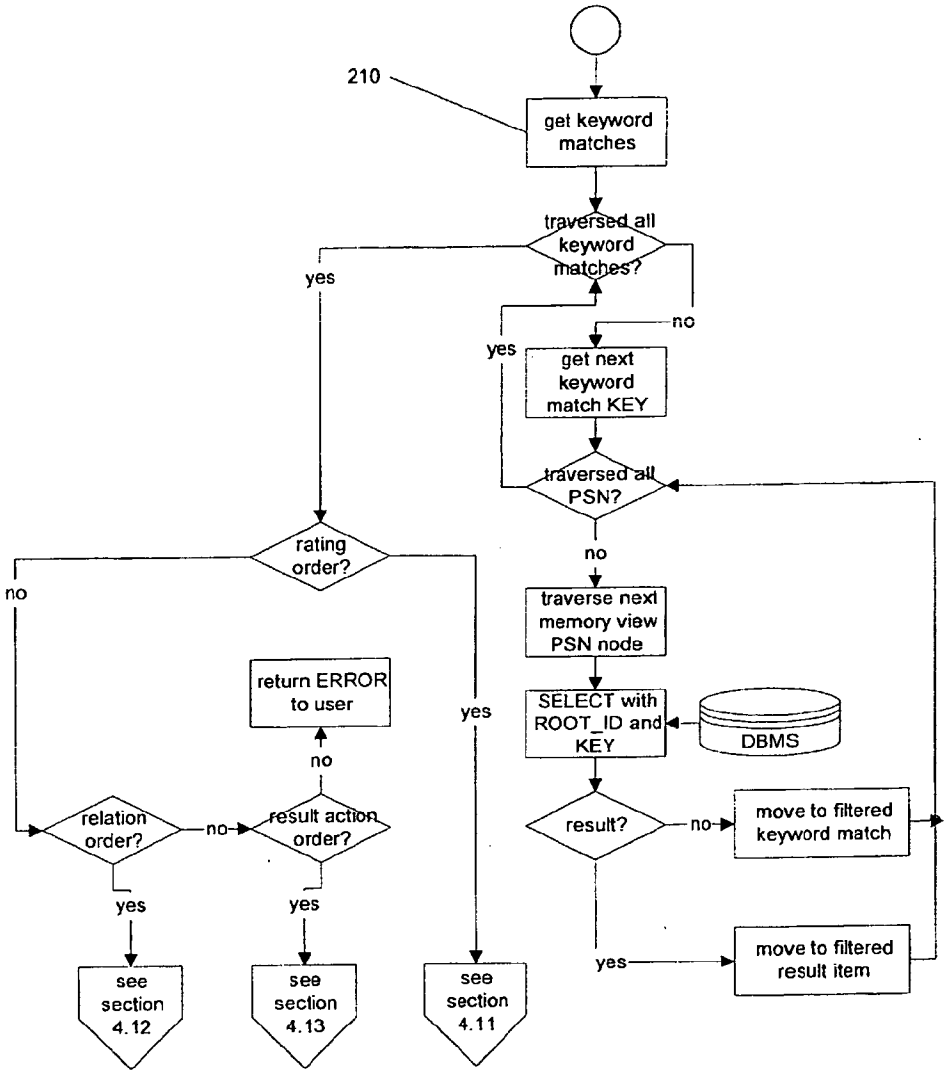


Figure 15

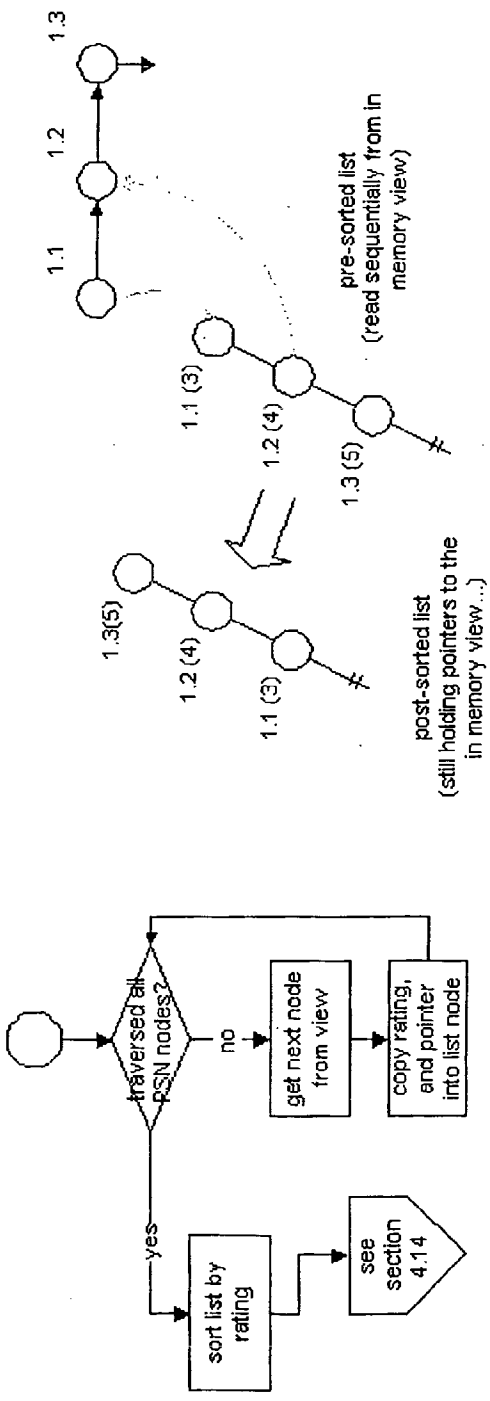


Figure 16

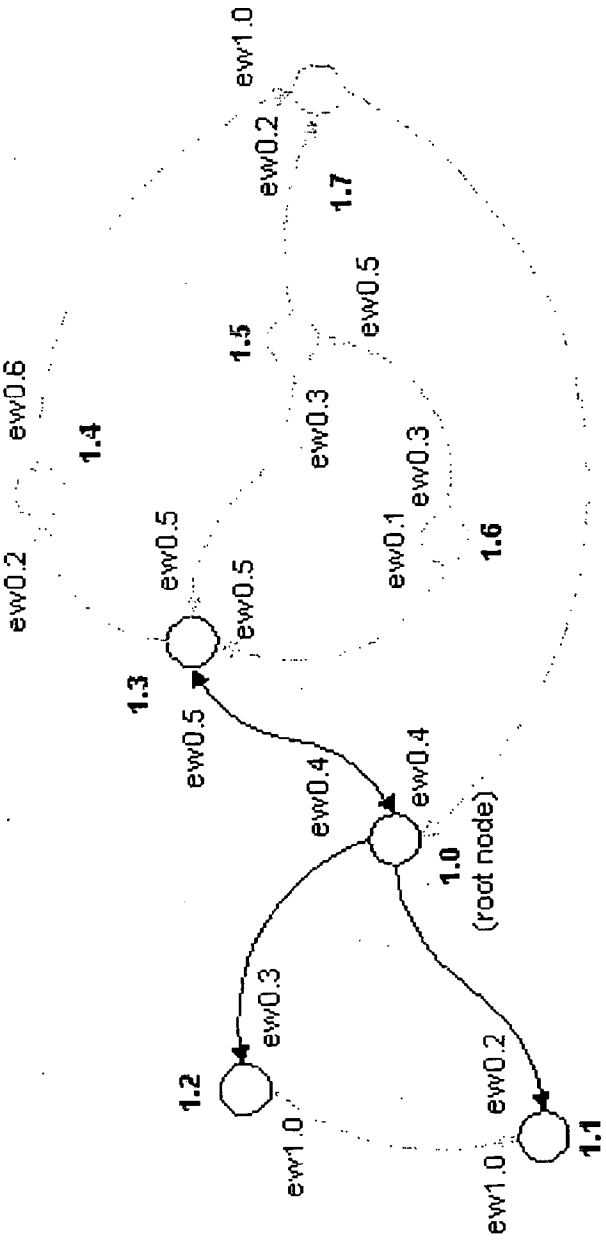


Figure 17

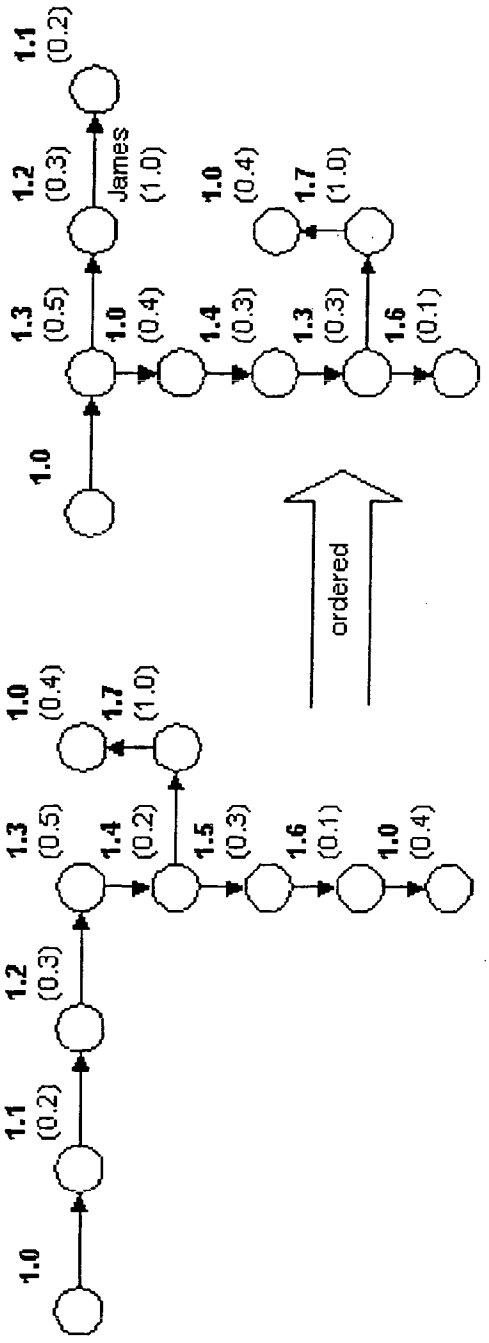


Figure 18

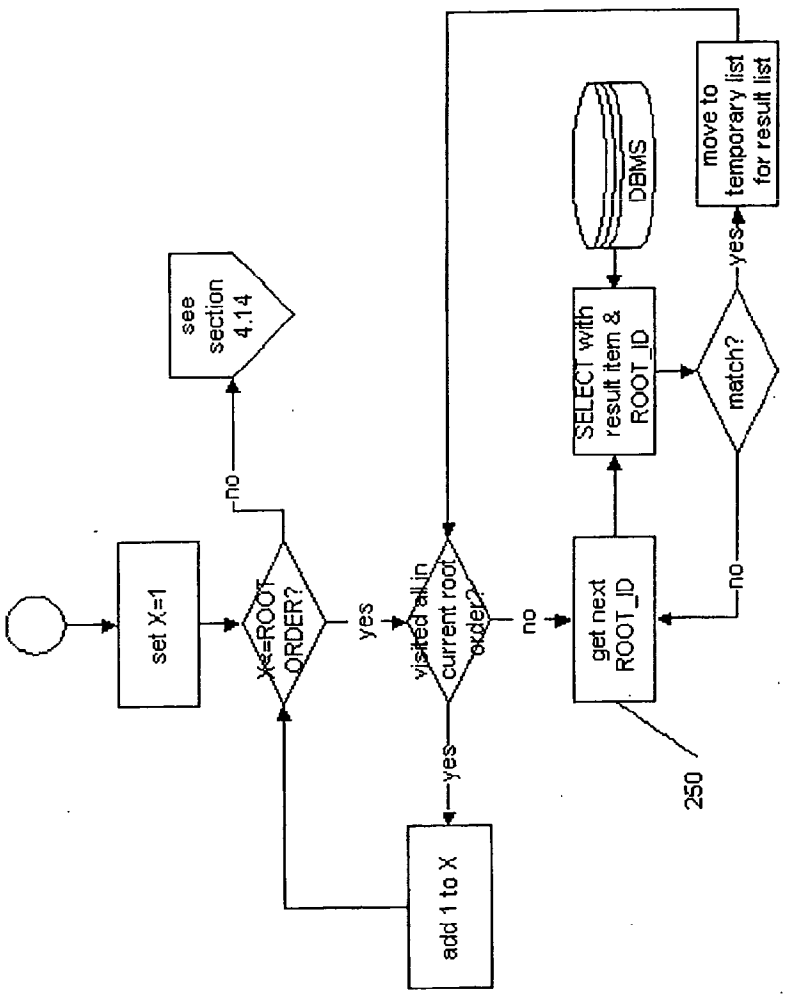


Figure 19

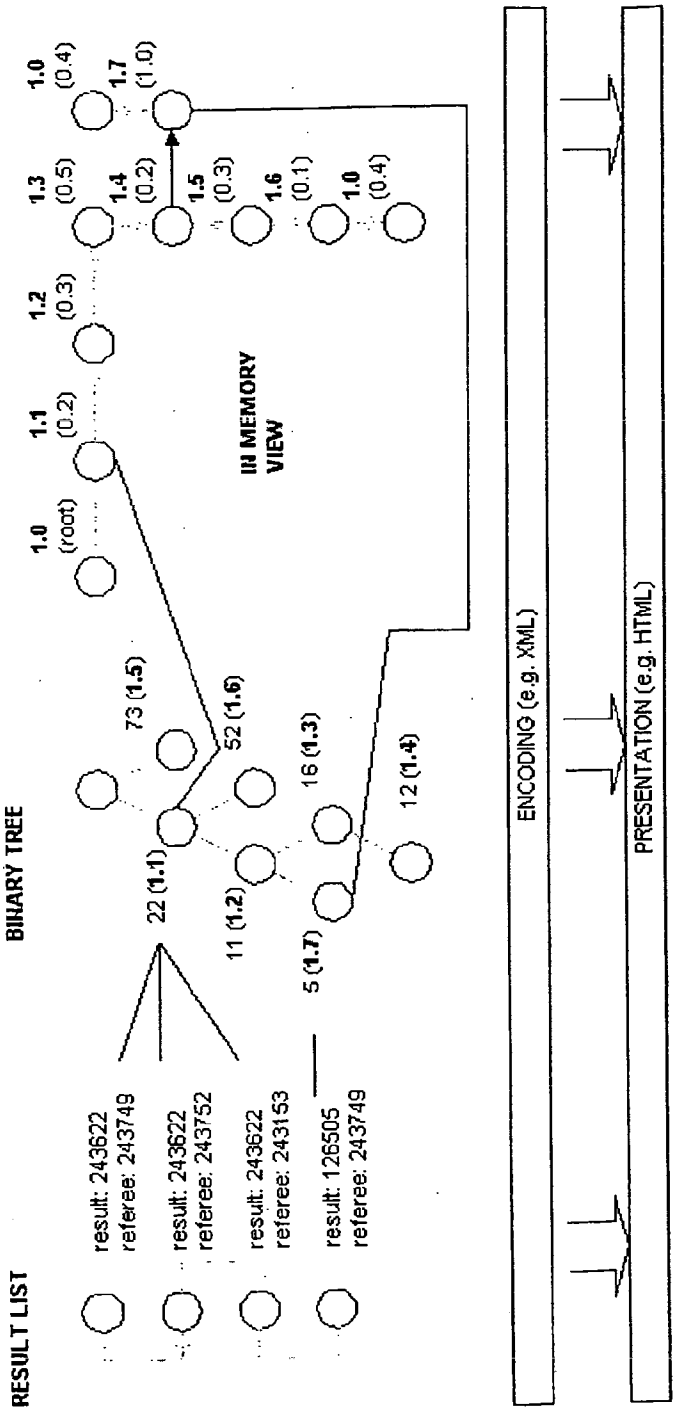


Figure 20

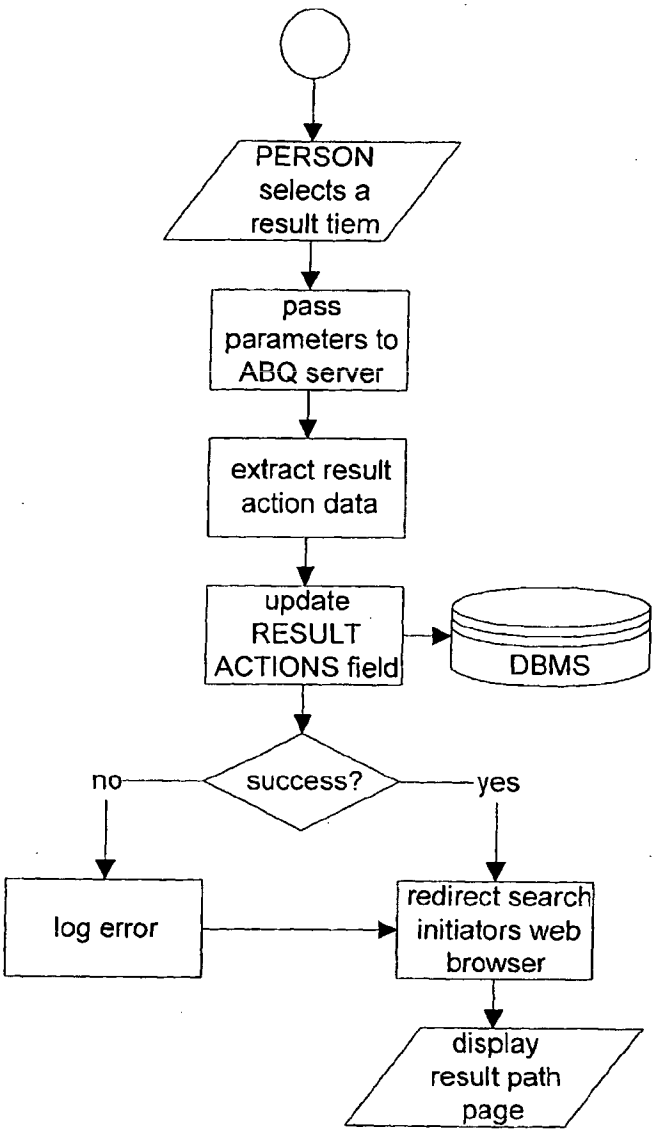

PERSON selects a result tiem
pass parameters to ABQ server
extract result action data
update RESULT ACTIONS field
DBMS
success?
no
yes
log error
redirect search initiators web browser
display result path page

Figure 21

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 01 30 7210

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	KAUTZ H ET AL: "The hidden web" AI MAGAZINE, SUMMER 1997, AMERICAN ASSOC. ARTIFICIAL INTELLIGENCE, USA, vol. 18, no. 2, pages 27-36, XP002202174 ISSN: 0738-4602 * abstract * * page 28, right-hand column, line 47 - page 29, left-hand column, line 14 * * page 31, right-hand column, line 4 - page 32, left-hand column, line 52 * * page 32, right-hand column, line 48 - page 33, left-hand column, line 3 *	1-23, 28-31	G06F17/30
Y	WO 00 04474 A (NET PERCEPTIONS INC) 27 January 2000 (2000-01-27) * abstract * * page 7, line 28 - page 9, line 15 * * page 10, line 1 - page 10, line 15 * * page 17, line 2 - page 17, line 23 * * page 25, line 8 - page 25, line 29 *	1-23, 28-31	
A	KUMAR R ET AL: "Recommendation systems: a probabilistic analysis" FOUNDATIONS OF COMPUTER SCIENCE, 1998. PROCEEDINGS. 39TH ANNUAL SYMPOSIUM ON PALO ALTO, CA, USA 8-11 NOV. 1998, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 8 November 1998 (1998-11-08), pages 664-673, XP010318861 ISBN: 0-8186-9172-7 * page 664, left-hand column, line 18 - page 664, right-hand column, line 34 * * page 665, left-hand column, line 16 - page 665, left-hand column, line 50 * * page 666, right-hand column, line 11 - page 666, right-hand column, line 40 *	1-31	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G06F

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	13 June 2002	Boyadzhiev, Y

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 01 30 7210

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	EP 0 751 471 A (MASSACHUSETTS INST TECHNOLOGY) 2 January 1997 (1997-01-02) * abstract * * page 2, line 1 - page 2, line 42 * * page 3, line 32 - page 4, line 17 * * page 5, line 43 - page 6, line 21 *	1-31	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	13 June 2002	Boyadzhiev, Y

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 01 30 7210

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

13-06-2002

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0004474 A	27-01-2000	US 6334127 B1	25-12-2001
		AU 4973999 A	07-02-2000
		EP 1099177 A1	16-05-2001
		WO 0004474 A1	27-01-2000
EP 0751471 A	02-01-1997	AU 6282596 A	05-02-1997
		CA 2225790 A1	23-01-1997
		EP 0751471 A1	02-01-1997
		JP 11509019 T	03-08-1999
		US 6041311 A	21-03-2000
		US 6049777 A	11-04-2000
		US 6092049 A	18-07-2000
		US 6112186 A	29-08-2000
		WO 9702537 A1	23-01-1997
		US 5872850 A	16-02-1999

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

[redacted]

Attorney Docket No. AOL0142

DECLARATION FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name;

I believe I am the original, first, and sole inventor (if only one name is listed below) or an original, first, and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

SEARCH ENHANCEMENT SYSTEM HAVING PERSONAL SEARCH PARAMETERS

the specification of which (check one) ___ is attached hereto, or _X_ was filed on 10/14/2003 as Application Serial No. 10/685,749 and was amended on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

==========

I hereby claim foreign priority benefits under Title 35, United Sates Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)	Priority Claimed Yes	No
_____	____	____
Number Country Day/Month/Year Filed		
_____	____	____
Number Country Day/Month/Year Filed		

==========

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith:

MICHAEL A. GLENN, Reg. No. 30,176
JAMES R. BRAMSON, Reg. No. 41,632
DONALD M. HENDRICKS, Reg. No. 40,355
CHRISTOPHER PEIL, Reg. No. 45,005
IVY Y. MEI, Reg.No. 46,925
FLORIN CORIE, Reg. No. 43,244
JULIA A. THOMAS, Reg. No. 52,283

SEND CORRESPONDENCE TO:

GLENN PATENT GROUP, 3475 Edison Way, Suite L, Menlo Park, CA 94025

==========

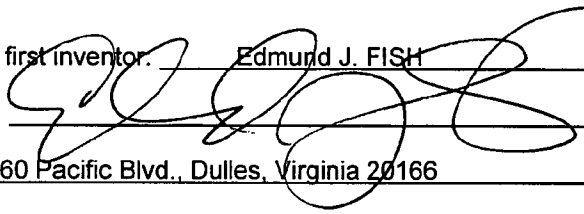
I hereby claim the benefit under Title 35, United States code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Ser. No.	Filing Date	Status: Patented, Pending, Abandoned

==========

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor: Edmund J. FISH

Inventor's signature  10/14/2003
Date

Residence 22260 Pacific Blvd., Dulles, Virginia 20166

Post Office Address Same

Citizenship Unjited States of America

Express Mail mailing label no. EV 189444323 US Date of Deposit: March 19, 2003
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Mail Stop IDS, P.O. Box 1450, Alexandria, VA 22313-1450.

Rhonda Dunn

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Fish
Docket No.: AOL01042
Serial No. : 10/685,749
Art Unit: 2155
Filed: 10/14/2003
Examiner: Unassigned
Title: Search Enhancement System Having Personal Search Parameters

Commissioner for Patents
Mail Stop IDS
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Sir:

This Information Disclosure Statement is submitted:

(X) under 37 CFR 1.97(b), or
(within three months of filing national application; or date of entry of international application; or before mailing date of first office action on the merits; whichever occurs last)

( ) under 37 CFR 1.97(c) together with either a:
- ( ) Certification under 37 CFR 1.97(e), or
- ( ) a $220.00 fee under 37 CFR 1.17(p), or

(After the CFR 1.97(b) time period, but before final action or notice of allowance, whichever occurs first)

( ) under 37 CFR 1.97(d) together with a:
- ( ) Certification under 37 CFR 1.97(e), and
- ( ) a $220.00 fee under 37 CFR 1.17(d)(2)(ii), and
- ( ) a $130.00 petition fee set forth in 37 CFR 1.17(i)(1)

(Filed after final action or notice of allowance, whichever occurs first, but before payment of the issue fee)

(X) The Commissioner is authorized to charge any additional fees or credit any overpayment to Deposit Account No. 07-1445 (Order No. AOL0142). A copy of this sheet is enclosed for accounting purposes.

(X) Applicant(s) submit herewith Form PTO 1449 -- Information Disclosure Citation together with copies of patents, publications or other information of which applicant(s) are aware, which applicant(s) believe(s) may be material to the examination of this application and for which there may be a duty to disclose in accordance with 37 CFR 1.25.

( ) A concise explanation of the relevance of foreign language patents, foreign language publications and other foreign language information listed on PTO Form 1449, as presently understood by the individual(s) designated in 37 CFR 156(c) most knowledgeable about the content is given on the attached sheet, or where a foreign language patent is cited in a search report or other action by a foreign patent office in a counterpart foreign application, an English language version of the search report or action which indicates the degree of relevance found by the foreign office is listed on form PTO 1449 and is enclosed herewith.

It is requested that the information disclosed herein be made of record in this application.

Respectfully Submitted,

Michael Glenn
Attorney For Applicant
Reg. No. 30,176

Customer No. 22862

Attorney Docket No. AOL0142

Form 1449 (Modified)	Atty. Docket No.: AOL0142	Serial No.: 10/685,749
Information Disclosure	Applicant: Fish	Group: 2155
Statement By Applicant		
	Filing Date: 10/14/2003	Examiner: Unassigned

U.S. Patent Documents

Examiner Initial	No.	Patent No.	Date	Patentee	Class	Sub-class	Filing Date
	A	5,642,502	6/24/97	Driscoll	395	606	12/6/94
	B	5,848.396	12/08/98	Gerace	705	10	4/26/96
	C	6,029,195	2/22/00	Herz	709	219	12/5/97
	D	6,314,420	11/06/01	Lang et al.	707	3	12/3/98
	E	6,363,377	3/26/02	Kravets et al.	707	4	12/22/98
	F	6,397,212	5/28/02	Biffar	707	5	3/4/99
	G	6, 539,377	3/25/03	Culliss	707	5	10/6/00
	H	5,983,214	11/09/99	Lang et al.	707	1	11/5/98
	I	6,029,161	2/22/00	Lang et al.	707	1	11/19/98
	J	6,047,327	4/04/00	Tso et al.	709	232	2/16/96
	K	6,182,068	1/30/01	Culliss	707	5	3/1/99
	L	6,308,175	10/23/01	Lang et al.	707	10	11/19/98
	M	6,349,307	2/19/02	Chen	707	103	12/28/98
	N	6,360,215	3/19/02	Judd et al.	707	3	11/3/98
	O	6,377,945	4/23/02	Risvik	707	3	7/9/99
	P	6,404,884	6/11/02	Marwell et al.	379	265.13	11/24/99
	Q	6/453,315	9/17/02	Weissman et al.	707	5	11/1/99
	R	6,466,970	10/15/02	Lee et al.	709	217	1/27/99
	S	6,484,162	11/19/02	Edlund et al.	707	3	6/29//99
	T	6,498,795	12/24/02	Zhang et al.	370	400	5/28/99
	U	6,505,194	1/07/03	Nicolovska et al.	707	3	3/29/00
	V	6,550,057	4/15/03	Bowman-Amuah	717	126	8/31/99

Foreign Patent or Published Foreign Patent Application

Examiner Initial	No.	Document No.	Publication Date	Country or Patent Office	Class	Sub-class	Translation Yes	Translation No
	W	1 072 982	1/31/01	EPO	G06F	17/30	X	
	X	1 095 326	1/30/02	EPO	G06F	1/00	X	
	Y	1 284 461	2/19/03	EPO	G06F	17/30	X	
	Z	1 288 795	3/05/03	EPO	G06F	17/30	X	

Publications and Other Documents

Examiner Initial	No.	Author, Title, Date, Place (e.g. Journal) of Publication
	AA	PITKOW, J. et al.; Personalized Search; Communications of the ACM, vol. 45, no. 9, p. 50-5; September 2002
	AB	MCGOWAN, J.P. et al.; Who Do You Want To Be Today? Web Personae for Personalised Information Access; Adaptive Hypermedia and Adaptive Web-Based Systems. Second International Conference, AH 2002. Proceedings (Lecture Notes in Computer Science Vol. 2347), p. 514-17; 2002
	AC	KALAJDZISKI, S. et al.; Intelligent Recommendation in Digital Library; Proceedings of the IASTED International Conference Intelligent Systems and Control, p. 408-12; ACTA Press, Anaheim, CA, USA; 2001
	AD	KERSCHBERG, L. et al.; A Semantic Taxonomy-Based Personalizable Meta-Search Agent; Proceedings of the Second International Conference on Web Information Systems Engineering, vol. 1, p. 41-50; IEEE Comput. Soc., Los Alamitos, CA, USA; 2002

	AE	DICHEV, C.; A Framework for Context-Driven Web Resource Discovery; Modeling and Using Context. Third International and Interdisciplinary Conference, Context 2001, Proceedings (Lecture Notes in Artificial Intelligence, vol. 2116), p. 433-6; Springer-Verlag, Berlin, Germany; 2001
	AF	MENG, X. et al.; Feasibility of Adding Filtering Process in Web Browser to Improve Web Search Accuracy; Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, PDPTA 2000, vol. 4, p. 1809-15; CSREA Press, Athens, GA, USA; 2000
	AG	KIM, K. et al.; Development of a Personalized Link-Based Search Engine Using Fuzzy Concept Network; Journal of KISS: Computing Practices, vol. 7, no. 3, p. 211-19; Korea Inf. Sci. Soc.; June 2001
	AH	YANG, C. et al.; A Hybrid Document Clustering for a Web Agent; Journal of KISS: Software and Applications, vol. 28, no. 5, p. 422-30; Korea Inf. Sci. Soc., May 2001
	AI	KIM, K. et al.; A Personalized Web Search Engine Using Fuzzy Concept Network with Link Structure; Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), vol. 1, p. 81-6; IEEE, Piscataway, NJ, USA; 2001
	AJ	SCIME, A. et al.; WebSifter: An Ontology-Based Personalizable Search Agent for the Web; Proceedings 2000 Kyoto International Conference on Digital Libraries: Research and Practice, p. 203-10; IEEE Comput. Soc., Los Alamitos, CA, USA; 2000
	AK	WEI-FENG, Z. et al.; Personalizing Search Result Using Agent; Mini-Micro Systems, vol. 22, no. 6, p. 724-7; Mini-Micro Syst., China; June 2001
	AL	CHEN, P. et al.; An Information Retrieval System Based on a User Profile; Journal of Systems and Software, vol. 54, no. 1, p. 3-8; Elsevier; September 30, 2000
	AM	MENG, X. et al.; Personalized Web Search Using Information On Client's Side; Fifth International Conference for Young Computer Scientists, ICYCS'99, Advances in Computer Science and Technology, vol. 2, p. 985-92; Int. Acad. Publishers, Beijing, China; 1999
	AN	CHEN, P. et al.; A Personalized Information Retrieval System; Computational Intelligence for Modelling, Control and Automation, Intelligent Image Processing, Data Analysis and Information Retrieval (Concurrent Systems Engineering Series, vol. 56), p. 247-53; IOS Press, Amsterdam, Netherlands; 1999
	AO	LAINE-CRUZEL, S. et al.; Improving Information Retrieval by Combining User Profile and Document Segmentation; Information Processing & Management, vol. 32, no. 3, p. 305-15; Elsevier; May 1996
	AP	BOLEY, D. et al.; Document Categorization and Query Generation on the World Wide Web Using WebACE; Department of Computer Science and Engineering, University of Minnesota
	AQ	PRETSCHNER, A.; Ontology Based Personalized Search; Dipl.-Inform., RWTH Aachen, Germany, 1998
	AR	CALLAN, J. et al.; Document Filtering with Inference Networks; Computer Science Department, University of Massachusetts
	AS	GOKER, A.; Capturing Information Need by Learning User Context; School of Computer and Mathematical Sciences; The Robert Gordon University
	AT	CHEN, L. et al.; WebMate: A Personal Agent for Browsing and Searching; The Robotics Institute, Carnegie Mellon Institute; September 30, 1997
	AU	COOLEY, R. et al.; Web Mining: Information and Pattern Discovery on the World Wide Web; Department of Computer Science and Engineering, University of Minnesota
	AV	SIMONS, J.; Using a Semantic User Model to Filter the World Wide Web Proactively; Nijmegen Institute for Cognition and Information, University of Nijmegen, The Netherlands
	AW	TANUDJAJA, F. et al.; Persona: A Contextualized and Personalized Web Search; Laboratory of Computer Science at MIT, Cambridge, MA; June 1, 2001
	AX	YAN, T. et al.; SIFT - A Tool for Wide-Area Information Dissemination; Department of Computer Science, Stanford University, February 16, 1995
	AY	BIANCHI-BERTHOUZE, N.; Mining Multimedia Subjective Feedback; Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies, vol. 19, no. 1, p. 43-59; Kluwer Academic Publishers; July 2002
	AZ	WIDYANTORO, D.H. et al.; A fuzzy Ontology-Based Abstract Search Engine and Its User Studies; 10th IEEE International Conference on Fuzzy Systems. (Cat. No.01CH37297), vol. 2, p. 1291-4; IEEE, Piscataway, NJ, USA; 2001

	BA	WIDYANTORO, D.H. et al.; Using Fuzzy Ontology for Query Refinement in a Personalized Abstract Search Engine; Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), vol. 1, p. 610-15; IEEE , Piscataway, NJ, USA; 2001
	BB	HO, M. et al.; A GA-Based Dynamic Personalized Filtering for Internet Search Service on Multi-Search Engine; Canadian Conference on Electrical and Computer Engineering 2001, Conference Proceedings (Cat. No.01TH8555) vol. 1, p. 271-6; IEEE , Piscataway, NJ, USA; 2001
	BC	POGAENIK, M. et al.; Layered Agent System Architecture for Personalized Retrieval of Information from Internet; Signal Processing X Theories and Applications. Proceedings of EUSIPCO 2000. Tenth European Signal Processing Conference, vol. 1, p. 421-4; Tampere Univ. Technology, Tampere, Finland; 2000
	BD	HO, M. et al.; An Agent-Based Personalized Search on a Multi-Search Engine Based on Internet Search Service; Intelligent Data Engineering and Automated - IDEAL 2000, Data Mining, Financial Engineering, and Intelligent Agents, Second International Conference, Proceedings (Lecture Notes in Computer Science Vol.1983), p. 404-9; Springer-Verlag, Berlin, Germany; 2000
	BE	OVERMEER, M.A.C.J.; My Personal Search Engine; Computer Networks, vol. 31, no. 21, p. 2271-9; Elsevier, November 10, 1999
	BF	LEE, E.S. et al.; Agent-Based Support for Personalized Information with Web Search Engines; Design of Computing Systems: Cognitive Considerations. Proceedings of the Seventh International Conference on Human-Computer Interaction (HCI International '97), vol. 2, p. 783-6; Elsevier, Amsterdam, Netherlands; 1997
	BG	BERGER, F.C. et al.; Personalized Search Support for Networked Document Retrieval Using Link Inference; Database and Expert Systems Applications. 7th International Conference, DEXA '96 Proceedings, p. 802-11; Springer-Verlag, Berlin, Germany; 1996

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	1	@ad<"20031014" and "6,282,548".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:02
S2	1	@ad<"20031014" and "5,812,776".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:02
S3	2	@ad<"20031014" and "6,256,639".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:04
S4	1	@ad<"20031014" and "6,311,214".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:05
S5	2	@ad<"20031014" and "6,314,423".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:05
S6	2	@ad<"20031014" and "6,324,566".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:06
S7	1	@ad<"20031014" and "6,405,222".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:08
S8	2	@ad<"20031014" and "6,480,853".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:09
S9	2	@ad<"20031014" and "6,493,702".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:19
S10	1	@ad<"20031014" and (integrat$3 or incoporat$3) with (stored or saved) with parameter with query and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:20
S11	14	@ad<"20031014" and (integrat$3 or incoporat$3) same (stored or saved) with (term or keyword or parameter) with query and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:21

S12	6	@ad<"20031014" and (integrat$3 or incoporat$3) with (stored or saved) with (term or keyword or parameter) with query and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:23
S13	2	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:25
S14	211	@ad<"20031014" and (integrat$3 or incoporat$3 or select$3) with (term or keyword or parameter) with (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:26
S15	9	@ad<"20031014" and (integrat$3 or incoporat$3 or select$3) with (term or keyword or parameter) with (search adj query) and (subject adj matter) with (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:30
S16	70	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:31
S17	7	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (query) and (subject adj matter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:15
S18	15	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (query) and (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:21
S19	62	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or addW$3) with (term or keyword or parameter) with (search adj query) and refin$5 and result and input and (subject or criteria) and user near (personal$4 or information)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:22
S20	69	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and refin$5 and result and input and (subject or criteria) and user near (personal$4 or information)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:22
S21	19	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and refin$5 with result and input and (subject or criteria) and user near (personal$4 or information)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:37

S22	97	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:50
S23	17	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) near result and (subject or criteria or parameter) and user near (personal$7 or information or profile)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:38
S24	10	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near source and (rat$3 or rank$3)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:59
S25	2	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near source and recipient	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:58
S26	10	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near source	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:58
S27	10	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near (recipient or source or server or provider) and (rat$3 or rank$3)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 22:01

S28	16	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3 or rat$3 or rank$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near (recipient or source or server or provider)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 22:01

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

22862 7590 07/13/2005

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

EXAMINER	
NGUYEN, QUANG N	

ART UNIT	PAPER NUMBER
2141	

DATE MAILED: 07/13/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

Office Action Summary	Application No.	Applicant(s)	
	10/685,749	FISH, EDMUND J.	
	Examiner	Art Unit	
	Quang N. Nguyen	2141	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *14 October 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-33* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-33* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *14 October 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *20040319*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

Detailed Action

1. This Office Action is in response to the Application SN 10/685,749 filed on 10/14/2003. Claims 1-33 are presented for examination.

Claim Rejections - 35 USC § 102

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

3. Claims 1-33 are rejected under 35 U.S.C. 102(b) as being anticipated by Harris (US 2002/0059204 A1).

4. As to claim 1, Harris teaches a process, comprising the steps of:

storing a set of search parameters specified by a user *(storing log files of received query information, customized search queries, generated search results, etc., in a dictionary 18 selected by initiating device)* (Harris, paragraphs [0055 and 0062]); and

integrating selected ones of the stored search parameters into a search query based upon a subject matter of the search query *(the query translator 34 can utilize the specified dictionary 18 including one or more of relevant content and/or values from a data source 20; configuration settings, profiles according to users and/or groups of users, business rules, terminology definitions or specifications, synonyms, a language translator, a natural language processor, etc., and query information to generate a customized search or query)* (Harris, paragraphs [0059, 0062-0063 and 0069]).

5. As to claim 2, Harris teaches the process of claim 1, further comprising the step of returning results of the search query to the user *(the customized search results can be provided to server 16 for display)* (Harris, paragraphs [0062-0063]).

6. As to claims 3-6, Harris teaches the process of claim 1, further comprising the step of sending results of the search query to a recipient selected by the user *(i.e., based on information input by the user)* or based on the results of the search query *(the query search results 108 can be transferred to the requested entity or another entity can be designated or otherwise specified to receive the query results 108, for example, if the query search results needs to be additionally filtered/formatted/aggregated then it can be returned to the dictionary 18 or the system server 14 for formatting, arranging, aggregating before transferring, forwarding or displaying to the initiating device 12 or to another recipient)* (Harris, paragraph [0031 and 0060]).

7. As to claims 7-8, Harris teaches the process of claim 1, further comprising the steps of:

receiving information from a recipient comprising an interest in the results of the search query *(receiving interest in verifying a quantity of a stock from the administrator and/or another interested party)*; and

selectively sending results of the search query to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user *(once a stock quantity reaches a predetermined value, an email can be sent to the administrator and/or another interested party)* (Harris, paragraph [0056]).

8. As to claim 9, Harris teaches the process of claim 1, wherein the subject matter is explicitly indicated by the user *(the query information initially submitted by the user)* (Harris, paragraph [0069]).

9. As to claim 10, Harris teaches the process of claim 1, wherein the subject matter is inferentially determined based upon user input (Harris, paragraph [0046]).

10. As to claim 11, Harris teaches the process of claim 1, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, further comprising the step of: determining the selected ones of the stored search parameters based upon the selectable applicability (Harris, paragraphs [0052-0053 and 0069]).

11. As to claim 12, Harris teaches a process, comprising the steps of:

receiving at least one search parameter specified by a user *(receiving the dictionary selection and/or the list of subscriber servers with the initial query information)* (Harris, paragraphs [0040 and 0062-0063]);

receiving a search query from the user *(the query information "How far is bank XXX from home?" initially submitted by the user)* (Harris, paragraph [0069]);

determining the subject matter of the search query *(i.e., parsing the received information, performing context analysis, etc.)* (Harris, paragraphs [0046 and 0069]);

selecting one or more user-specified search parameters based on the determined subject matter *(selecting data pertinent to the user "home" and data related to another sensor or entity such as mapping program or distance calculator, selecting synonyms or phonetic equivalents, etc.)* (Harris, paragraphs [0046 and 0069]); and

integrating the selected search parameters with the search query *(the query translator 34 can utilize the specified dictionary 18 including one or more of relevant content and/or values from a data source 20, configuration settings, profiles according to users and/or groups of users, business rules, terminology definitions or specifications, synonyms, a language translator, a natural language processor, etc., and query information to generate a customized search or query)* (Harris, paragraphs [0059, 0062-0063 and 0069]).

12. Claims 13-22 are corresponding process claims of claims 2-11; therefore, they are rejected under the same rationale.

13. Claims 23-33 are corresponding system claims of process claims 12-22; therefore, they are rejected under the same rationale.

14. Further references of interest are cited on Form PTO-892, which is an attachment to this office action.

15. A shortened statutory period for reply to this action is set to expire THREE (3) months from the mailing date of this communication. See 37 CFR 1.134.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Quang N. Nguyen whose telephone number is (571) 272-3886.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's SPE, Rupal Dharia, can be reached at (571) 272-3880. The fax phone number for the organization is (571) 273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

RUPAL DHARIA
SUPERVISORY PATENT EXAMINER

Attorney Docket No. AOL0142

Form 1449 (Modified)	Atty. Docket No.: AOL0142	Serial No.: 10/685,749
Information Disclosure	Applicant: Fish	Group: 2155
Statement By Applicant		
	Filing Date: 10/14/2003	Examiner: Unassigned

U.S. Patent Documents

Examiner Initial	No.	Patent No.	Date	Patentee	Class	Sub-class	Filing Date
QF	A	5,642,502	6/24/97	Driscoll	395	606	12/6/94
	B	5,848.396	12/08/98	Gerace	705	10	4/26/96
	C	6,029,195	2/22/00	Herz	709	219	12/5/97
	D	6,314,420	11/06/01	Lang et al.	707	3	12/3/98
	E	6,363,377	3/26/02	Kravets et al.	707	4	12/22/98
	F	6,397,212	5/28/02	Biffar	707	5	3/4/99
	G	6, 539,377	3/25/03	Culliss	707	5	10/6/00
	H	5,983,214	11/09/99	Lang et al.	707	1	11/5/98
	I	6,029,161	2/22/00	Lang et al.	707	1	11/19/98
	J	6,047,327	4/04/00	Tso et al.	709	232	2/16/96
	K	6,182,068	1/30/01	Culliss	707	5	3/1/99
	L	6,308,175	10/23/01	Lang et al.	707	10	11/19/98
	M	6,349,307	2/19/02	Chen	707	103	12/28/98
	N	6,360,215	3/19/02	Judd et al.	707	3	11/3/98
	O	6,377,945	4/23/02	Risvik	707	3	7/9/99
	P	6,404,884	6/11/02	Marwell et al.	379	265.13	11/24/99
	Q	6/453,315	9/17/02	Weissman et al.	707	5	11/1/99
	R	6,466,970	10/15/02	Lee et al.	709	217	1/27/99
	S	6,484,162	11/19/02	Edlund et al.	707	3	6/29//99
	T	6,498,795	12/24/02	Zhang et al.	370	400	5/28/99
	U	6,505,194	1/07/03	Nicolovska et al.	707	3	3/29/00
QF	V	6,550,057	4/15/03	Bowman-Amuah	717	126	8/31/99

Foreign Patent or Published Foreign Patent Application

Examiner Initial	No.	Document No.	Publication Date	Country or Patent Office	Class	Sub-class	Translation Yes	Translation No
QF	W	1 072 982	1/31/01	EPO	G06F	17/30	X	
QF	X	1 095 326	1/30/02	EPO	G06F	1/00	X	
QF	Y	1 284 461	2/19/03	EPO	G06F	17/30	X	
QF	Z	1 288 795	3/05/03	EPO	G06F	17/30	X	

Publications and Other Documents

Examiner Initial	No.	Author, Title, Date, Place (e.g. Journal) of Publication
QF	AA	PITKOW, J. et al.; Personalized Search; Communications of the ACM, vol. 45, no. 9, p. 50-5; September 2002
QF	AB	MCGOWAN, J.P. et al.; Who Do You Want To Be Today? Web Personae for Personalised Information Access; Adaptive Hypermedia and Adaptive Web-Based Systems. Second International Conference, AH 2002. Proceedings (Lecture Notes in Computer Science Vol. 2347), p. 514-17; 2002
QF	AC	KALAJDZISKI, S. et al.; Intelligent Recommendation in Digital Library; Proceedings of the IASTED International Conference Intelligent Systems and Control, p. 408-12; ACTA Press, Anaheim, CA, USA; 2001
QF	AD	KERSCHBERG, L. et al.; A Semantic Taxonomy-Based Personalizable Meta-Search Agent; Proceedings of the Second International Conference on Web Information Systems Engineering, vol. 1, p. 41-50; IEEE Comput. Soc., Los Alamitos, CA, USA; 2002

QW	AE	DICHEV, C.; A Framework for Context-Driven Web Resource Discovery; Modeling and Using Context. Third International and Interdisciplinary Conference, Context 2001, Proceedings (Lecture Notes in Artificial Intelligence, vol. 2116), p. 433-6; Springer-Verlag, Berlin, Germany; 2001
	AF	MENG, X. et al.; Feasibility of Adding Filtering Process in Web Browser to Improve Web Search Accuracy; Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, PDPTA 2000, vol. 4, p. 1809-15; CSREA Press, Athens, GA, USA; 2000
	AG	KIM, K. et al.; Development of a Personalized Link-Based Search Engine Using Fuzzy Concept Network; Journal of KISS: Computing Practices, vol. 7, no. 3, p. 211-19; Korea Inf. Sci. Soc.; June 2001
	AH	YANG, C. et al.; A Hybrid Document Clustering for a Web Agent; Journal of KISS: Software and Applications, vol. 28, no. 5, p. 422-30; Korea Inf. Sci. Soc., May 2001
	AI	KIM, K. et al.; A Personalized Web Search Engine Using Fuzzy Concept Network with Link Structure; Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), vol. 1, p. 81-6; IEEE, Piscataway, NJ, USA; 2001
	AJ	SCIME, A. et al.; WebSifter: An Ontology-Based Personalizable Search Agent for the Web; Proceedings 2000 Kyoto International Conference on Digital Libraries: Research and Practice, p. 203-10; IEEE Comput. Soc., Los Alamitos, CA, USA; 2000
	AK	WEI-FENG, Z. et al.; Personalizing Search Result Using Agent; Mini-Micro Systems, vol. 22, no. 6, p. 724-7; Mini-Micro Syst., China; June 2001
	AL	CHEN, P. et al.; An Information Retrieval System Based on a User Profile; Journal of Systems and Software, vol. 54, no. 1, p. 3-8; Elsevier; September 30, 2000
	AM	MENG, X. et al.; Personalized Web Search Using Information On Client's Side; Fifth International Conference for Young Computer Scientists, ICYCS'99, Advances in Computer Science and Technology, vol. 2, p. 985-92; Int. Acad. Publishers, Beijing, China; 1999
	AN	CHEN, P. et al.; A Personalized Information Retrieval System; Computational Intelligence for Modelling, Control and Automation, Intelligent Image Processing, Data Analysis and Information Retrieval (Concurrent Systems Engineering Series, vol. 56), p. 247-53; IOS Press, Amsterdam, Netherlands; 1999
	AO	LAINE-CRUZEL, S. et al.; Improving Information Retrieval by Combining User Profile and Document Segmentation; Information Processing & Management, vol. 32, no. 3, p. 305-15; Elsevier; May 1996
	AP	BOLEY, D. et al.; Document Categorization and Query Generation on the World Wide Web Using WebACE; Department of Computer Science and Engineering, University of Minnesota
	AQ	PRETSCHNER, A.; Ontology Based Personalized Search; Dipl.-Inform., RWTH Aachen, Germany, 1998
	AR	CALLAN, J. et al.; Document Filtering with Inference Networks; Computer Science Department, University of Massachusetts
	AS	GOKER, A.; Capturing Information Need by Learning User Context; School of Computer and Mathematical Sciences; The Robert Gordon University
	AT	CHEN, L. et al.; WebMate: A Personal Agent for Browsing and Searching; The Robotics Institute, Carnegie Mellon Institute; September 30, 1997
	AU	COOLEY, R. et al.; Web Mining: Information and Pattern Discovery on the World Wide Web; Department of Computer Science and Engineering, University of Minnesota
	AV	SIMONS, J.; Using a Semantic User Model to Filter the World Wide Web Proactively; Nijmegen Institute for Cognition and Information, University of Nijmegen, The Netherlands
	AW	TANUDJAJA, F. et al.; Persona: A Contextualized and Personalized Web Search; Laboratory of Computer Science at MIT, Cambridge, MA; June 1, 2001
	AX	YAN, T. et al.; SIFT - A Tool for Wide-Area Information Dissemination; Department of Computer Science, Stanford University, February 16, 1995
	AY	BIANCHI-BERTHOUZE, N.; Mining Multimedia Subjective Feedback; Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies, vol. 19, no. 1, p. 43-59; Kluwer Academic Publishers; July 2002
QN	AZ	WIDYANTORO, D.H. et al.; A fuzzy Ontology-Based Abstract Search Engine and Its User Studies; 10th IEEE International Conference on Fuzzy Systems. (Cat. No.01CH37297), vol. 2, p. 1291-4; IEEE, Piscataway, NJ, USA; 2001

QN	BA	WIDYANTORO, D.H. et al.; Using Fuzzy Ontology for Query Refinement in a Personalized Abstract Search Engine; Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), vol. 1, p. 610-15; IEEE , Piscataway, NJ, USA; 2001
	BB	HO, M. et al.; A GA-Based Dynamic Personalized Filtering for Internet Search Service on Multi-Search Engine; Canadian Conference on Electrical and Computer Engineering 2001, Conference Proceedings (Cat. No.01TH8555) vol. 1, p. 271-6; IEEE , Piscataway, NJ, USA; 2001
	BC	POGAENIK, M. et al.; Layered Agent System Architecture for Personalized Retrieval of Information from Internet; Signal Processing X Theories and Applications. Proceedings of EUSIPCO 2000. Tenth European Signal Processing Conference, vol. 1, p. 421-4; Tampere Univ. Technology, Tampere, Finland; 2000
	BD	HO, M. et al.; An Agent-Based Personalized Search on a Multi-Search Engine Based on Internet Search Service; Intelligent Data Engineering and Automated - IDEAL 2000, Data Mining, Financial Engineering, and Intelligent Agents, Second International Conference, Proceedings (Lecture Notes in Computer Science Vol.1983), p. 404-9; Springer-Verlag, Berlin, Germany; 2000
	BE	OVERMEER, M.A.C.J.; My Personal Search Engine; Computer Networks, vol. 31, no. 21, p. 2271-9; Elsevier, November 10, 1999
	BF	LEE, E.S. et al.; Agent-Based Support for Personalized Information with Web Search Engines; Design of Computing Systems: Cognitive Considerations. Proceedings of the Seventh International Conference on Human-Computer Interaction (HCI International '97), vol. 2, p. 783-6; Elsevier, Amsterdam, Netherlands; 1997
QN	BG	BERGER, F.C. et al.; Personalized Search Support for Networked Document Retrieval Using Link Inference; Database and Expert Systems Applications. 7th International Conference, DEXA '96 Proceedings, p. 802-11; Springer-Verlag, Berlin, Germany; 1996

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

Nguyen Q 07/07/05

Notice of References Cited	Application/Control No. 10/685,749	Applicant(s)/Patent Under Reexamination FISH, EDMUND J.	
	Examiner Quang N. Nguyen	Art Unit 2141	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-2002/0059204	05-2002	Harris, Larry R.	707/3
	B	US-2005/0028156	02-2005	Hammond et al.	718/100
	C	US-2005/0071328	03-2005	Lawrence, Stephen R.	707/003
	D	US-6,256,639	07-2001	Himmel et al.	709/219
	E	US-6,314,423	11-2001	Himmel et al.	709/218
	F	US-6,324,566	11-2001	Himmel et al.	709/203
	G	US-6,480,853	11-2002	Jain, Anuj Kumar	709/225
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Index of Claims

Application/Control No.	Applicant(s)/Patent under Reexamination
10/685,749	FISH, EDMUND J.
Examiner	**Art Unit**
Quang N. Nguyen	2141

√	Rejected	–	(Through numeral) Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

Claim		Date
Final	Original	7/7/05
	1	√
	2	√
	3	√
	4	√
	5	√
	6	√
	7	√
	8	√
	9	√
	10	√
	11	√
	12	√
	13	√
	14	√
	15	√
	16	√
	17	√
	18	√
	19	√
	20	√
	21	√
	22	√
	23	√
	24	√
	25	√
	26	√
	27	√
	28	√
	29	√
	30	√
	31	√
	32	√
	33	√
	34	
	35	
	36	
	37	
	38	
	39	
	40	
	41	
	42	
	43	
	44	
	45	
	46	
	47	
	48	
	49	
	50	

Claim		Date
Final	Original	
	51	
	52	
	53	
	54	
	55	
	56	
	57	
	58	
	59	
	60	
	61	
	62	
	63	
	64	
	65	
	66	
	67	
	68	
	69	
	70	
	71	
	72	
	73	
	74	
	75	
	76	
	77	
	78	
	79	
	80	
	81	
	82	
	83	
	84	
	85	
	86	
	87	
	88	
	89	
	90	
	91	
	92	
	93	
	94	
	95	
	96	
	97	
	98	
	99	
	100	

Claim		Date
Final	Original	
	101	
	102	
	103	
	104	
	105	
	106	
	107	
	108	
	109	
	110	
	111	
	112	
	113	
	114	
	115	
	116	
	117	
	118	
	119	
	120	
	121	
	122	
	123	
	124	
	125	
	126	
	127	
	128	
	129	
	130	
	131	
	132	
	133	
	134	
	135	
	136	
	137	
	138	
	139	
	140	
	141	
	142	
	143	
	144	
	145	
	146	
	147	
	148	
	149	
	150	

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

Bib Data Sheet

CONFIRMATION NO. 6161

SERIAL NUMBER	FILING DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
10/685,749	10/14/2003 RULE	709	2141	AOL0142

APPLICANTS

Edmund J. Fish, Dulles, VA;

** CONTINUING DATA **********

** FOREIGN APPLICATIONS **********

IF REQUIRED, FOREIGN FILING LICENSE GRANTED
** 01/21/2004

Foreign Priority claimed ☐ yes ☑ no 35 USC 119 (a-d) conditions met ☐ yes ☑ no ☐ Met after Allowance Verified and Acknowledged Examiner's Signature Initials	STATE OR COUNTRY VA	SHEETS DRAWING 15	TOTAL CLAIMS 33	INDEPENDENT CLAIMS 3

ADDRESS
22862
GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK , CA
94025

TITLE
Search enhancement system having personal search parameters

FILING FEE RECEIVED 1134	FEES: Authority has been given in Paper No. _____ to charge/credit DEPOSIT ACCOUNT No. _____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

Search Notes

Application/Control No.	Applicant(s)/Patent under Reexamination	
10/685,749	FISH, EDMUND J.	
Examiner	**Art Unit**	
Quang N. Nguyen	2141	

SEARCHED			
Class	Subclass	Date	Examiner
709	203	7/5/2005	QN
709	218,219	7/5/2005	QN
709	225	7/5/2005	QN
707	3	7/5/2005	QN

INTERFERENCE SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	DATE	EXMR
East Search (see notes)	7/5/2005	QN

10-17-05

1fw

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Edmund J. Fish **Docket No.:** AOL0142

Serial No.: 10/685,749 **Art Unit:** 2141

Filed: 14 October 2003 **Examiner:** Nguyen, Quang N

Title: Search Enhancement System Having Personal Search Parameters

13 October 2005

Commissioner for Patents
Mail Stop Amendment
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT IN RESPONSE TO OFFICE ACTION

Sir:

In response to the Office Action of July 13, 2005, Applicant responds as follows:

Applicant has also entered eighteen new dependent Claims. The Application as originally filed included 33 total claims, with 3 independent Claims. The Application as currently prosecuted includes 51 total claims, with 3 independent claims. Applicant therefore authorizes payment in compliance with 37 CFR 1.16(i) of $900 for 18 additional total claims, and any other fees that may be due to be charged to the Glenn Patent Group Deposit Account No. 07-1445, Customer No. 22862.

REMARKS

35 U.S.C. § 102. Claim Rejections.

3. Claims 1-33 are rejected under 35 U.S.C. §102(b) as being anticipated by Harris (U.S. Patent Publication No. 2002/0059204 A1).

4-10. In regard to Claim 1, the Office Action states that "Harris teaches a process, comprising the steps of:

storing a set of search parameters specified by a user (storing log files of received query information, customized search queries, generated search results, etc., in a dictionary 18 selected by initiating device)(Harris, paragraphs [0055 and 0062]); and

integrating selected ones of the stored search parameters into a search query based upon a subject matter of the search query (the query translator 34 can utilize the specified dictionary 18 including one or more of relevant content and/or values from a data source 20, configuration settings, profiles according to users and/or groups of users, business rules, terminology definitions or specifications, synonyms, a language translator, a natural language processor, etc., and query information to generate a customized search query)(Harris, paragraphs [0059, 0062-0063 and 0069])."

Applicant disagrees that Claim 1 is anticipated by Harris. Applicant reserves the right to present the same or similar claims in a related Application. For sake of convenience in prosecution, Applicant has amended Claim 1, as discussed below, to particularly point out and distinctly claim some of the preferred embodiments of the invention.

Harris describes a "distributed search system and method" seen at least in the abstract, wherein:

> "A method and system for providing distributing a query to devices on a communications network, such as servers on the internet, using an application that can survey a subscriber's server that can include text documents and databases, and use the formatting information and data from the survey to create at least one dictionary customized to the

subscriber's data sources. A user seeking information can initiate a search or query from an initiating device using keywords, natural language terms, connectors, expressions, etc., and the query can be transmitted to various subscriber customized dictionaries. The customized dictionaries can customize the query based on respective subscriber databases and text documents, text search engines, etc., to produce an customized query result. The query results can be filtered and integrated for presentation to the initiating device. Search results can be customized using user preference or profile information."

Applicant has amended independent Claim 1, to claim a process, comprising the steps of:

soliciting from a user a set of one or more search parameters specified by the user;

associatively storing the set of solicited search parameters specified by the user for at least one future search;

receiving a search query from the user;

determining the subject matter of the search query;

selecting one or more of the associatively stored solicited search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query; and

integrating the selected ones of the associatively stored solicited search parameters with the search query.

Support is seen in the Application as filed, at least on page 23, line 23 to page 29, line 14, in Claims 1-33; and in Figures 12-16.

While Harris describes that "search results can be customized using user preference or profile information", as seen at least in the abstract, Harris does not disclose a process which, *inter alia*, comprises the steps of:

associatively storing a set of solicited search parameters specified by the user for at least one future search;

receiving a search query from the user;

determining the subject matter of the search query;

selecting one or more of the associatively stored solicited search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query; and

integrating the selected ones of the associatively stored solicited search parameters with the search query.

Applicant therefore submits that independent Claim 1, as amended, overcomes the rejection under 35 U.S.C. §102(b) as being unpatentable over Harris.

The Examiner bears the burden of establishing a *prima facie* case of anticipation (In re King, 801 F.2d 1324, 1327, 231 USPQ 136, 138-139 (Fed. Cir. 1986)). The prior art reference must disclose each element of the claimed invention, as correctly interpreted, and as arranged in the claim (Lindermann Maschinefabrik Gmbh v. American Hoist & Derrick Co., 730 F.2d 1452, 221 USPQ 481, 485 (Fed. Cir. 1984)). A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. The identical invention must be shown in as complete detail as is contained in the claim (MPEP 2131).

As dependent claims 2-11, as and new claims 40-45 as discussed below, depend from amended independent Claim 1, and inherently contain all the limitations of the claims they depend from, they are seen to be patentable as well.

11-13. In regard to Claim 12, the Office Action states that "Harris teaches a process, comprising the steps of:

receiving at least one search parameter specified by a user (receiving the dictionary selection and/or the list of subscriber servers with the initial query information)(Harris, paragraphs [0040 and 0062-0063]);

receiving a search query from the user (the query information "How far is bank XXX from home" initially submitted by the user)(Harris, paragraph [0069]);

determining the subject matter of the search query (i.e. parsing the received information, performing context analysis, etc.)(Harris, paragraphs [0046 and 0069]);

selecting one or more user-specified search parameters based on the determined subject matter (selecting data pertinent to the user 'home" and data related to another sensor or entity such as a mapping program or distance calculator, selecting synonyms or phonetic equivalents, etc.)(Harris, paragraphs [0046 and 0069]); and

integrating the selected search parameters with the search query (the query translator 34 can utilize the specified dictionary 18 including one or more of relevant content and/or values from a data source 20, configuration settings, profiles according to users and/or groups of users, business riles, terminology definitions or specifications, synonyms, a language translator, a natural language processor, etc., and query information to generate a customized search or query)(Harris, paragraphs [0059, 0062-0063 and 0069])."

In regard to Claim 23, the Office Action states that "Claims 23-33 are corresponding claims of process claims 12-22; therefore, they are rejected under the same rationale."

Applicant disagrees that Claim 12 and Claim 23 are anticipated by Harris. Applicant reserves the right to present the same or similar claims in a related Application. For sake of convenience in prosecution, Applicant has amended Claim 12 and Claim 23, as discussed below, to particularly point out and distinctly claim some of the preferred embodiments of the invention.

Applicant has amended independent Claim 12, to claim a process, comprising the steps of:

soliciting from a user at least one user-specified search parameter;

associating the solicited user-specified search parameters with the user;

storing the associated solicited user-specified search parameters for use in at least one future search;

receiving a search query from the user;

determining the subject matter of the received search query;

selecting one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored

associated solicited user-specified search parameters with the determined subject matter of the received search query; and

integrating the selected stored associated solicited user-specified search parameters with a search based on the received search query.

Applicant has amended independent Claim 23, to claim a system, comprising:

means for soliciting from a user at least one user-specified search parameter;

means for associating solicited user-specified search parameters with the user;

means for storing the associated solicited user-specified search parameters for use in at least one future search;

an initial search query received from the user;

a determination of subject matter of the search query;

means for selection of one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited search parameters with the determined subject matter of the received search query; and

means for integrating the selected stored associated solicited user-specified search parameters with a search based on the received search query.

Support is seen in the Application as filed, at least on page 23, line 23 to page 29, line 14, in Claims 1-33; and in Figures 12-16.

In regard to Claim 12 as amended, while Harris describes that "search results can be customized using user preference or profile information", as seen at least in the abstract, Harris does not disclose a process that, *inter alia*, comprises the steps of:

associating solicited user-specified search parameters with the user;

storing the associated solicited user-specified search parameters for use in at least one future search;

receiving a search query from the user;

determining the subject matter of the received search query;

selecting one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored

associated solicited user-specified search parameters with the determined subject matter of the received search query; and

integrating the selected stored associated solicited user-specified search parameters with a search based on the received search query.

Applicant therefore submits that independent Claim 12, as amended, overcomes the rejection under 35 U.S.C. §102(b) as being unpatentable over Harris.

In regard to Claim 23 as amended, while Harris describes that "search results can be customized using user preference or profile information", as seen at least in the abstract, Harris does not disclose a system that, *inter alia*, comprises:

means for associating solicited user-specified search parameters with the user;

means for storing the associated solicited user-specified search parameters for use in at least one future search;

an initial search query received from the user;

a determination of subject matter of the search query;

means for selection of one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited search parameters with the determined subject matter of the received search query; and

means for integrating the selected stored associated solicited user-specified search parameters with a search based on the received search query.

Applicant therefore submits that independent Claim 23, as amended, overcomes the rejection under 35 U.S.C. §102(b) as being unpatentable over Harris.

The Examiner bears the burden of establishing a *prima facie* case of anticipation (In re King, 801 F.2d 1324, 1327, 231 USPQ 136, 138-139 (Fed. Cir. 1986)). The prior art reference must disclose each element of the

claimed invention, as correctly interpreted, and as arranged in the claim (Lindermann Maschinefabrik Gmbh v. American Hoist & Derrick Co., 730 F.2d 1452, 221 USPQ 481, 485 (Fed. Cir. 1984)). A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. The identical invention must be shown in as complete detail as is contained in the claim (MPEP 2131).

As dependent claims 13-22, and new claims 46-51 as discussed below, depend from amended independent Claim 12, and as dependent claims 24-33, and new claims 34-39 as discussed below, depend from amended independent Claim 23, and inherently contain all the limitations of the claims they depend from, they are seen to be patentable as well.

Other Amendments.

Applicant has amended Figure 2 and Figure 3, to correct minor errors. Applicant has amended the Specification, to correct minor grammatical errors. Applicant has amended Claims 11, 22 and 33, to provide proper antecedent terminology.

Applicant has added dependent Claims 34-51, to more particularly point out and distinctly claim preferred embodiments of the claimed invention. Support for Claims 34, 40 and 46 is seen in the Application as filed, at least on page 26, line 29 to page 27, line 15. Support for Claims 35, 41 and 47 is seen in the Application as filed, at least on page 26, lines 33-37. Support for Claims 36, 37, 42, 43, 48 and 49 is seen in the Application as filed, at least on page 26, lines 9-14 and 22-27, and in Figure 15. Support for Claims 38, 44 and 50 is seen in the Application as filed, at least on page 27, lines 1-4, and in Figure 15. Support for Claims 39, 45 and 51 is seen in the Application as filed, at least on page 23, line 23 to page 29, line 14, and in Figures 1, and in Figures 1-7 and 15.

CONCLUSION

For the foregoing reasons, the claims in the present application are patentably distinguished over the cited references. Applicant also submits that the amendments do not introduce new matter into the Application. Based on the foregoing, Applicant considers the invention to be in condition for allowance. Applicant earnestly solicits the Examiner's withdrawal of the rejections set forth in the prior Office Action, such that a Notice of Allowance is forwarded to Applicant, and the present application is therefore allowed to issue as a United States Patent.

Respectfully Submitted,

Michael A. Glenn
Reg. No. 30,176

Customer No. 22862

Amendments to the Drawings

In Figure 2, please replace the reference character "22a" with "22b", as shown on the replacement sheet.

In Figure 3, please replace the text legend "Delivery or Enhanced Results" associated with reference character 41 with ""Delivery of Enhanced Results"", as shown on the replacement sheet.

Amendments to the Specification

In page 24, lines 30-33 of the Application, please replace the paragraph with the following paragraph, as amended:

Figure 13 is a schematic diagram 160 of ~~an~~ a user-specified search parameter 148 and subject validities, *i.e.* rankings 170. A search parameter 148 typically comprises a parameter value 166 entered by a user USR, such as within a parameter value entry window 164 within a user interface 142 (FIG. 12).

In page 25, lines 8-13 of the Application, please replace the paragraph with the following paragraph, as amended:

Based upon the determined subject matter 168 of a search query 13, a user specified parameter 148 having a ~~rankings~~ ranking 170 which is determined to be applicable to the search may either be used in conjunction with primary search terms, *i.e.* to further limit search results, or may be used to sort the results of a query based upon primary search criteria, wherein the sort is based upon the applicability ranking 170 of one or more generalized search parameters 148.

Status of the Claims

1. (Currently Amended) A process, comprising the steps of:

soliciting from a user a set of one or more search parameters specified by the user;

associatively storing ~~a~~ the set of solicited search parameters specified by ~~a~~ the user for at least one future search;

receiving a search query from the user;

determining the subject matter of the search query;

selecting one or more of the associatively stored solicited search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query; and

integrating the selected ones of the associatively stored solicited search parameters ~~into a~~ with the search query ~~based upon subject matter of the search query~~.

2. (Original) The process of Claim 1, further comprising the step of:

returning results of the search query to the user.

3. (Original) The process of Claim 1, further comprising the step of:

sending results of the search query to a recipient.

4. (Original) The process of Claim 3, wherein the recipient is selected by the user.

5. (Original) The process of Claim 3, wherein the recipient is determined based on the results of the search query.

6. (Original) The process of Claim 3, wherein the recipient is determined based on information input by the user.

7. (Original) The process of Claim 1, further comprising the steps of:

receiving information from a recipient; and

selectively sending results of the search query to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user.

8. (Original) The process of Claim 7, wherein the received information comprises an interest in the results of the search query by the recipient.

9. (Original) The process of Claim 1, wherein the subject matter is explicitly indicated by the user.

10. (Original) The process of Claim 1, wherein the subject matter is inferentially determined based upon user input.

11. (Currently Amended) The process of Claim 1, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, further comprising the step of:

determining the selected ones of the associatively stored solicited search parameters based upon the selectable applicability.

12. (Currently Amended) A process, comprising the steps of:

soliciting from a user at least one user-specified search parameter;

~~receiving at least one search parameter specified by a user;~~

associating the solicited user-specified search parameters with the user;

storing the associated solicited user-specified search parameters for use in at least one future search;

receiving a search query from the user;

determining the subject matter of the received search query;

selecting one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited user-specified search parameters with the determined subject matter of the received search query; and

integrating the selected stored associated solicited user-specified search parameters with a search based on the received search query.

13. (Original) The process of Claim 12, further comprising the step of:

returning results of the search query to the user.

14. (Original) The process of Claim 12, further comprising the step of:

sending results of the search query to a recipient.

15. (Original) The process of Claim 14, wherein the recipient is selected by the user.

16. (Original) The process of Claim 14, wherein the recipient is determined based on the results of the search query.

17. (Original) The process of Claim 14, wherein the recipient is determined based on information input by the user.

18. (Original) The process of Claim 12, further comprising the steps of:

receiving information from a recipient; and

selectively sending results of the search query to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user.

19. (Original) The process of Claim 18, wherein the received information comprises an interest in the results of the search query by the recipient.

20. (Original) The process of Claim 12, wherein the subject matter is explicitly indicated by the user.

21. (Original) The process of Claim 12, wherein the subject matter is inferentially determined based upon user input.

22. (Currently Amended) The process of Claim 12, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, and further comprising the step of:

determining the selected ones of the stored <u>associated solicited user-specified</u> search parameters based upon the selectable applicability.

23. (Currently Amended) A system, comprising:

means for soliciting from a user at least one user-specified search parameter ~~specified by a user~~;

means for associating solicited user-specified search parameters with the user;

means for storing the associated solicited user-specified search parameters for use in at least one future search;

an initial search query received from the user;

a determination of subject matter of the search query;

means for selection of one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited search parameters with the determined subject matter of the received search query; and

means for integrating ~~the initial search query with~~ the selected stored associated solicited user-specified search parameters with a search based on the received search query.

24. (Original) The system of Claim 23, wherein results of the integrated search query are sent to the user.

25. (Original) The system of Claim 23, wherein results of the integrated search query are sent to a recipient.

26. (Original) The system of Claim 25, wherein the recipient is selectable by the user.

27. (Original) The system of Claim 25, wherein the recipient is based on the results.

28. (Original) The system of Claim 25, wherein the recipient is based on information input by the user.

29. (Original) The system of Claim 23, further comprising:

information received from a recipient; and

a selective transmission of the results of the integrated search query to the recipient based upon any of the received information and a selection of the recipient by the user.

30. (Original) The system of Claim 29, wherein the received information comprises an interest in the enhanced results.

31. (Original) The system of Claim 23, wherein the determination of the subject matter is explicitly indicated by the user.

32. (Original) The system of Claim 23, wherein the determination of the subject matter is inferentially determined based upon user input.

33. (Currently Amended) The system of Claim 23, wherein the <u>stored associated solicited</u> user-specified search parameters comprise a selectable applicability based upon subject matter, and wherein the means for selection is based upon the selectable applicability.

34. (New) The system of Claim 25, wherein the recipient is any of a selected peer, a friend, a family relative, a student, a neighbor, any of a person or entity within a zip code region, and a member of a network of people associated with the user.

35. (New) The system of Claim 34, wherein the network of people comprises any of people with similar interests and a chain of friends.

36. (New) The system of Claim 23, further comprising:

means for soliciting the user to provide at least one additional parameter after receipt of the query at the time of the search.

37. (New) The system of Claim 36, wherein the solicited additional parameter is a missing parameter.

38. (New) The system of Claim 23, further comprising:

means for receiving at least one source selection from the user; and

means for modifying the search query based upon information from the selected source.

39. (New) The system of Claim 23, wherein the step of integrating the selected ones of the stored associated solicited user-specified search parameters with the search query comprises any of an inclusion of the selected ones of the stored associated solicited user-specified search parameters with the search query, a refinement of a search result using the selected ones of the stored associated solicited user-specified parameters, and a sorting of a search result using the selected ones of the stored associated solicited user-specified search parameters.

40. (New) The process of Claim 3, wherein the recipient is any of a selected peer, a friend, a family relative, a student, a neighbor, any of a person or entity within a zip code region, and a member of a network of people associated with the user.

41. (New) The process of Claim 40, wherein the network of people comprises any of people with similar interests and a chain of friends.

42. (New) The process of Claim 1, further comprising the steps of:

soliciting the user to provide at least one parameter after receipt of the query at the time of the search.

43. (New) The process of Claim 42, wherein the solicited parameter is a missing parameter.

44. (New) The process of Claim 1, further comprising the steps of:

receiving at least one source selection from the user; and

modifying the search query based upon information from the selected source.

45. (New) The process of Claim 1, wherein the step of integrating the selected ones of the associatively stored solicited search parameters with the search query comprises any of an inclusion of the selected ones of the associatively stored solicited search parameters with the search query, a refinement of a search result using the selected ones of the associatively stored solicited search parameters, and a sorting of a search result using the selected ones of the associatively stored solicited search parameters.

46. (New) The process of Claim 14, wherein the recipient is any of a selected peer, a friend, a family relative, a student, a neighbor, any of a person or entity within a zip code region, and a member of a network of people associated with the user.

47. (New) The process of Claim 46, wherein the network of people comprises any of people with similar interests and a chain of friends.

48. (New) The process of Claim 12, further comprising the steps of:

soliciting the user to provide at least one parameter after receipt of the query at the time of the search.

49. (New) The process of Claim 48, wherein the solicited parameter is a missing parameter.

50. (New) The process of Claim 12, further comprising the steps of:

receiving at least one source selection from the user; and

modifying the search query based upon information from the selected source.

51. (New) The process of Claim 12, wherein the step of integrating the selected ones of the stored associated solicited user-specified search parameters with the search query comprises any of an inclusion of the selected ones of the stored associated solicited user-specified search parameters with the search query, a refinement of a search result using the selected ones of the stored associated solicited user-specified search parameters, and a sorting of a search result using the selected ones of the stored associated solicited user-specified search parameters.

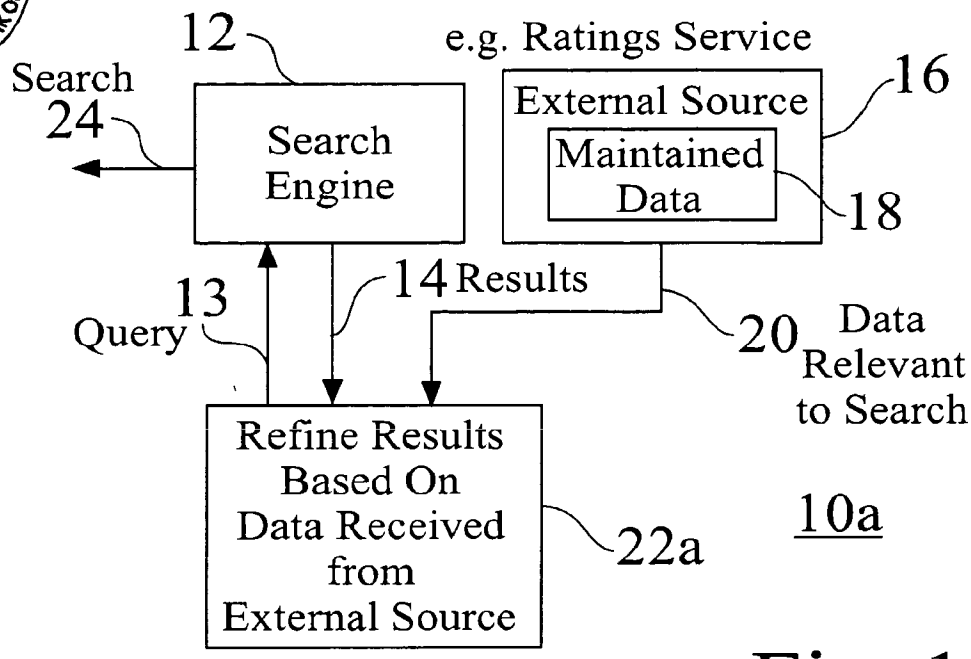


Fig. 1

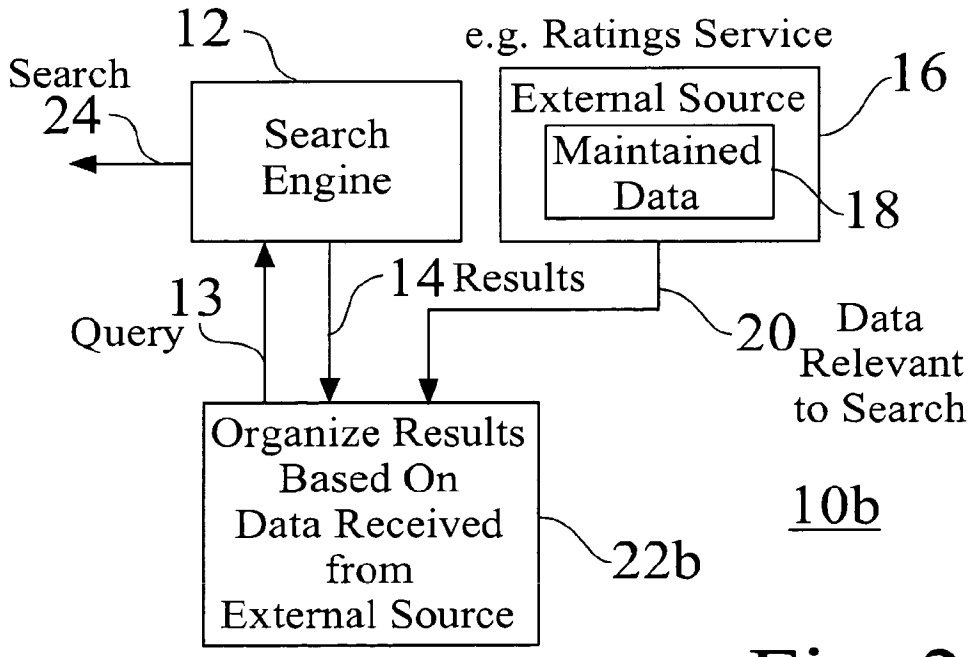


Fig. 2

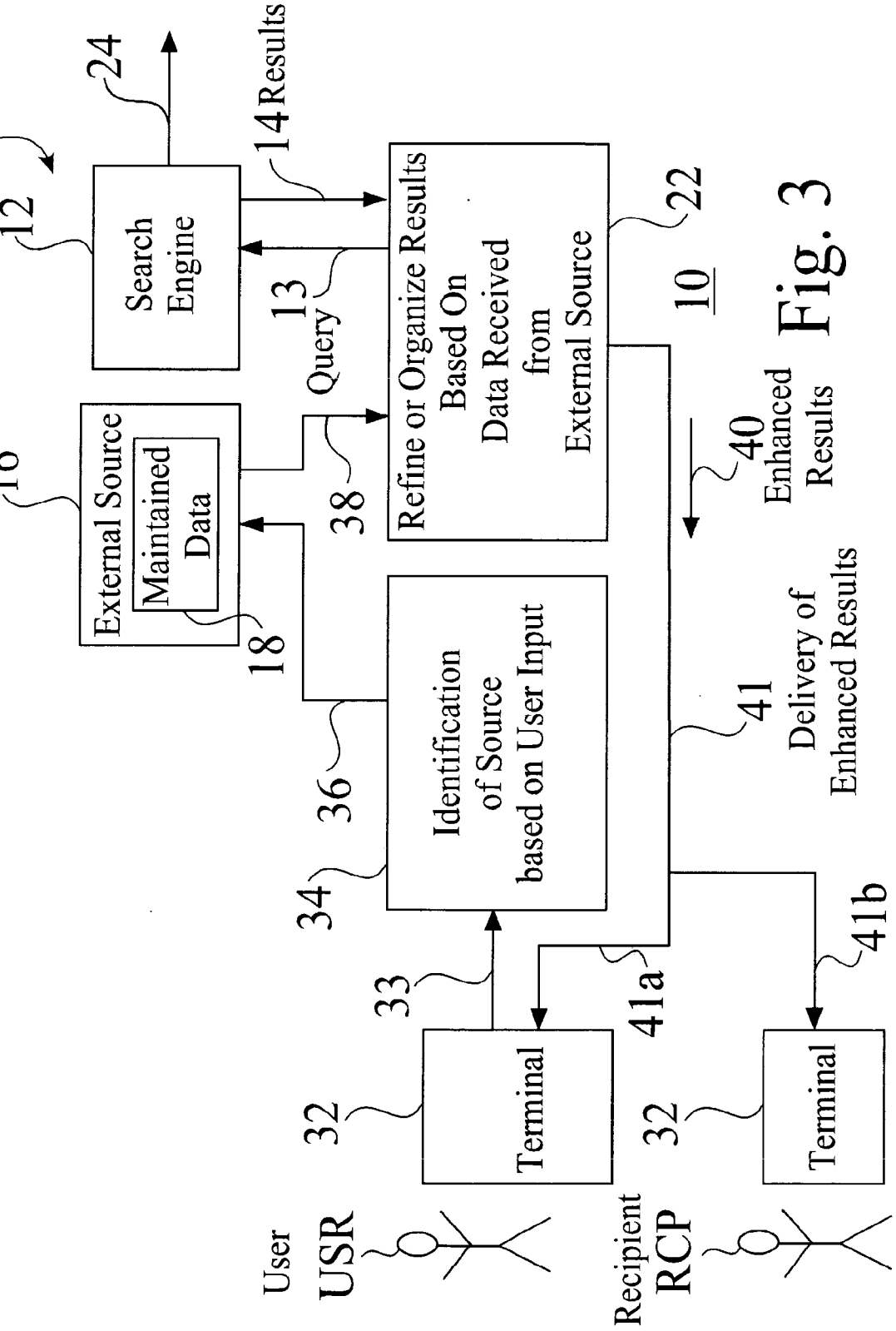

30
12
Search Engine
24
14 Results
13 Query
Refine or Organize Results Based On Data Received from External Source
22
10
16
External Source
Maintained Data
18
38
40 Enhanced Results
36
Identification of Source based on User Input
34
41 Delivery of Enhanced Results
33
41a
41b
32
Terminal
32
Terminal
User
USR
Recipient
RCP

Fig. 3

PTO/SB/92 (12-97)
Approved for use through 9/30/00. OMB 0651-0031
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Certificate of Mailing under 37 CFR 1.10

Express Mail Label No. EV 636913125 US

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as Express Mail in an envelope addressed to:

COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, VA 22313-1450

Docket No. AOL0142

on October 13, 2005
Date

Signature

Kristianne Serrano
Typed or printed name of person signing Certificate

Note: Each paper must have its own certificate of mailing, or this certificate must identify each submitted paper.

Attached to this coversheet please find the following documents:

- Certificate of Mailing (1 page);
- Response (19 pages);
- Replacement Drawings (2 pages); and
- Return Postcard

PATENT APPLICATION FEE DETERMINATION RECORD
Effective October 1, 2003

Application or Docket Number: 10685749

CLAIMS AS FILED - PART I

	(Column 1)	(Column 2)
TOTAL CLAIMS	33	
FOR	NUMBER FILED	NUMBER EXTRA
TOTAL CHARGEABLE CLAIMS	33 minus 20=	* 13
INDEPENDENT CLAIMS	3 minus 3 =	* 0
MULTIPLE DEPENDENT CLAIM PRESENT	☐	

* If the difference in column 1 is less than zero, enter "0" in column 2

SMALL ENTITY TYPE ☐			OR	OTHER THAN SMALL ENTITY	
RATE	FEE			RATE	FEE
BASIC FEE	385.00		OR	BASIC FEE	770.00
X$ 9=			OR	X$18=	234
X43=			OR	X86=	
+145=			OR	+290=	
TOTAL			OR	TOTAL	1004

CLAIMS AS AMENDED - PART II

10-13-05

AMENDMENT A

	(Column 1)		(Column 2)	(Column 3)
	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA
Total	* 51	Minus	** 33	= 18
Independent	* 3	Minus	*** 3	= 0
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM	☐			

1, 12, 23,

SMALL ENTITY			OR	OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE			RATE	ADDITIONAL FEE
X$ 9=			OR	X$18=	360.40
X43=			OR	X86=	
+145=			OR	+290=	
TOTAL ADDIT. FEE			OR	TOTAL ADDIT. FEE	

AMENDMENT B

	(Column 1)		(Column 2)	(Column 3)
	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM	☐			

RATE	ADDITIONAL FEE			RATE	ADDITIONAL FEE
X$ 9=			OR	X$18=	
X43=			OR	X86=	
+145=			OR	+290=	
TOTAL ADDIT. FEE			OR	TOTAL ADDIT. FEE	

AMENDMENT C

	(Column 1)		(Column 2)	(Column 3)
	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM	☐			

RATE	ADDITIONAL FEE			RATE	ADDITIONAL FEE
X$ 9=			OR	X$18=	
X43=			OR	X86=	
+145=			OR	+290=	
TOTAL ADDIT. FEE			OR	TOTAL ADDIT. FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20."
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3."
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

A

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

22862 7590 11/25/2005

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

EXAMINER	
NGUYEN, QUANG N	

ART UNIT	PAPER NUMBER
2141	

DATE MAILED: 11/25/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/685,749	FISH, EDMUND J.	
	Examiner	Art Unit	
	Quang N. Nguyen	2141	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *13 October 2005*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-51* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-51* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *13 October 2005* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

Detailed Action

1. This Office Action is in response to the Amendment filed on 10/13/2005. Claims 1, 11-12, 22-23 and 33 have been amended. Claims 34-51 have been added as new claims. Claims 1-51 are presented for examination.

Drawings

2. The drawings were received on 10/13/2005. These drawings are acceptable.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

4. Claims 1-51 are rejected under 35 U.S.C. 102(b) as being anticipated by Harris (US 2002/0059204 A1).

5. As to claim 1, Harris teaches a process, comprising the steps of:

soliciting from a user a set of one or more search parameters specified by the user *(data pertaining to received query information, customized search queries, etc., from a user of the initiating device 12 can be an input to a dictionary 18)* (Harris, paragraphs [0044, 0052 and 0059]);

associatively storing the set of solicited search parameters specified by the user for at least one future search *(storing log files of received query information, customized search queries, generated search results, etc., from the user of initiating device 12 in a dictionary 18 which is later used by the query translator 34 to generate a customized query)* (Harris, paragraphs [0044, 0055 and 0062]);

receiving a search query from the user *(the query information "How far is bank XXX from home?" initially submitted by the user)* (Harris, paragraph [0069]);

determining the subject matter of the search query *(i.e., parsing the received information, performing context analysis, etc.)* (Harris, paragraphs [0046 and 0069]);

selecting one or more of the associatively stored solicited search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query *(selecting data pertinent to the user "home" and data related to another sensor or entity such as mapping program or distance calculator, selecting synonyms or phonetic equivalents, etc., based upon received profile and/or identity information to prevent the query information from being applied to unauthorized data source)* (Harris, paragraphs [0046, 0052-0053 and 0069]); and

integrating the selected ones of the associatively stored solicited search parameters with a search query *(such data pertinent to the user "home" and data related to another sensor or entity such as mapping program or distance calculator, selecting synonyms or phonetic equivalents, etc., based upon received profile and/or identity information can be integrated to generate a customized search or query)* (Harris, paragraphs [0062-0063 and 0069]).

6. As to claim 2, Harris teaches the process of claim 1, further comprising the step of returning results of the search query to the user *(the customized search results can be provided to server 16 for display)* (Harris, paragraphs [0062-0063]).

7. As to claims 3-6, Harris teaches the process of claim 1, further comprising the step of sending results of the search query to a recipient selected by the user *(i.e., based on information input by the user)* or based on the results of the search query *(the query search results 108 can be transferred to the requested entity or another entity designated or otherwise specified to receive the query results 108, for example, if the query search results needs to be additionally filtered/formatted/aggregated then it can be returned to the dictionary 18 or the system server 14 for formatting, arranging, aggregating before transferring, forwarding or displaying to the initiating device 12 or to another recipient)* (Harris, paragraph [0031 and 0060]).

8. As to claims 7-8, Harris teaches the process of claim 1, further comprising the steps of:

receiving information from a recipient comprising an interest in the results of the search query *(receiving interest in verifying a quantity of a stock from the administrator and/or another interested party)*; and

selectively sending results of the search query to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user *(once a stock quantity reaches a predetermined value, an email can be sent to the administrator and/or another interested party)* (Harris, paragraph [0056]).

9. As to claim 9, Harris teaches the process of claim 1, wherein the subject matter is explicitly indicated by the user *(the query information initially submitted by the user)* (Harris, paragraph [0069]).

10. As to claim 10, Harris teaches the process of claim 1, wherein the subject matter is inferentially determined based upon user input (Harris, paragraph [0046]).

11. As to claim 11, Harris teaches the process of claim 1, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, further comprising the step of: determining the selected ones of the associatively stored solicited search parameters based upon the selectable applicability *(selecting data pertinent to the user "home" and data related to another sensor or entity such as*

mapping program or distance calculator, selecting synonyms or phonetic equivalents, etc., based upon received profile and/or identity information to prevent the query information from being applied to unauthorized data source) (Harris, paragraphs [0046, 0052-0053 and 0069]).

12. As to claim 12, Harris teaches a process, comprising the steps of:

soliciting from a user at least one user-specified search parameter *(data pertaining to received query information, customized search queries, etc., from a user of the initiating device 12 can be an input to a dictionary 18)* (Harris, paragraphs [0044, 0052 and 0059]);

associating the solicited user-specified search parameters with the user *(customization of a dictionary 18 can be performed and applied according to a particular user, or a group of users)* (Harris, paragraphs [0057-0059]);

storing the associated solicited user-specified search parameters for use in at least one future search *(storing log files of received query information, customized search queries, generated search results, etc., from the user of initiating device 12 in a dictionary 18 which is later used by the query translator 34 to generate a customized query)* (Harris, paragraphs [0044, 0055 and 0062]);

receiving a search query from the user *(the query information "How far is bank XXX from home?" initially submitted by the user)* (Harris, paragraph [0069]);

determining the subject matter of the received search query *(i.e., parsing the received information, performing context analysis, etc.)* (Harris, paragraphs [0046 and 0069]);

selecting one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited user-specified search parameters with the determined subject matter of the received search query *(selecting data pertinent to the user "home" and data related to another sensor or entity such as mapping program or distance calculator, selecting synonyms or phonetic equivalents, etc., based upon received profile and/or identity information to prevent the query information from being applied to unauthorized data source)* (Harris, paragraphs [0046, 0052-0053 and 0069]); and

integrating the selected stored associated solicited user-specified search parameters with a search based on the received search query *(such data pertinent to the user "home" and data related to another sensor or entity such as mapping program or distance calculator, selecting synonyms or phonetic equivalents, etc., based upon received profile and/or identity information can be integrated to generate a customized search or query)* (Harris, paragraphs [0062-0063 and 0069]).

13. Claims 13-22 contain similar limitations as claims 2-11; therefore, they are rejected under the same rationale.

14. Claims 23-33 are corresponding system claims of process claims 12-22; therefore, they are rejected under the same rationale.

15. Claims 34-35 contain similar limitations as claims 7-8; therefore, they are rejected under the same rationale.

16. As to claims 36-37, Harris teaches the system of claim 23, further comprising means for soliciting the user to provide at least one additional parameter (or missing parameter) after receipt of the query of the time of search *(in another embodiment, a user can be prompted for preference criteria for further searching/filtering, wherein the preferences can indicate suppliers, price, and other search-related criteria that can be submitted with the natural language search terms)* (Harris, paragraph [0075]).

17. As to claim 38, Harris teaches the system of claim 23, further comprising:

means for receiving at least one source selection from the user *(e.g., receiving a user-specific preference criteria indicating suppliers, price, and other search-related criteria)*; and

means for modifying the search query based upon information from the selected source *(the selected user-specific preference indicating suppliers, price, and/or other search-related criteria that can be submitted with the natural language search terms)* (Harris, paragraph [0075]).

18. As to claim 39, Harris teaches the system of claim 23, wherein the step of integrating the selected ones of the stored associated solicited user-specified search parameters with the search query comprises any of an inclusion of the selected ones of the stored associated solicited user-specified search parameters with the search query, a refinement of a search result using the selected ones of the stored associated solicited user-specified parameters *(in another embodiment, a user can be prompted for a user-specific preference criteria to be included in the search query for further searching/filtering, wherein the selected user-specific preference can indicate suppliers, price, and/or other search-related criteria that can be submitted with the natural language search terms)*, and a sorting of a search result using the selected ones of the stored associated user-specified search parameters *(search results can be displayed according to a user identification and/or user-specific preferences)* (Harris, paragraph [0075]).

19. Claims 40-45 contain similar limitations as claims 34-39; therefore, they are rejected under the same rationale.

20. Claims 46-51 contain similar limitations as claims 34-39; therefore, they are rejected under the same rationale.

21. Applicant's arguments as well as request for reconsideration filed on 10/13/2005 have been fully considered but they are not deemed to be persuasive.

22. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

23. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Quang N. Nguyen whose telephone number is (571) 272-3886.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's SPE, Rupal Dharia, can be reached at (571) 272-3880. The fax phone number for the organization is (571) 273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

RUPAL DHARIA
SUPERVISORY PATENT EXAMINER

Index of Claims

Application/Control No.	Applicant(s)/Patent under Reexamination
10/685,749	FISH, EDMUND J.
Examiner	**Art Unit**
Quang N. Nguyen	2141

√	Rejected	–	(Through numeral) Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

Claim		Date	
Final	Original	7/7/05	11/20/05
	1	√	√
	2	√	√
	3	√	√
	4	√	√
	5	√	√
	6	√	√
	7	√	√
	8	√	√
	9	√	√
	10	√	√
	11	√	√
	12	√	√
	13	√	√
	14	√	√
	15	√	√
	16	√	√
	17	√	√
	18	√	√
	19	√	√
	20	√	√
	21	√	√
	22	√	√
	23	√	√
	24	√	√
	25	√	√
	26	√	√
	27	√	√
	28	√	√
	29	√	√
	30	√	√
	31	√	√
	32	√	√
	33	√	√
	34		√
	35		√
	36		√
	37		√
	38		√
	39		√
	40		√
	41		√
	42		√
	43		√
	44		√
	45		√
	46		√
	47		√
	48		√
	49		√
	50		√

Claim		Date
Final	Original	11/20/05
	51	√
	52	
	53	
	54	
	55	
	56	
	57	
	58	
	59	
	60	
	61	
	62	
	63	
	64	
	65	
	66	
	67	
	68	
	69	
	70	
	71	
	72	
	73	
	74	
	75	
	76	
	77	
	78	
	79	
	80	
	81	
	82	
	83	
	84	
	85	
	86	
	87	
	88	
	89	
	90	
	91	
	92	
	93	
	94	
	95	
	96	
	97	
	98	
	99	
	100	

Claim		Date
Final	Original	
	101	
	102	
	103	
	104	
	105	
	106	
	107	
	108	
	109	
	110	
	111	
	112	
	113	
	114	
	115	
	116	
	117	
	118	
	119	
	120	
	121	
	122	
	123	
	124	
	125	
	126	
	127	
	128	
	129	
	130	
	131	
	132	
	133	
	134	
	135	
	136	
	137	
	138	
	139	
	140	
	141	
	142	
	143	
	144	
	145	
	146	
	147	
	148	
	149	
	150	

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

Bib Data Sheet

CONFIRMATION NO. 6161

SERIAL NUMBER	FILING DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
10/685,749	10/14/2003 RULE	709	2141	AOL0142

APPLICANTS

Edmund J. Fish, Dulles, VA;

** CONTINUING DATA ********** QN 11/20/05

** FOREIGN APPLICATIONS ********** QN 11/20/05

IF REQUIRED, FOREIGN FILING LICENSE GRANTED
** 01/21/2004

		STATE OR	SHEETS	TOTAL	INDEPENDENT
Foreign Priority claimed	☐ yes ☑ no				
35 USC 119 (a-d) conditions met	☐ yes ☑ no ☐ Met after Allowance	COUNTRY	DRAWING	CLAIMS	CLAIMS
Verified and Acknowledged	Examiner's Signature / Initials QN	VA	15	33	3

ADDRESS
22862
GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK , CA
94025

TITLE
Search enhancement system having personal search parameters

FILING FEE RECEIVED 1134	FEES: Authority has been given in Paper No. _____ to charge/credit DEPOSIT ACCOUNT No. _____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

RECEIVED
CENTRAL FAX CENTER
JAN 19 2006

FAX TO (571) 272-3886 / 273-8300

PTOL-413A (08-03)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Applicant Initiated Interview Request Form

Application No.: 10 / 685,749 First Named Applicant: Edmund J. Fish
Examiner: Nguyen, Quang N. Art Unit: 2141 Status of Application: Pending

Tentative Participants:
(1) Donald M. Hendricks (831) 656-0598 (2) _____

(3) _____ (4) _____

Proposed Date of Interview: 20 Jan 2006 **Proposed Time:** 3:00PM EST **(AM/PM)**

Type of Interview Requested:
(1) [x] **Telephonic** (2) [ ] **Personal** (3) [ ] **Video Conference**

Exhibit To Be Shown or Demonstrated: [ ] **YES** [x] **NO**
If yes, provide brief description: _____

Issues To Be Discussed

Issues (Rej., Obj., etc)	Claims/ Fig. #s	Prior Art	Discussed	Agreed	Not Agreed
(1) Rej.	C-1,12,23	Harris	[ ]	[ ]	[ ]
(2) for Issue (1)	F-1-10,15	Harris	[ ]	[ ]	[ ]
(3)			[ ]	[ ]	[ ]
(4)			[ ]	[ ]	[ ]

[ ] Continuation Sheet Attached

Brief Description of Arguments to be Presented:
Discussion of possible amendments to independent Claims 1,12 & 23, as patentably distinguishable from cited art.

An interview was conducted on the above-identified application on _____.

NOTE:
This form should be completed by applicant and submitted to the examiner in advance of the interview (see MPEP § 713.01).
This application will not be delayed from issue because of applicant's failure to submit a written record of this interview. Therefore, applicant is advised to file a statement of the substance of this interview (37 CFR 1.133(b)) as soon as possible.

Donald M. Hendricks

(Applicant/Applicant's Representative Signature) (Examiner/SPE Signature)

RECEIVED
CENTRAL FAX CENTER

JAN 2 4 2006

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Edmund J. Fish **Docket No.:** AOL0142

Serial No.: 10/685,749 **Art Unit:** 2141

Filed: 14 October 2003 **Examiner:** Nguyen, Quang N.

Title: Search Enhancement System Having Personal Search Parameters

24 January 2006

Commissioner for Patents
Mail Stop RCE
P.O. Box 1450
Alexandria, VA 22313-1450

PRELIMINARY AMENDMENT

Sir:

This is in response to the Office Action of 25 November 2005, for the above-identified patent application. The Examiner made the Action Final. Applicant Requests Continued Examination (RCE) in accordance with 37 CFR 1.114, and authorizes payment in compliance with 37 CFR 1.17(e) of $790, to be charged to the Glenn Patent Group Deposit Account No. 07-1445, Customer No. 22862.

An Examiner Interview was conducted by telephone on 20 January 2006. Donald M. Hendricks, Applicant's representative, and the Examiner discussed the present invention, as disclosed in the Application as filed and as proposed to be claimed in Claims 1, 12 and 23, in contrast to the Cited Art of record. Applicant specifically discussed what are believed to be patentable features of Claim 1, as proposed to be amended. The Examiner stated that the scope of the proposed Claims may require further Search and Examination.

REMARKS

35 U.S.C. § 102. Claim Rejections.

4-20. Claims 1-33 are rejected under 35 U.S.C. §102(b) as being anticipated by Harris (U.S. Patent Publication No. 2002/0059204 A1).

Hilton Davis / Festo Statement

Applicant has amended Claims 1, 12 and 23, for convenience in prosecution, and reserves the right to present the same or similar claims in a related Application. The amendments herein were not made for any reason related to patentability.

5-11. Applicant disagrees that Claim 1 is anticipated by Harris. Applicant reserves the right to present the same or similar claims in a related Application. For sake of convenience in prosecution, Applicant has amended Claim 1, as discussed below, to particularly point out and distinctly claim some of the preferred embodiments of the invention.

Applicant has amended independent Claim 1, to claim a process, comprising the steps of:

soliciting from a user a set of one or more general search parameters specified by the user;

soliciting from the user at least one ranking explicitly specified by the user, the ranking based on any of applicability and validity of at least one of the solicited general search parameters in regard to at least one subject matter;

associatively storing the set of solicited general search parameters and the rankings specified by the user for at least one future search initiated by the user;

receiving a search query from the user, the search query comprising one or more query parameters other than the previously solicited general search parameters;

determining the subject matter of the received search query;

selecting one or more of the associatively stored solicited general search parameters based upon any of applicability and validity of the

associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user; and

providing a refined search, wherein the refinement comprises any of

using the selected ones of the associatively stored solicited general search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the associatively stored solicited general search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the associatively stored solicited general search parameters.

Support is seen in the Application as filed, at least on page 12, lines 12-21; on page 13, line 38 to page 14, line 23; on page 15, line 1 to page 16, line 3; on page 16, line 29-38; on page 23, line 23 to page 29, line 14; in Claims 1-33; and in Figures 12-16.

Harris describes a "distributed search system and method" seen at least in the abstract, wherein:

> "A method and system for providing distributing a query to devices on a communications network, such as servers on the internet, using an application that can survey a subscriber's server that can include text documents and databases, and use the formatting information and data from the survey to create at least one dictionary customized to the subscriber's data sources. A user seeking information can initiate a search or query from an initiating device using keywords, natural language terms, connectors, expressions, etc., and the query can be transmitted to various subscriber customized dictionaries. The customized dictionaries can customize the query based on respective subscriber databases and text documents, text search engines, etc., to produce an customized query result. The query results can be filtered

and integrated for presentation to the initiating device. Search results can be customized using user preference or profile information."

Applicant respectfully submits that Claim 1, as amended, is patentably distinguished from the distributed search system and method of Harris.

Harris describes, at least in the Abstract, that the "query results can be filtered and integrated for presentation to the initiating device. Search results can be customized using user preference or profile information". Details of filtering schemes as described by Harris are seen at least in paragraph [0052], wherein:

"[0052] A user or system manager associated with a server 16 can provide or otherwise designate filtering schemes for providing search results. A subscriber, for example, can determine to exclude products from particular users based upon a received profile (i.e., age, etc.). Those with ordinary skill in the art will recognize that there are many parameters by which a subscriber can filter or otherwise customize the search results."

Applicant submits that, while a user in Harris may "provide or otherwise designate filtering schemes for providing search results", Harris does not solicit and/or receive a ranking explicitly specified by the user that is based on any of applicability and validity of at least one solicited general search parameter in regard to at least one subject matter. As well, Harris does not associatively store such an explicit ranking with the user for use in a future search initiated by the user, nor does Harris select one or more of the associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user.

Furthermore, Harris does not provide a refined search, wherein the refinement comprises any of

using the selected ones of the associatively stored solicited general search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the associatively stored solicited general search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the associatively stored solicited general search parameters;

wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user.

Harris describes that the "query information from a user of the initiating device 12 can be an input to a dictionary 18, and the dictionary output can be a customized SQL query that uses terminology, abbreviations, etc., derived from the survey engine" as seen at least in paragraph [0044], wherein:

"[0044] For example, if the data source 20 is a database, the survey engine can identify labels of tables, rows, and columns, and abbreviations of labels, when necessary. This survey information can be incorporated into a dictionary 18 to allow received query information to be properly translated for the database. In the case of a database, query information from a user of the initiating device 12 can be an input to a dictionary 18, and the dictionary output can be a customized SQL query that uses terminology, abbreviations, etc., derived from the survey engine. In some embodiments, the dictionary output can be a customized HTTP search string that can utilize a general access method that can be created for the data source (e.g., the HTTP search string can be formatted based on drop-down menus/boxes, radio button selections, and/or other general access provisions). In an example of a database embodiment, a database can be configured with columns or rows that relate to colors that are abbreviated, such as "Rd" for Red, "Bl" for Blue, etc. If query information is submitted with

> the word "Blue", the customized dictionary can cause a customized query to be formatted using "Bl" according to the survey information. Accordingly, row information and column information can be understood herein to include a characterization of the database information that can include header information, element information, extraneous information that can otherwise provide insight to the database, etc."

Applicant submits in regard to Harris that, while "the query information from a user of the initiating device 12 can be an input to a dictionary 18, and the dictionary output can be a customized SQL query that uses terminology, abbreviations, etc., derived from the survey engine", the described structure and process is significantly different that Claim 1, as amended, as Harris does not associatively store an explicit ranking specified by the user for use in a future search initiated by the user, nor does Harris select one or more of the associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user.

Harris describes the results of a search query, at least in paragraphs [0051]-[0052], wherein:

> "[0051] In some embodiments, the requested query information can provide significant search results. The methods and systems herein can provide the search results using a format that includes categories and sub-categories from which a user at the initiating device 12, for example, can further select. Aggregate data can also be presented with hyperlinks to detailed information to allow users to retrieve further information without providing further details, follow-up questions, etc. Those with ordinary skill in the art will thus recognize that although the methods and systems presented herein are provided with respect to an

internet illustration that includes browsers, etc., other interfaces can be utilized to retrieve query information and present query results.

[0052] A user or system manager associated with a server 16 can provide or otherwise designate filtering schemes for providing search results. A subscriber, for example, can determine to exclude products from particular users based upon a received profile (i.e., age, etc.). Those with ordinary skill in the art will recognize that there are many parameters by which a subscriber can filter or otherwise customize the search results."

Applicant submits that, while Harris describes that "the requested query information can provide significant search results", Harris does not provide a refined search, wherein the refinement comprises any of

using the selected ones of the associatively stored solicited general search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the associatively stored solicited general search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the associatively stored solicited general search parameters;

wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user.

As well, while Harris describes that a "user or system manager associated with a server 16 can provide or otherwise designate filtering schemes for providing search results", there is no disclosure or suggestion, express or implied, that Harris solicits and/or receives a ranking explicitly specified by the user that is based on any of applicability and validity of at least one solicited general search parameter in regard to at least one subject matter.

Furthermore, Harris does not associatively store a ranking explicitly specified by the user for use in a future search initiated by the user, nor does he select one or more associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user.

Harris also describes, in regard to system security, that query information can be received or otherwise associated with identity information, as seen at least in [0053], wherein:

> "[0053] The methods and systems can operate with security measures that can be established by a system manager related to a server 16. For example, query information can be received or otherwise associated with identity information. A dictionary 18 can be configured to prevent the query information from being applied to a data source for which the user is not allowed to otherwise access. In such an embodiment, the methods and systems herein can be incorporated behind a firewall."

Applicant submits that, while Harris describes that "query information can be received or otherwise associated with identity information", the described relationship between identity information and query information for security purposes by Harris lacks any disclosure or suggestion in regard to a selection of one or more of the associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one ranking that is explicitly specified by the user.

Harris also describes the use further details in regard to a dictionary (18), as seen at least in paragraph [0059], wherein:

> "[0059] Accordingly, a dictionary 18 can include one or more of relevant content and/or values from a data source 20, a representation of schemas, relationships, and category hierarchies from the data source 20, configuration settings according to users and/or groups of users, business rules, terminology definitions or specifications, synonyms, a language translator, a natural language processor, an output data formatter, and a log file generator."

Applicant submits that, while Harris describes "configuration settings according to users and/or groups of users, business rules, terminology definitions or specifications, synonyms, a language translator, a natural language processor, an output data formatter, and a log file generator", there is no disclosure or suggestion, express or implied, that Harris solicits and/or receives a ranking that us explicitly specified by the user that is based on any of applicability and validity of at least one solicited general search parameter in regard to at least one subject matter.

As well, Harris does not associatively store a ranking explicitly specified by the user for use in a future search initiated by the user, nor does Harris select one or more associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user.

Harris describes further details in regard to queries, as seen at least in paragraph [0069], wherein:

> "[0069] Accordingly, it can be understood that the methods and systems disclosed herein can be applied to a variety of queries. For example, the query information initially submitted can be a question, such as "How far is bank XXX from home?", such that data pertinent to the user (e.g., "home") and data related to a another sensor or entity (e.g., mapping program or distance calculator) can be integrated with

> the query information to provide an appropriate response. Questions can be presented in succession, and results from one question can be a basis as input to the dictionary for subsequent questions."

Applicant submits that, while Harris describes that "data pertinent to the user (e.g., "home") and data related to a another sensor or entity (e.g., mapping program or distance calculator) can be integrated with the query information to provide an appropriate response", Harris does not solicit and/or receive a ranking explicitly specified by the user that is based on any of applicability and validity of at least one solicited general search parameter in regard to at least one subject matter. As well, Harris does not associatively store such an explicit ranking with the user for use in a future search initiated by the user, nor does Harris select one or more of the associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user.

Harris describes further details of system and process embodiments, at least in [0075], wherein:

> "[0075] Although the methods and systems have been described relative to specific embodiments thereof, the methods and systems are not so limited. Obviously many modifications and variations may become apparent in light of the above teachings. For example, although the illustrated embodiment presented herein related to the internet, the methods and systems can be applied to searching other devices on other communications networks. Although HTML and XML languages were utilized to facilitate the searching, other languages may be utilized. The survey program can be run locally on a server, or remotely from another platform. Similarly, a dictionary can be accessed and/or managed locally or remotely using wired or wireless communications systems and methods. Results of customized queries can be aggregated at a server having multiple data sources, and

thereafter transferred to the system server that can aggregate the results. In another embodiment, an application residing on the initiating device can receive and aggregate the results for presentation on the device. In an embodiment, search results can be displayed according to a user identification and/or information stored in a central database or local memory that includes user-specific preferences. The preferences can indicate suppliers, price, and other search-relevant criteria that can be submitted with the natural language search terms. In embodiments utilizing a user account for preferences, the respective subscriber dictionaries can integrate the user's preferences with the search terms to increase the probability of satisfying a search query for a user. In another embodiment, a user can be prompted for preference criteria, while in another embodiment, user identification data can be entered and submitted with the search criteria. Upon receiving the user identification data, a dictionary can cause a pre-stored user profile to be extracted and incorporated into the search."

Applicant submits that, while Harris describes that "search results can be displayed according to a user identification and/or information stored in a central database or local memory that includes user-specific preferences", there is no disclosure or suggestion, express or implied, that any of the user identification, information or user-specific preferences comprise solicited and associatively stored rankings explicitly specified by the user that are based on any of applicability and validity of at least one solicited general search parameter in regard to at least one subject matter.

Applicant also submits that, while Harris describes that "a user can be prompted for preference criteria", there is no disclosure or suggestion, express or implied, that "preference criteria" prompted from the user comprises a ranking explicitly specified by the user that is based on any of applicability and validity of at least one solicited general search parameter in regard to at least one subject matter.

Furthermore, Applicant submits that, while Harris describes that "user identification data can be entered and submitted with the search criteria. Upon receiving the user identification data, a dictionary can cause a pre-stored user profile to be extracted and incorporated into the search," there is no disclosure or suggestion Harris, express or implied, that either "user identification data" that is "submitted with the search criteria" or "a pre-stored user profile" comprises solicited and associatively stored rankings that are explicitly specified by the user, that are based on any of applicability and validity of at least one solicited general search parameter in regard to at least one subject matter.

Applicant therefore submits that independent Claim 1, as amended, overcomes the rejection under 35 U.S.C. §102(b) as being unpatentable over Harris (U.S. Patent Publication No. 2002/0059204 A1).

The Examiner bears the burden of establishing a *prima facie* case of anticipation (In re King, 801 F.2d 1324, 1327, 231 USPQ 136, 138-139 (Fed. Cir. 1986)). The prior art reference must disclose each element of the claimed invention, as correctly interpreted, and as arranged in the claim (Lindemann Maschinefabrik Gmbh v. American Hoist & Derrick Co., 730 F.2d 1452, 221 USPQ 481, 485 (Fed. Cir. 1984)). A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. The identical invention must be shown in as complete detail as is contained in the claim (MPEP 2131).

As dependent claims 2-11 and 40-45 depend from amended independent Claim 1, and inherently contain all the limitations of the claims they depend from, they are seen to be patentable as well.

12-15. Applicant disagrees that Claim 12 and Claim 23 are anticipated by Harris. Applicant reserves the right to present the same or similar claims in a related Application. For sake of convenience in prosecution, Applicant has amended Claim 12 and Claim 23, as discussed below, to particularly point out and distinctly claim some of the preferred embodiments of the invention.

Applicant has amended independent Claim 12, to claim a process, comprising the steps of:

soliciting from a user at least one user-specified search parameter;

soliciting from the user at least one ranking explicitly specified by the user, the rankings based on any of applicability and validity of at least one of the solicited user-specified search parameters in regard to at least one subject matter;

associating the solicited user-specified search parameters and the explicitly specified rankings with the user;

storing the associated solicited user-specified search parameters and the explicitly specified rankings for use in at least one future search initiated by the user;

receiving a search query from the user, the search query comprising one or more query parameters other than the previously solicited user-specified search parameters;

determining the subject matter of the received search query;

selecting one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited user-specified search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user; and

providing a refined search, wherein the refinement comprises any of

using the selected ones of the stored associated solicited user-specified search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the stored associated solicited user-specified search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the stored associated solicited user-specified search parameters.

Applicant has amended independent Claim 23, to claim a system, comprising:

means for soliciting from a user at least one user-specified search parameter;

means for soliciting from the user one or more rankings ranking explicitly specified by the user, the rankings based on any of applicability and validity of at least one of the solicited user-specified search parameters in regard to at least one subject matter;

means for associating the solicited user-specified search parameters and the explicitly specified rankings with the user;

means for storing the associated solicited user-specified search parameters and the explicitly specified rankings for use in at least one future search initiated by the user;

means for receiving an initial search query from the user, the search query comprising one or more query parameters other than the previously solicited user-specified search parameters;

means for determining the subject matter of the received initial search query;

means for selection of one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user; and

means for providing a refined search, wherein the refinement comprises any of

means for using the selected ones of the stored associated solicited user-specified search parameters in conjunction with the received query parameters to perform the search;

means for performing the search with the received query parameters and subsequently refining search results with the selected ones of the stored associated solicited user-specified search parameters; and

means for performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the stored associated solicited user-specified search parameters.

Support is seen in the Application as filed, at least on page 12, lines 12-21; on page 13, line 38 to page 14, line 23; on page 15, line 1 to page 16, line 3; on page 16, line 29-38; on page 23, line 23 to page 29, line 14; in Claims 1-33; and in Figures 12-16.

In regard to Claim 12 as amended, while Harris describes that "search results can be customized using user preference or profile information", as seen at least in the abstract, [0044], [0051-0053], [0059], [0069] and [0075], Harris does not solicit and/or receive a ranking explicitly specified by the user that is based on any of applicability and validity of at least one solicited user-specified search parameters in regard to at least one subject matter. As well, Harris does not associatively store such an explicit ranking with the user for use in a future search initiated by the user. In addition, Harris does not select one or more of the associatively stored solicited user-specified search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user. Furthermore, Harris does not provide a refined search, wherein the refinement comprises any of

using the selected ones of the associatively stored solicited user-specified search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the associatively stored solicited user-specified search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the associatively stored solicited user-specified search parameters,

wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user.

Applicant therefore submits that independent Claim 12, as amended, overcomes the rejection under 35 U.S.C. §102(b) as being unpatentable over Harris (U.S. Patent Publication No. 2002/0059204 A1).

In regard to Claim 23 as amended, while Harris describes that "search results can be customized using user preference or profile information", as seen at least in the abstract, [0044], [0051-0053], [0059], [0069] and [0075], Harris does not provide means for soliciting and/or receiving a ranking explicitly specified by the user that is based on any of applicability and validity of at least one solicited user-specified search parameters in regard to at least one subject matter. As well, Harris does not provide means for associatively storing such an explicit ranking with the user for use in a future search initiated by the user. In addition, Harris does not provide means for selecting one or more of the associatively stored solicited user-specified search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user. Furthermore, Harris does not provide means for providing a refined search, wherein the refinement comprises any of

using the selected ones of the associatively stored solicited user-specified search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the associatively stored solicited user-specified search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the associatively stored solicited user-specified search parameters,

wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user.

Applicant therefore submits that independent Claim 23, as amended, overcomes the rejection under 35 U.S.C. §102(b) as being unpatentable over Harris (U.S. Patent Publication No. 2002/0059204 A1).

The Examiner bears the burden of establishing a *prima facie* case of anticipation (In re King, 801 F.2d 1324, 1327, 231 USPQ 136, 138-139 (Fed. Cir. 1986)). The prior art reference must disclose each element of the claimed invention, as correctly interpreted, and as arranged in the claim (Lindermann Maschinefabrik Gmbh v. American Hoist & Derrick Co., 730 F.2d 1452, 221 USPQ 481, 485 (Fed. Cir. 1984)). A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. The identical invention must be shown in as complete detail as is contained in the claim (MPEP 2131).

As dependent claims 13-22 and 46-51 depend from amended independent Claim 12, and as dependent claims 24-33 and 34-39 depend from amended independent Claim 23, and inherently contain all the limitations of the claims they depend from, they are seen to be patentable as well.

Other Amendments.

Applicant has amended Claims 2-3, 5, 7-8, 13-14, 16, 18-19, 24-25, 27 and 29-30, to provide proper antecedent terminology.

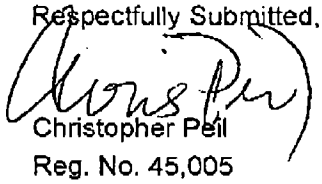
Applicant has amended dependent Claims 39, 45 and 51, to more particularly point out and distinctly claim preferred embodiments of the claimed invention. Support is seen in the Application as filed, at least on page 27, lines 6-9.

Applicant has also amended the Application to include a cross reference to a related application, which is also currently pending. Applicant submits that the prior omission of this cross-reference was inadvertent.

CONCLUSION

For the foregoing reasons, the claims in the present application are patentably distinguished over the cited reference. Applicant also submits that the amendments do not introduce new matter into the Application. Based on the foregoing, Applicant considers the invention to be in condition for allowance. Applicant earnestly solicits the Examiner's withdrawal of the rejections set forth in the prior Office Action, such that a Notice of Allowance is forwarded to Applicant, and the present application is therefore allowed to issue as a United States Patent.

Respectfully Submitted,



Christopher Peil
Reg. No. 45,005

Customer No. 22862

Amendments to the Specification

On page 1, line 3 of the Application, please insert the following Section, as entered:

CROSS REFERENCE TO RELATED APPLICATION

This Application is related to Edmund J. Fish, U.S. Application No. 10/685,747, entitled Search Enhancement System with Information from a Selected Source, filed 14 October 2003.

Status of the Claims

1. (Currently Amended) A process, comprising the steps of:

soliciting from a user a set of one or more general search parameters specified by the user;

soliciting from the user at least one ranking explicitly specified by the user, the rankings based on any of applicability and validity of at least one of the solicited general search parameters in regard to at least one subject matter;

associatively storing the set of solicited general search parameters and the rankings specified by the user for at least one future search initiated by the user;

receiving a search query from the user, the search query comprising one or more query parameters other than the previously solicited general search parameters;

determining the subject matter of the received search query;

selecting one or more of the associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user; and

~~integrating the selected ones of the associatively stored solicited search parameters with the~~ providing a refined search ~~query~~, wherein the refinement comprises any of

using the selected ones of the associatively stored solicited general search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the associatively stored solicited general search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the associatively stored solicited general search parameters.

2. (Currently Amended) The process of Claim 1, further comprising the step of:

returning results of the refined search ~~query~~ to the user.

3. (Currently Amended) The process of Claim 1, further comprising the step of:

sending results of the refined search ~~query~~ to a recipient.

4. (Original) The process of Claim 3, wherein the recipient is selected by the user.

5. (Currently Amended) The process of Claim 3, wherein the recipient is determined based on the results of the refined search ~~query~~.

6. (Original) The process of Claim 3, wherein the recipient is determined based on information input by the user.

7. (Currently Amended) The process of Claim 1, further comprising the steps of:

receiving information from a recipient; and

selectively sending results of the refined search ~~query~~ to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user.

8. (Currently Amended) The process of Claim 7, wherein the received information comprises an interest in the results of the refined search ~~query~~ by the recipient.

9. (Original) The process of Claim 1, wherein the subject matter is explicitly indicated by the user.

10. (Original) The process of Claim 1, wherein the subject matter is inferentially determined based upon user input.

11. (Previously Presented) The process of Claim 1, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, further comprising the step of:

determining the selected ones of the associatively stored solicited search parameters based upon the selectable applicability.

12. (Currently Amended) A process, comprising the steps of:

soliciting from a user at least one user-specified search parameter;

soliciting from the user at least one ranking explicitly specified by the user, the rankings based on any of applicability and validity of at least one of the solicited user-specified search parameters in regard to at least one subject matter;

associating the solicited user-specified search parameters and the explicitly specified rankings with the user;

storing the associated solicited user-specified search parameters and the explicitly specified rankings for use in at least one future search initiated by the user;

receiving a search query from the user, the search query comprising one or more query parameters other than the previously solicited user-specified search parameters;

determining the subject matter of the received search query;

selecting one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited user-specified search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user; and

~~integrating the selected stored associated solicited user-specified search parameters with a search based on the received~~ providing a refined search ~~query~~, wherein the refinement comprises any of

using the selected ones of the stored associated solicited user-specified search parameters in conjunction with the received query parameters to perform the search;

performing the search with the received query parameters and subsequently refining search results with the selected ones of the stored associated solicited user-specified search parameters; and

performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search

results with the selected ones of the stored associated solicited user-specified search parameters.

13. (Currently Amended) The process of Claim 12, further comprising the step of:

returning results of the refined search ~~query~~ to the user.

14. (Currently Amended) The process of Claim 12, further comprising the step of:

sending results of the refined search ~~query~~ to a recipient.

15. (Original) The process of Claim 14, wherein the recipient is selected by the user.

16. (Currently Amended) The process of Claim 14, wherein the recipient is determined based on the results of the refined search ~~query~~.

17. (Original) The process of Claim 14, wherein the recipient is determined based on information input by the user.

18. (Currently Amended) The process of Claim 12, further comprising the steps of:

receiving information from a recipient; and

selectively sending results of the refined search ~~query~~ to the recipient based upon any of the received information from the recipient and a selection of the recipient by the user.

19. (Currently Amended) The process of Claim 18, wherein the received information comprises an interest in the results of the refined search ~~query~~ by the recipient.

20. (Original) The process of Claim 12, wherein the subject matter is explicitly indicated by the user.

21. (Original) The process of Claim 12, wherein the subject matter is inferentially determined based upon user input.

22. (Previously Presented) The process of Claim 12, wherein the user-specified search parameters comprise a selectable applicability based upon subject matter, and further comprising the step of:

determining the selected ones of the stored associated solicited user-specified search parameters based upon the selectable applicability.

23. (Currently Amended) A system, comprising:

means for soliciting from a user at least one user-specified search parameter;

means for soliciting from the user one or more rankings ranking explicitly specified by the user, the rankings based on any of applicability and validity of at least one of the solicited user-specified search parameters in regard to at least one subject matter;

means for associating the solicited user-specified search parameters and the explicitly specified rankings with the user;

means for storing the associated solicited user-specified search parameters and the explicitly specified rankings for use in at least one future search initiated by the user;

means for receiving an initial search query ~~received~~ from the user, the search query comprising one or more query parameters other than the previously solicited user-specified search parameters;

means for determining the ~~a determination of~~ subject matter of the received initial search query;

means for selection of one or more of the stored associated solicited user-specified search parameters based on any of applicability and validity of the stored associated solicited search parameters with the determined subject matter of the received search query, wherein the selection is at least partially based on at least one of the solicited user-specified search parameters having a ranking explicitly specified by the user; and

means for ~~integrating the selected stored associated solicited user-specified search parameters with a search based on the received~~ providing a refined search, wherein the refinement comprises any of

means for using the selected ones of the stored associated solicited user-specified search parameters in conjunction with the received query parameters to perform the search;

means for performing the search with the received query parameters and subsequently refining search results with the selected ones of the stored associated solicited user-specified search parameters; and

means for performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the stored associated solicited user-specified search parameters.

24. (Currently Amended) The system of Claim 23, wherein results of the ~~integrated~~ refined search ~~query~~ are sent to the user.

25. (Currently Amended) The system of Claim 23, wherein results of the ~~integrated~~ refined search ~~query~~ are sent to a recipient.

26. (Original) The system of Claim 25, wherein the recipient is selectable by the user.

27. (Currently Amended) The system of Claim 25, wherein the recipient is based on the results of the refined search.

28. (Original) The system of Claim 25, wherein the recipient is based on information input by the user.

29. (Currently Amended) The system of Claim 23, further comprising:

information received from a recipient; and

a selective transmission of the results of the ~~integrated~~ refined search ~~query~~ to the recipient based upon any of the received information and a selection of the recipient by the user.

30. (Currently Amended) The system of Claim 29, wherein the received information comprises an interest in the ~~enhanced~~ results of the refined search.

31. (Original) The system of Claim 23, wherein the determination of the subject matter is explicitly indicated by the user.

32. (Original) The system of Claim 23, wherein the determination of the subject matter is inferentially determined based upon user input.

33. (Previously Presented) The system of Claim 23, wherein the stored associated solicited user-specified search parameters comprise a selectable

applicability based upon subject matter, and wherein the means for selection is based upon the selectable applicability.

34. (Previously Presented) The system of Claim 25, wherein the recipient is any of a selected peer, a friend, a family relative, a student, a neighbor, any of a person or entity within a zip code region, and a member of a network of people associated with the user.

35. (Previously Presented) The system of Claim 34, wherein the network of people comprises any of people with similar interests and a chain of friends.

36. (Previously Presented) The system of Claim 23, further comprising:

means for soliciting the user to provide at least one additional parameter after receipt of the query at the time of the search.

37. (Previously Presented) The system of Claim 36, wherein the solicited additional parameter is a missing parameter.

38. (Previously Presented) The system of Claim 23, further comprising:

means for receiving at least one source selection from the user; and

means for modifying the search query based upon information from the selected source.

39. (Currently Amended) The system of Claim 23, wherein the ~~step of integrating the selected ones of the stored associated solicited user-specified search parameters with the search query comprises any of an inclusion of the selected ones of the stored associated solicited user-specified search parameters with the search query, a refinement of a search result using the selected ones of the stored associated solicited user-specified parameters, and a sorting of a search result using the selected ones of the stored associated solicited~~ means for soliciting the user-specified search parameters corresponds to any of before and after a particularized search.

40. (Previously Presented) The process of Claim 3, wherein the recipient is any of a selected peer, a friend, a family relative, a student, a neighbor, any

of a person or entity within a zip code region, and a member of a network of people associated with the user.

41. (Previously Presented) The process of Claim 40, wherein the network of people comprises any of people with similar interests and a chain of friends.

42. (Previously Presented) The process of Claim 1, further comprising the steps of:

soliciting the user to provide at least one parameter after receipt of the query at the time of the search.

43. (Previously Presented) The process of Claim 42, wherein the solicited parameter is a missing parameter.

44. (Previously Presented) The process of Claim 1, further comprising the steps of:

receiving at least one source selection from the user; and

modifying the search query based upon information from the selected source.

45. (Currently Amended) The process of Claim 1, wherein the step of ~~integrating the selected ones of the associatively stored solicited search parameters with the search query comprises any of an inclusion of the selected ones of the associatively stored solicited search parameters with the search query, a refinement of a search result using the selected ones of the associatively stored solicited search parameters, and a sorting of a search result using the selected ones of the associatively stored solicited~~ soliciting the general search parameters from the user occurs any of before and after a particularized search.

46. (Previously Presented) The process of Claim 14, wherein the recipient is any of a selected peer, a friend, a family relative, a student, a neighbor, any of a person or entity within a zip code region, and a member of a network of people associated with the user.

47. (Previously Presented) The process of Claim 46, wherein the network of people comprises any of people with similar interests and a chain of friends.

48. (Previously Presented) The process of Claim 12, further comprising the steps of:

soliciting the user to provide at least one parameter after receipt of the query at the time of the search.

49. (Previously Presented) The process of Claim 48, wherein the solicited parameter is a missing parameter.

50. (Previously Presented) The process of Claim 12, further comprising the steps of:

receiving at least one source selection from the user; and

modifying the search query based upon information from the selected source.

51. (Currently Amended) The process of Claim 12, wherein the step of ~~integrating the selected ones of the stored associated solicited user specified search parameters with the search query comprises any of an inclusion of the selected ones of the stored associated solicited user specified search parameters with the search query, a refinement of a search result using the selected ones of the stored associated solicited user specified search parameters, and a sorting of a search result using the selected ones of the stored associated solicited user specified search parameters~~ soliciting the user-specified search parameters from the user occurs any of before and after a particularized search.

RECEIVED
CENTRAL FAX CENTER
JAN 2 4 2006

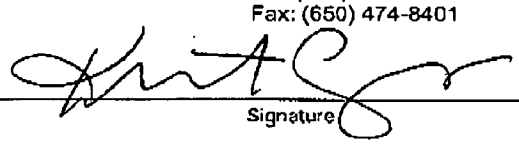
Glenn Patent Group
3475 Edison Way, Ste. L
Menlo Park, CA 94025
Tel. (650) 474-0400
Fax. (650) 474-8401
Customer No. 22,852

Certificate of Transmission under 37 CFR 1.8

I hereby certify that this correspondence is being facsimile transmitted to the United States Patent and Trademark Office

on January 24, 2006
Date

Attn:Examiner Q.N. Nguyen
Fax: (571) 273-8300
From: GLENN PATENT GROUP
Tel: (650) 474-8400
Fax: (650) 474-8401



Signature

Kristianne Serrano

Typed or printed name of person signing Certificate

Note: Each paper must have its own certificate of transmission, or this certificate must identify each submitted paper.

Re: Serial No. 10/685,749
Atty. Docket No.AOL0142
TOTAL PAGES (Incl. cover sheet) - 31
The following documents are enclosed:

- Certificate of Facsimile Transmission (1 page);
- Response (28 pages).
- Request for Continued Examination (1 page, in duplicate)

RECEIVED
CENTRAL FAX CENTER
JAN 24 2006

PTO/SB/30 (09-03)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Request for Continued Examination (RCE) Transmittal

Address to:
Mail Stop RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Application Number	10/685,749
Filing Date	October 14, 2003
First Named Inventor	Fish
Art Unit	2141
Examiner Name	Nguyen, Quang
Attorney Docket Number	AOL0142

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.
Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, or to any design application. See Instruction Sheet for RCEs (not to be submitted to the USPTO) on page 2.

1. Submission required under 37 CFR 1.114 Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).
   - a. ☐ Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.
     - i. ☐ Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
     - ii. ☐ Other _____
   - b. ☐ Enclosed
     - i. ☐ Amendment/Reply
     - ii. ☐ Affidavit(s)/ Declaration(s)
     - iii. ☐ Information Disclosure Statement (IDS)
     - iv. ☐ Other _____
2. Miscellaneous
   - a. ☐ Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)
   - b. ☐ Other _____
3. Fees The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.
   - a. ☑ The Director is hereby authorized to charge the following fees, or credit any overpayments, to Deposit Account No. 07-1445, Glenn Patent Group
     - i. ☑ RCE fee required under 37 CFR 1.17(e) $790 Large Entity
     - ii. ☐ Extension of time fee (37 CFR 1.136 and 1.17)
     - iii. ☐ Other _____
   - b. ☐ Check in the amount of $ _____ enclosed
   - c. ☐ Payment by credit card (Form PTO-2038 enclosed)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED

Name (Print/Type)	Christopher Pei	Registration No. (Attorney/Agent)	45,005
Signature		Date	January 24, 2006

CERTIFICATE OF MAILING OR TRANSMISSION

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date shown below.

Name (Print/Type)	Kristianne Serrano		
Signature		Date	January 24, 2006

COPY

PTO/SB/30 (09-03)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Request for Continued Examination (RCE) Transmittal

Address to:
Mail Stop RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Application Number	10/685,749
Filing Date	October 14, 2003
First Named Inventor	Fish
Art Unit	2141
Examiner Name	Nguyen, Quang
Attorney Docket Number	AOL0142

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.
Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, or to any design application. See Instruction Sheet for RCEs (not to be submitted to the USPTO) on page 2.

1. Submission required under 37 CFR 1.114 Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).

   a. ☐ Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.
      i. ☐ Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
      ii. ☐ Other _____

   b. ☐ Enclosed
      i. ☐ Amendment/Reply
      ii. ☐ Affidavit(s)/ Declaration(s)
      iii. ☐ Information Disclosure Statement (IDS)
      iv. ☐ Other _____

2. Miscellaneous

   a. ☐ Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)
   b. ☐ Other _____

3. Fees The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.

   a. ☑ The Director is hereby authorized to charge the following fees, or credit any overpayments, to Deposit Account No. 07-1445, Glenn Patent Group
      i. ☑ RCE fee required under 37 CFR 1.17(e) $790 Large Entity
      ii. ☐ Extension of time fee (37 CFR 1.136 and 1.17)
      iii. ☐ Other _____

   b. ☐ Check in the amount of $ _____ enclosed
   c. ☐ Payment by credit card (Form PTO-2038 enclosed)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED

Name (Print/Type)	Christopher Pal	Registration No. (Attorney/Agent)	45,005
Signature		Date	January 24, 2006

CERTIFICATE OF MAILING OR TRANSMISSION

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date shown below.

Name (Print/Type)	Kristianne Serrano		
Signature		Date	January 24, 2006

This collection of information is required by 37 CFR 1.114. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop RCE, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

United States Patent and Trademark Office
- Sales Receipt -

02/09/2006 SFORD1 00000001 071445 10685749

01 FC:1202 900.00 DA

RECEIVED
CENTRAL FAX CENTER
JAN 2 4 2006

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Edmund J. Fish
Docket No.: AOL0142

Serial No.: 10/685,749
Art Unit: 2141

Filed: 14 October 2003
Examiner: Nguyen, Quang N.

Title: Search Enhancement System Having Personal Search Parameters

24 January 2006

Commissioner for Patents
Mail Stop RCE
P.O. Box 1450
Alexandria, VA 22313-1450

PRELIMINARY AMENDMENT

Sir:

This is in response to the Office Action of 25 November 2005, for the above-identified patent application. The Examiner made the Action Final. Applicant Requests Continued Examination (RCE) in accordance with 37 CFR 1.114, and authorizes payment in compliance with 37 CFR 1.17(e) of $790, to be charged to the Glenn Patent Group Deposit Account No. 07-1445, Customer No. 22862.

An Examiner Interview was conducted by telephone on 20 January 2006. Donald M. Hendricks, Applicant's representative, and the Examiner discussed the present invention, as disclosed in the Application as filed and as proposed to be claimed in Claims 1, 12 and 23, in contrast to the Cited Art of record. Applicant specifically discussed what are believed to be patentable features of Claim 1, as proposed to be amended. The Examiner stated that the scope of the proposed Claims may require further Search and Examination.

United States Patent and Trademark Office
- Sales Receipt -

02/09/2006 SFORD1 00000002 071445 10685749

01 FC:1801 790.00 DA

RECEIVED
CENTRAL FAX CENTER
JAN 24 2006

PTO/SB/30 (09-03)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Request for Continued Examination (RCE) Transmittal

Address to:
Mail Stop RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Application Number	10/683,749
Filing Date	October 14, 2003
First Named Inventor	Fish
Art Unit	2141
Examiner Name	Nguyen, Quang
Attorney Docket Number	AOL0142

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.
Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, or to any design application. See Instruction Sheet for RCEs (not to be submitted to the USPTO) on page 2.

1. Submission required under 37 CFR 1.114 Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).
   - a. ☐ Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.
     - i. ☐ Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
     - ii. ☐ Other _____
   - b. ☐ Enclosed
     - i. ☐ Amendment/Reply
     - ii. ☐ Affidavit(s)/ Declaration(s)
     - iii. ☐ Information Disclosure Statement (IDS)
     - iv. ☐ Other _____
2. Miscellaneous
   - a. ☐ Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)
   - b. ☐ Other _____
3. Fees The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.
   - a. ☑ The Director is hereby authorized to charge the following fees, or credit any overpayments, to Deposit Account No. 07-1445, Glenn Patent Group
     - i. ☑ RCE fee required under 37 CFR 1.17(e) $790 Large Entity
     - ii. ☐ Extension of time fee (37 CFR 1.136 and 1.17)
     - iii. ☐ Other _____
   - b. ☐ Check in the amount of $ _____ enclosed
   - c. ☐ Payment by credit card (Form PTO-2038 enclosed)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED

Name (Print/Type)	Christopher Pell	Registration No. (Attorney/Agent)	45,005
Signature		Date	January 24, 2006

CERTIFICATE OF MAILING OR TRANSMISSION

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date shown below.

Name (Print/Type)	Kristianne Serrano		
Signature		Date	January 24, 2006

This collection of information is required by 37 CFR 1.114. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop RCE, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

10685749

PATENT APPLICATION FEE DETERMINATION RECORD
Effective October 1, 2003

Application or Docket Number 10685749

CLAIMS AS FILED - PART I

	(Column 1)	(Column 2)
TOTAL CLAIMS	33	
FOR	NUMBER FILED	NUMBER EXTRA
TOTAL CHARGEABLE CLAIMS	33 minus 20=	* 13
INDEPENDENT CLAIMS	3 minus 3 =	* 0
MULTIPLE DEPENDENT CLAIM PRESENT	☐	

* If the difference in column 1 is less than zero, enter "0" in column 2

SMALL ENTITY TYPE ☐	RATE	FEE	OR	OTHER THAN SMALL ENTITY	RATE	FEE
	BASIC FEE	385.00	OR		BASIC FEE	770.00
	XS 9=		OR		XS18=	936
	X43=		OR		X86=	
	+145=		OR		+290=	
	TOTAL		OR		TOTAL	1006

CLAIMS AS AMENDED - PART II

10/13/05

10-13-05

AMENDMENT A

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	33 ~~35~~	Minus	** 33	= 18
Independent	* 3	Minus	*** 3	= 0
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM	☐			

1/12/23/

1/24/06

SMALL ENTITY	RATE	ADDITIONAL FEE	OR	OTHER THAN SMALL ENTITY	RATE	ADDITIONAL FEE
	XS 9=		OR		X$18=	360.00
	X43=		OR		X86=	
	+145=		OR		+290=	
	TOTAL ADDIT. FEE		OR		TOTAL ADDIT. FEE	

AMENDMENT B

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	* 51	Minus	** 33	= 18
Independent	* 3	Minus	*** 3	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM	☐			

1/12/23/

	RATE	ADDITIONAL FEE	OR		RATE	ADDITIONAL FEE
	XS 9=		OR		X$18= 50	900
	X43=		OR		X86=	
	+145=		OR		+290=	
	TOTAL ADDIT. FEE		OR		TOTAL ADDIT. FEE	900

AMENDMENT C

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	*	Minus	**	=
Independent	*	Minus	***	=
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM	☐			

	RATE	ADDITIONAL FEE	OR		RATE	ADDITIONAL FEE
	XS 9=		OR		X$18=	
	X43=		OR		X86=	
	+145=		OR		+290=	
	TOTAL ADDIT. FEE		OR		TOTAL ADDIT. FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20."
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3."
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	2	(fish or edmund) and (search or query) and (applicability and validity) same rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:12
L2	0	"709"/.ccls. and (search or query) and (specif$4 or indicat$3 or select$3 or identif$4) same (applicability and validity) same rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:13
L3	37	"707"/.ccls. and (search or query) and (specif$4 or indicat$3 or select$3 or identif$4) same (applicability or validity) same rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:13
L4	8	"707"/.ccls. and (search or query) and (specif$4 or indicat$3 or select$3 or identif$4) with (applicability or validity) with rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:14
L5	5	"707"/.ccls. and (search or query or parameter) same (specif$4 or indicat$3 or select$3 or identif$4) with (applicability or validity) with rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:15
L6	4	@ad<"20031014" and (search or query or parameter) same (specif$4 or indicat$3 or select$3 or identif$4) with (applicability or validity) with rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:16
L7	4	@ad<"20031014" and (search or query or parameter) same (specif$4 or indicat$3 or select$3 or identif$4 or solicit$3) with (applicability or validity) with rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:16
L8	6	(search or query or parameter) same (specif$4 or indicat$3 or select$3 or identif$4 or solicit$3) with (applicability or validity) with rank$3	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2006/02/10 20:16
S1	1	@ad<"20031014" and "6,282,548".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:02
S2	1	@ad<"20031014" and "5,812,776".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:02
S3	2	@ad<"20031014" and "6,256,639".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:04
S4	1	@ad<"20031014" and "6,311,214".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:05

S5	2	@ad<"20031014" and "6,314,423".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:05
S6	2	@ad<"20031014" and "6,324,566".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:06
S7	1	@ad<"20031014" and "6,405,222".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:08
S8	2	@ad<"20031014" and "6,480,853".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:09
S9	2	@ad<"20031014" and "6,493,702".pn. and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:19
S10	1	@ad<"20031014" and (integrat$3 or incoporat$3) with (stored or saved) with parameter with query and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:20
S11	14	@ad<"20031014" and (integrat$3 or incoporat$3) same (stored or saved) with (term or keyword or parameter) with query and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:21
S12	6	@ad<"20031014" and (integrat$3 or incoporat$3) with (stored or saved) with (term or keyword or parameter) with query and (search$3 or query or parameter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:23
S13	2	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:25
S14	211	@ad<"20031014" and (integrat$3 or incoporat$3 or select$3) with (term or keyword or parameter) with (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:26
S15	9	@ad<"20031014" and (integrat$3 or incoporat$3 or select$3) with (term or keyword or parameter) with (search adj query) and (subject adj matter) with (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:30
S16	70	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 20:31

S17	7	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (query) and (subject adj matter)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:15
S18	15	@ad<"20031014" and (integrat$3 or incoporat$3) with (term or keyword or parameter) with (query) and (search adj query)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:21
S19	62	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or addW$3) with (term or keyword or parameter) with (search adj query) and refin$5 and result and input and (subject or criteria) and user near (personal$4 or information)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:22
S20	69	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and refin$5 and result and input and (subject or criteria) and user near (personal$4 or information)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:22
S21	19	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and refin$5 with result and input and (subject or criteria) and user near (personal$4 or information)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:37
S22	97	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:50
S23	17	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) near result and (subject or criteria or parameter) and user near (personal$7 or information or profile)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:38

S24	10	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near source and (rat$3 or rank$3)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:59
S25	2	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near source and recipient	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:58
S26	10	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near source	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 21:58
S27	10	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near (recipient or source or server or provider) and (rat$3 or rank$3)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 22:01
S28	16	@ad<"20031014" and (integrat$3 or incoporat$3 or includ$3 or add$3) with (term or keyword or parameter) with (search adj query) and (refin$5 or organiz$3 or order$3 or rat$3 or rank$3) with result and (subject or criteria or parameter) and user near (personal$7 or information or profile) and (external or ID) near (recipient or source or server or provider)	US-PGPUB; USPAT; EPO; JPO; DERWENT	OR	ON	2005/07/05 22:01

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

22862 7590 02/23/2006

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

EXAMINER	
NGUYEN, QUANG N	
ART UNIT	PAPER NUMBER
2141	

DATE MAILED: 02/23/2006

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

TITLE OF INVENTION: A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER--

APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	NO	$1400	$300	$1700	05/23/2006

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE REFLECTS A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE APPLIED IN THIS APPLICATION. THE PTOL-85B (OR AN EQUIVALENT) MUST BE RETURNED WITHIN THIS PERIOD EVEN IF NO FEE IS DUE OR THE APPLICATION WILL BE REGARDED AS ABANDONED.

HOW TO REPLY TO THIS NOTICE:

I. Review the SMALL ENTITY status shown above.

If the SMALL ENTITY is shown as YES, verify your current SMALL ENTITY status:

A. If the status is the same, pay the TOTAL FEE(S) DUE shown above.

B. If the status above is to be removed, check box 5b on Part B - Fee(s) Transmittal and pay the PUBLICATION FEE (if required) and twice the amount of the ISSUE FEE shown above, or

If the SMALL ENTITY is shown as NO:

A. Pay TOTAL FEE(S) DUE shown above, or

B. If applicant claimed SMALL ENTITY status before, or is now claiming SMALL ENTITY status, check box 5a on Part B - Fee(s) Transmittal and pay the PUBLICATION FEE (if required) and 1/2 the ISSUE FEE shown above.

II. PART B - FEE(S) TRANSMITTAL should be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). Even if the fee(s) have already been paid, Part B - Fee(s) Transmittal should be completed and returned. If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.

PTOL-85 (Rev. 01/06) Approved for use through 04/30/2007.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

22862 7590 02/23/2006

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission
I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

TITLE OF INVENTION: A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER--

APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	NO	$1400	$300	$1700	05/23/2006

EXAMINER	ART UNIT	CLASS-SUBCLASS
NGUYEN, QUANG N	2141	709-246000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).
☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) the names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.
1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☐ Corporation or other private group entity ☐ Government

4a. The following fee(s) are enclosed:
☐ Issue Fee
☐ Publication Fee (No small entity discount permitted)
☐ Advance Order - # of Copies _____

4b. Payment of Fee(s):
☐ A check in the amount of the fee(s) is enclosed.
☐ Payment by credit card. Form PTO-2038 is attached.
☐ The Director is hereby authorized by charge the required fee(s), or credit any overpayment, to Deposit Account Number _____ (enclose an extra copy of this form).

5. **Change in Entity Status** (from status indicated above)
☐ a. Applicant claims SMALL ENTITY status. See 37 CFR 1.27. ☐ b. Applicant is no longer claiming SMALL ENTITY status. See 37 CFR 1.27(g)(2).

The Director of the USPTO is requested to apply the Issue Fee and Publication Fee (if any) or to re-apply any previously paid issue fee to the application identified above. NOTE: The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

This collection of information is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

22862 7590 02/23/2006

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

EXAMINER	
NGUYEN, QUANG N	
ART UNIT	PAPER NUMBER
2141	

DATE MAILED: 02/23/2006

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment to date is 211 day(s). If the issue fee is paid on the date that is three months after the mailing date of this notice and the patent issues on the Tuesday before the date that is 28 weeks (six and a half months) after the mailing date of this notice, the Patent Term Adjustment will be 211 day(s).

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

PTOL-85 (Rev. 01/06) Approved for use through 04/30/2007.

Notice of Allowability	Application No.	Applicant(s)	
	10/685,749	FISH, EDMUND J.	
	Examiner	Art Unit	
	Quang N. Nguyen	2141	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to *the Amendment filed on 01/24/2006*.

2. ☒ The allowed claim(s) is/are *1-51*.

3. ☒ The drawings filed on *13 October 2005* are accepted by the Examiner.

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
   a) ☐ All b) ☐ Some* c) ☐ None of the:
      1. ☐ Certified copies of the priority documents have been received.
      2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
      3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.

6. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
      1) ☐ hereto or 2) ☐ to Paper No./Mail Date _____.
   (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

7. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☐ Notice of Draftperson's Patent Drawing Review (PTO-948)
3. ☐ Information Disclosure Statements (PTO-1449 or PTO/SB/08), Paper No./Mail Date _____
4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
5. ☐ Notice of Informal Patent Application (PTO-152)
6. ☐ Interview Summary (PTO-413), Paper No./Mail Date _____ .
7. ☐ Examiner's Amendment/Comment
8. ☒ Examiner's Statement of Reasons for Allowance
9. ☐ Other _____.

Allowable Subject Matter

1. Claims 1-51 are allowable.

2. The following is an examiner's statement of reasons for allowance:

In interpreting the amended claims, in light of the specification and the applicant's arguments filed on 01/24/2006, the Examiner finds the claimed invention to be patentably distinct from the prior art of record.

Harris (US 2002/0059204 A1) teaches a system and method for providing distributing a search query to devices on a communications network, such as servers on the Internet, using an application that can survey a subscriber's server and use the formatting information and data from the survey to create at least one dictionary customized to the subscriber's data sources, wherein the customized dictionaries can customize the search query based on respective subscriber databases and text documents, text search queries, etc., to produce an customized result. The search query results can be filtered and integrated for presentation to the initiating client/device. Search results can be customized using user preference or profile information **(Harris, Abstract).**

Adar et al. (US 6,493,702), teach a method and system for searching and recommending documents in a collection using shared bookmarks employing the preferences and profiles of individual users and groups within a community of users, as well as information derived from shared document bookmarks, to augment Internet searches, re-rank search results, and provide recommendations for documents based on a subject-matter query **(Adar, Abstract).**

The prior art of record teaches the claimed invention substantially, but it fails to teach or suggest individually or in combination that a search enhancement system and method comprising the steps of: soliciting from a user a set of one or more general search parameters specified by the user; ***soliciting from the user at least one ranking explicitly specified by the user, the rankings based on any of applicability and validity of at least one of the solicited general search parameters in regard to at least one subject matter;*** associatively storing the set of solicited general search parameters and the rankings specified by the user for at least one future search initiated by the user; receiving a search query from the user, the search query comprising one or more query parameters other than the previously solicited general search parameters; determining the subject matter of the received search query; ***selecting one or more of the associatively stored solicited general search parameters based upon any of applicability and validity of the associatively stored solicited search parameters with the determined subject matter of the received search, wherein the selection is at least partially based on at least one of the rankings explicitly specified by the user;*** and providing a refined search, wherein the refinement comprises any of using

the selected ones of the associatively stored solicited general search parameters in conjunction with the received query parameters to perform the search; performing the search with the received query parameters and subsequently refining search results with the selected ones of the associatively stored solicited general search parameters; and performing the search with the received query parameters and subsequently providing any of organizing and sorting of the search results with the selected ones of the associatively stored solicited general search parameters as set forth in independent claims 1, 12 and 23. Claims 1-51 are allowed because of the combination of other limitations and the limitation listed above.

3. Pursuant to MPEP 606.01, the title has been changed to read:

-- A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER --

4. Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

5. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Quang N. Nguyen whose telephone number is (571) 272-3886.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Rupal Dharia can be reached on (571) 272-3880. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

RUPAL DHARIA
SUPERVISORY PATENT EXAMINER

Attorney Docket No. AOL0142

Form 1449 (Modified)	Atty. Docket No.: AOL0142	Serial No.: 10/685,749
Information Disclosure	Applicant: Fish	Group: 2155
Statement By Applicant		
	Filing Date: 10/14/2003	Examiner: Unassigned

U.S. Patent Documents

Examiner Initial	No.	Patent No.	Date	Patentee	Class	Sub-class	Filing Date
QN	A	5,642,502	6/24/97	Driscoll	395	606	12/6/94
	B	5,848,396	12/08/98	Gerace	705	10	4/26/96
	C	6,029,195	2/22/00	Herz	709	219	12/5/97
	D	6,314,420	11/06/01	Lang et al.	707	3	12/3/98
	E	6,363,377	3/26/02	Kravets et al.	707	4	12/22/98
	F	6,397,212	5/28/02	Biffar	707	5	3/4/99
	G	6, 539,377	3/25/03	Culliss	707	5	10/6/00
	H	5,983,214	11/09/99	Lang et al.	707	1	11/5/98
	I	6,029,161	2/22/00	Lang et al.	707	1	11/19/98
	J	6,047,327	4/04/00	Tso et al.	709	232	2/16/96
	K	6,182,068	1/30/01	Culliss	707	5	3/1/99
	L	6,308,175	10/23/01	Lang et al.	707	10	11/19/98
	M	6,349,307	2/19/02	Chen	707	103	12/28/98
	N	6,360,215	3/19/02	Judd et al.	707	3	11/3/98
	O	6,377,945	4/23/02	Risvik	707	3	7/9/99
	P	6,404,884	6/11/02	Marwell et al.	379	265.13	11/24/99
	Q	6/453,315	9/17/02	Weissman et al.	707	5	11/1/99
	R	6,466,970	10/15/02	Lee et al.	709	217	1/27/99
	S	6,484,162	11/19/02	Edlund et al.	707	3	6/29//99
	T	6,498,795	12/24/02	Zhang et al.	370	400	5/28/99
	U	6,505,194	1/07/03	Nicolovska et al.	707	3	3/29/00
QN	V	6,550,057	4/15/03	Bowman-Amuah	717	126	8/31/99

Foreign Patent or Published Foreign Patent Application

Examiner Initial	No.	Document No.	Publication Date	Country or Patent Office	Class	Sub-class	Translation	
							Yes	No
QN	W	1 072 982	1/31/01	EPO	G06F	17/30	X	
QN	X	1 095 326	1/30/02	EPO	G06F	1/00	X	
QN	Y	1 284 461	2/19/03	EPO	G06F	17/30	X	
QN	Z	1 288 795	3/05/03	EPO	G06F	17/30	X	

Publications and Other Documents

Examiner Initial	No.	Author, Title, Date, Place (e.g. Journal) of Publication
QN	AA	PITKOW, J. et al.; Personalized Search: Communications of the ACM, vol. 45, no. 9, p. 50-5; September 2002
QN	AB	MCGOWAN, J.P. et al.; Who Do You Want To Be Today? Web Personae for Personalised Information Access; Adaptive Hypermedia and Adaptive Web-Based Systems. Second International Conference, AH 2002. Proceedings (Lecture Notes in Computer Science Vol. 2347), p. 514-17; 2002
QN	AC	KALAJDZISKI, S. et al.; Intelligent Recommendation in Digital Library; Proceedings of the IASTED International Conference Intelligent Systems and Control, p. 408-12; ACTA Press, Anaheim, CA, USA; 2001
QN	AD	KERSCHBERG, L. et al.; A Semantic Taxonomy-Based Personalizable Meta-Search Agent; Proceedings of the Second International Conference on Web Information Systems Engineering, vol. 1, p. 41-50; IEEE Comput. Soc., Los Alamitos, CA, USA; 2002

Issue Classification	Application/Control No.	Applicant(s)/Patent under Reexamination	
	10/685,749	FISH, EDMUND J.	
	Examiner	Art Unit	
	Quang N. Nguyen	2141	

ISSUE CLASSIFICATION

ORIGINAL			CROSS REFERENCE(S)						
CLASS	SUBCLASS	CLASS	SUBCLASS (ONE SUBCLASS PER BLOCK)						
709	246	709	203	219	228				
INTERNATIONAL CLASSIFICATION									
G 0 6 F	15/16								
	/								
	/								
	/								
	/								

Quang N. Nguyen 02/10/06
(Assistant Examiner) (Date)

Sharon Blymond 2/21/06
(Legal Instruments Examiner) (Date)

2/17/06
RUPAL DHARIA
SUPERVISORY PATENT EXAMINER
(Primary Examiner) (Date)

Total Claims Allowed: 51

O.G. Print Claim(s)	O.G. Print Fig.
1	14

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
1	1	43	31		61		91		121		151		181
2	2	44	32		62		92		122		152		182
3	3	45	33		63		93		123		153		183
4	4	46	34		64		94		124		154		184
5	5	47	35		65		95		125		155		185
6	6	48	36		66		96		126		156		186
7	7	49	37		67		97		127		157		187
8	8	50	38		68		98		128		158		188
9	9	51	39		69		99		129		159		189
10	10	12	40		70		100		130		160		190
11	11	13	41		71		101		131		161		191
18	12	14	42		72		102		132		162		192
19	13	15	43		73		103		133		163		193
20	14	16	44		74		104		134		164		194
21	15	17	45		75		105		135		165		195
22	16	29	46		76		106		136		166		196
23	17	30	47		77		107		137		167		197
24	18	31	48		78		108		138		168		198
25	19	32	49		79		109		139		169		199
26	20	33	50		80		110		140		170		200
27	21	34	51		81		111		141		171		201
28	22		52		82		112		142		172		202
35	23		53		83		113		143		173		203
36	24		54		84		114		144		174		204
37	25		55		85		115		145		175		205
38	26		56		86		116		146		176		206
39	27		57		87		117		147		177		207
40	28		58		88		118		148		178		208
41	29		59		89		119		149		179		209
42	30		60		90		120		150		180		210

Index of Claims

Application/Control No.	Applicant(s)/Patent under Reexamination
10/685,749	FISH, EDMUND J.
Examiner	**Art Unit**
Quang N. Nguyen	2141

√	Rejected	–	(Through numeral) Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

Claim		Date								
Final	Original	7/7/05	11/20/05	2/10/06						
	1	√	√	=						
	2	√	√	=						
	3	√	√	=						
	4	√	√	=						
	5	√	√	=						
	6	√	√	=						
	7	√	√	=						
	8	√	√	=						
	9	√	√	=						
	10	√	√	=						
	11	√	√	=						
	12	√	√	=						
	13	√	√	=						
	14	√	√	=						
	15	√	√	=						
	16	√	√	=						
	17	√	√	=						
	18	√	√	=						
	19	√	√	=						
	20	√	√	=						
	21	√	√	=						
	22	√	√	=						
	23	√	√	=						
	24	√	√	=						
	25	√	√	=						
	26	√	√	=						
	27	√	√	=						
	28	√	√	=						
	29	√	√	=						
	30	√	√	=						
	31	√	√	=						
	32	√	√	=						
	33	√	√	=						
	34		√	=						
	35		√	=						
	36		√	=						
	37		√	=						
	38		√	=						
	39		√	=						
	40		√	=						
	41		√	=						
	42		√	=						
	43		√	=						
	44		√	=						
	45		√	=						
	46		√	=						
	47		√	=						
	48		√	=						
	49		√	=						
	50		√	=						

Claim		Date								
Final	Original	7/7/05	11/20/05	2/10/06						
	51		√	=						
	52									
	53									
	54									
	55									
	56									
	57									
	58									
	59									
	60									
	61									
	62									
	63									
	64									
	65									
	66									
	67									
	68									
	69									
	70									
	71									
	72									
	73									
	74									
	75									
	76									
	77									
	78									
	79									
	80									
	81									
	82									
	83									
	84									
	85									
	86									
	87									
	88									
	89									
	90									
	91									
	92									
	93									
	94									
	95									
	96									
	97									
	98									
	99									
	100									

Claim		Date								
Final	Original									
	101									
	102									
	103									
	104									
	105									
	106									
	107									
	108									
	109									
	110									
	111									
	112									
	113									
	114									
	115									
	116									
	117									
	118									
	119									
	120									
	121									
	122									
	123									
	124									
	125									
	126									
	127									
	128									
	129									
	130									
	131									
	132									
	133									
	134									
	135									
	136									
	137									
	138									
	139									
	140									
	141									
	142									
	143									
	144									
	145									
	146									
	147									
	148									
	149									
	150									

Search Notes

Application/Control No.	Applicant(s)/Patent under Reexamination	
10/685,749	FISH, EDMUND J.	
Examiner	**Art Unit**	
Quang N. Nguyen	2141	

SEARCHED			
Class	Subclass	Date	Examiner
709	203	7/5/2005	QN
709	218	7/5/2005	QN
709	219	7/5/2005	QN
707	225	7/5/2005	QN
707	3	7/5/2005	QN

INTERFERENCE SEARCHED			
Class	Subclass	Date	Examiner
709	**203**	**2/10/2006**	**QN**
709	219	2/10/2006	QN
709	228	2/10/2006	QN
709/246		2/10/2006	QN

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	DATE	EXMR
East Search (see notes)	7/5/2005	QN
Inventor Name Search (See Notes)	2/10/2006	QN
East Search (See Notes)	2/10/2006	QN

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Fish
Serial No. : 10/685,749
Filed: October 14, 2003
Title: Search Enhancement System Having Personal Search Parameters

Docket No.: AOL0142
Art Unit: 2141
Examiner: Quang N. Nguyen

April 24, 2006

Assistant Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Sir:

This Information Disclosure Statement is submitted:

( ) under 37 CFR 1.97(b), or
(within three months of filing national application; or date of entry of international application; or before mailing date of first office action on the merits; whichever occurs last)

( ) under 37 CFR 1.97(c) together with either a:
- ( ) **Certification** under 37 CFR 1.97(e), or
- ( ) a $220.00 fee under 37 CFR 1.17(p), or

(After the CFR 1.97(b) time period, but before final action or notice of allowance, whichever occurs first)

(x) under 37 CFR 1.97(d) together with a:
- ( ) Certification under 37 CFR 1.97(e), and
- (x) a $180.00 fee under 37 CFR 1.17(p)), and
- (x) a $130.00 petition fee set forth in 37 CFR 1.17(i)(1.97)(d)

(Filed after final action or notice of allowance, whichever occurs first, but before payment of the issue fee)

(X) The Commissioner is authorized to charge any additional fees or credit any overpayment to Deposit Account No. 07-1445 (Order No. AOL0142). A copy of this sheet is enclosed for accounting purposes.

[redacted]

(X) Applicant(s) submit herewith Form PTO SB/08 -- Information Disclosure Citation together with copies of patents, publications or other information of which applicant(s) are aware, which applicant(s) believe(s) may be material to the examination of this application and for which there may be a duty to disclose in accordance with 37 CFR 1.25.

( ) A concise explanation of the relevance of foreign language patents, foreign language publications and other foreign language information listed on PTO Form 1449, as presently understood by the individual(s) designated in 37 CFR 156(c) most knowledgeable about the content is given on the attached sheet, or where a foreign language patent is cited in a search report or other action by a foreign patent office in a counterpart foreign application, an English language version of the search report or action which indicates the degree of relevance found by the foreign office is listed on form PTO 1449 and is enclosed herewith.

It is requested that the information disclosed herein be made of record in this application.

Respectfully Submitted,

Michael Glenn
Reg. No. 30,176

Customer No. 22862

[redacted]

COPY

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Fish
Serial No. : 10/685,749
Filed: October 14, 2003
Title: Search Enhancement System Having Personal Search Parameters

Docket No.: AOL0142
Art Unit: 2141
Examiner: Quang N. Nguyen

April 24, 2006

Assistant Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Sir:

This Information Disclosure Statement is submitted:

( ) under 37 CFR 1.97(b), or
(within three months of filing national application; or date of entry of international application; or before mailing date of first office action on the merits; whichever occurs last)

( ) under 37 CFR 1.97(c) together with either a:
- ( ) **Certification** under 37 CFR 1.97(e), or
- ( ) a $220.00 fee under 37 CFR 1.17(p), or

(After the CFR 1.97(b) time period, but before final action or notice of allowance, whichever occurs first)

(x) under 37 CFR 1.97(d) together with a:
- ( ) Certification under 37 CFR 1.97(e), and
- (x) a $180.00 fee under 37 CFR 1.17(p)), and
- (x) a $130.00 petition fee set forth in 37 CFR 1.17(i)(1.97)(d)

(Filed after final action or notice of allowance, whichever occurs first, but before payment of the issue fee)

(X) The Commissioner is authorized to charge any additional fees or credit any overpayment to Deposit Account No. 07-1445 (Order No. AOL0142). A copy of this sheet is enclosed for accounting purposes.

[redacted]

COPY

(X) Applicant(s) submit herewith Form PTO SB/08 -- Information Disclosure Citation together with copies of patents, publications or other information of which applicant(s) are aware, which applicant(s) believe(s) may be material to the examination of this application and for which there may be a duty to disclose in accordance with 37 CFR 1.25.

( ) A concise explanation of the relevance of foreign language patents, foreign language publications and other foreign language information listed on PTO Form 1449, as presently understood by the individual(s) designated in 37 CFR 1.56(c) most knowledgeable about the content is given on the attached sheet, or where a foreign language patent is cited in a search report or other action by a foreign patent office in a counterpart foreign application, an English language version of the search report or action which indicates the degree of relevance found by the foreign office is listed on form PTO 1449 and is enclosed herewith.

It is requested that the information disclosed herein be made of record in this application.

Respectfully Submitted,

Michael Glenn
Reg. No. 30,176

Customer No. 22862

[redacted]

Form 1449 (Modified) **Information Disclosure Statement By Applicant** (Use Several Sheets if Necessary)	Atty. Docket No. AOL0142 Applicant: Fish Filing Date: October 14, 2003	Serial No.: 10/685,749 Group: 2141

U.S. Patent Documents

Examiner Initial	No.	Patent No.	Date	Patentee	Class	Sub-class	Filing Date
	A	5,862,325	1/19/1999	Reed, et al	395	200.31	9/27/1996
	B	5.974 412	10/26/1999	Hazlehurst, et al	707	3	9/24/1997
	C	6,018,738	1/25/2000	Breese, et al	700	100	1/22/1998
	D	6,144,964	11/07/2000	Breese, et al	707	10	1/22/1998
	E	6,256,664B1	7/03/2001	Donoho, et al	709	204	3/19/1999
	F	6,263,362B1	7/17/2001	Donoho, et al	709	207	7/09/1999
	G	6.289,353B1	9/11/2001	Hazlehurst, et al	707	102	6/10/1999
	H	6,314,420B1	11/06/2001	Lang, et al	707	3	12/03/1998
	I	6,345,264B1	2/05/2002	Breese, et al	706	21	1/22/1998
	J	6,345,288B1	2/05/2002	Reed, et al	709	201	5/15/2000
	K	6,353,813B1	3/05/2002	Breese, et al	706	12	1/22/1998
	L	6,356,936B1	3/12/2002	Donoho, et al	709	206	5/20/1999
	M	6,493,637B1	12/10/2002	Steeg	702	19	9/24/1999
	N	6,510,458B1	1/21/2003	Berstis, et al	709	219	7/15/1999
	O	6,539,392B1	3/25/2003	Rebane	707	101	3/29/2000

Foreign Patent or Published Foreign Patent Application

Examiner Initial	No.	Document No.	Publication Date	Country or Patent Office	Class	Sub-class	Translation	
							Yes	No

Other Documents

Examiner Initial	No.	Author, Title, Date, Place (e.g. Journal) of Publication
	P	FREUND, Y. et al.; *An Efficient Boosting Algorithm for Combining Preferences*; AT&T Labs, MIT Laboratory for Computer Science;
	Q	SHAVLIK, J. et al.; *Building Intelligent Agents for Web-Based Tasks: A Theory-Refinement Approach*; University of Wisconsin-Madison;
	R	SHAVLIK, J. et al.; *Intelligent Agents for Web-based Tasks: An Advice-Taking Approach*; University of Wisconsin-Madison;

[redacted]

	S	BAUDISCH, P.; *The Profile Editor: Designing a Direct Manipulative Tool for Assembling Profiles*; Institute for Integrated Information and Publication Systems IPSI, German National Research Center for Information Technology GMD, Germany;
	T	BUDZIK, J. et al.; *User Interactions with Everyday Applications as Context for Just-in-time Information Access*; Intelligent Information Laboratory, Northwestern University;
	U	BUDZIK, J. et al.; *Watson: Anticipating and Contextualizing Information Needs*; Northwestern University;
	V	GLOVER, E.J. et al.; *Improving Category Specific Web Search by Learning Query Modifications*; NEC Research Institute, Princeton, NJ, EECS Department, University of Michigan, Ann Arbor, MI, Information Sciences and Technology, Pennsylvania State University;
	W	PAZZANI, M.J. et al.; *A Framework for Collaborative, Content-Based and Demographic Filtering*; Department of Information and Computer Science, University of California, Irvine;
	X	BAUER, T. et al.; *Real Time User Context Modeling for Information Retrieval Agents*; Computer Science Department, Indiana University;
	Y	SHAVLIK, J. et al.; *An Instructable, Adaptive Interface for Discovery and Monitoring Information on the World-Wide Web*; University of Wisconsin-Madison;
	Z	BUDZIK, J. et al.; *Watson: An Infrastructure for Providing Task-Relevant, Just-In-Time Information*; Department of Computer Science, Northwestern University;
	AA	NAHL, D.; *Ethnography Of Novices' First Use Of Web Search Engines: Affective Control In Cognitive Processing*; Internet Reference Services Quarterly, vol. 3, no. 2, p. 51-72; 1998;

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[redacted]

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 288 795 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
05.03.2003 Bulletin 2003/10

(51) Int Cl.[7]: **G06F 17/30**

(21) Application number: **01307210.3**

(22) Date of filing: **24.08.2001**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **BRITISH TELECOMMUNICATIONS public limited company**
London EC1A 7AJ (GB)

(72) Inventor: **The designation of the inventor has not yet been filed**

(74) Representative: **Maggs, Michael Norman et al**
Kilburn & Strode
20 Red Lion Street
London WC1R 4PJ (GB)

(54) Query systems

(57) A computer query system and method enhances the results produced by standard queries, for example in search engines, by filtering and optionally ordering the results on the basis of affinity with individuals known to the search initiator, or individuals within the search initiator's Personal Social Network. The system allows users searching databases for information to evaluate the results of a query with reference to feedback that has previously been provided by others within the user's social network. This allows users to make informed decisions based on how much they value the opinions of others within their personal network.

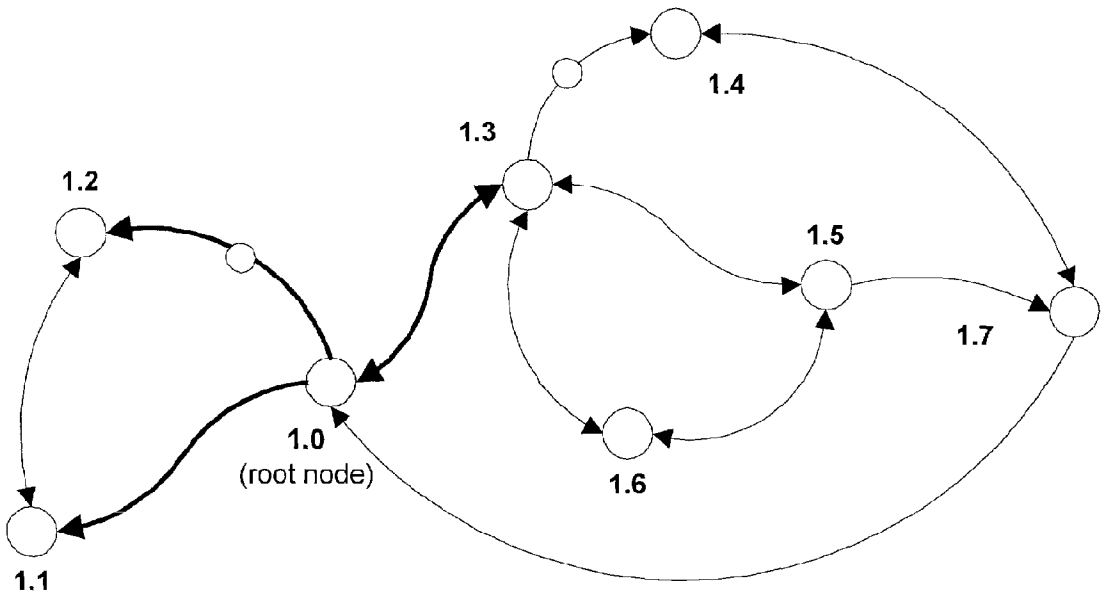


Figure 1

EP 1 288 795 A1

Description

[0001] The present invention relates to computer based queries systems and has applications in fields such as search engines, instant messaging services and knowledge management systems.

[0002] Keyword based query systems, such as for example those used by internet search engines, have been available for many years. Because of the large number of results or "hits" typically returned by such search engines, there has been increasing interest in finding ways of filtering and/or prioritising the results so that they can be presented to the user or "search initiator" in a meaningful form. Typically, weighting factors may be applied to the returned results in some way, so that the more interesting more relevant results appear towards the top of the returned list. One way of doing this, used by some internet search engines, is to assign "popularity scores" to URLs based upon the number of times the URL has been accessed by the search engine website.

[0003] Another approach to prioritisation is to use Collaborative Based Filtering (CBF) scheme. These systems find relevant documents by looking at ratings and recommendations (which may be explicit or implicit) made by others. Examples of Collaborative Based Filtering schemes include the following:

Information Forwarding Systems (IFS):

[0004] These systems allow people to forward information they find interesting to groups of people with a similar interest. These systems are sometimes called Active Collaborative Filtering systems because, here, users explicitly select the group of people to which they recommend an item. The system's role is simply to forward the material to a selected group.

Recommender Systems (RS):

[0005] These systems find items (documents, movies, products) relevant to a particular user by analysing the preferences of other users with similar profiles. These systems are sometimes called Passive Collaborative Filtering Systems since users do not recommend items to any one in particular. Instead, the system determines what items should be recommended to whom, based on similarities between user preferences. A licensable example of this technology can be found at NetPerceptions (www.netperceptions.com). This solution is used by e.g. online bookstores and provides a recommended reading list following a user's purchasing profile, weighed against other users who have purchased similar books.

"Top N" Systems (TNS):

[0006] These systems allow a user to navigate through a list of the most popular documents in a given collection. These systems differ from Recommender Systems insofar as their recommendation is not based on a priori knowledge about a user's particular interest, but only on the overall popularity of the documents themselves.

[0007] Other approaches allow authors to rate their own websites for others to use and to prioritise or filter based upon those ratings. Such system are used by service providers who wish to provide "safe" content, for example www.safesurf.com.

[0008] According to a first aspect of the present invention there is provided a computer query method comprising receiving a query from a search initiator, carrying out a search for information on the basis of the query and filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

[0009] According to a second aspect of the invention there is provided a computer query system comprising means for receiving a query from a search initiator, means for carrying out a search for information on the basis of the query, and means for filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

[0010] The invention further extends to a computer program for carrying out the above method and to a data carrier which carries such a computer program.

[0011] The method of the invention will be referred to within this patent application by the term "Affinity based queries" (ABQ). ABQ allows users searching databases, or the internet, an ability to evaluate the results returned by virtue of inter-personal relationships in the context of their social network. The user can, in that way, be presented preferentially with results which have been recommended by people known to him or her, or otherwise with in his or her social network.

[0012] In the preferred embodiment, results are categorised not only by the degree with which they fit the search criteria (this is currently done), but they are further categorised with metadata describing the relationship between the search initiator and an individual within the network who has refereed, rated or is otherwise acquainted with one of the results. This allows users to make informed decisions based on the extent to which they value the opinion, or the prior

actions, of other members of their social network.

[0013] The invention is preferably implemented through a combination of relational databases, instant messaging (or some other community tool holding information on the social network), along with other query software to implement the ABQ rules and to drive the process of traversing through the social network. This is preferably implemented through nested SQL queries and graphing techniques.

[0014] In a preferred embodiment, an instant messaging application or an internet search engine forms the key service entry point. A relational database backend holds information on the social network, and uses the internal relationships within the network to carry out filtering and/or sorting on the results of a standard search (a keyword-based search). Whenever information is publicly exchanged between individuals within the network (either explicitly or automatically/implicitly) the system preferably intercepts the exchange and records entries in the data base defining metadata links which indicate:

1. Where the information is,
2. Summaries of that information,
3. Keywords about the information,
4. Any ratings that may have been applied (e.g. to an article product or individual),
5. Details of the individual who referred it,
6. Details of the individual to whom it was referred,
7. Details of the time and date when it was referred.

[0015] Then, whenever a user initiates an ABQ, the system preferably uses that person's entry in the database as the root node in the search graph and begins to percolate out through the social network conducting nested queries and table joins between persons and content, to a predefined distance or order, to retrieve the information.

[0016] In order to encourage involvement of the individuals within the network, reward feedback may be provided to the individual who recommended to the user the search result that the user decides is "best". For example, if a user is looking for an article on a particular subject, and selects an article on the internet that has been recommended by an individual X, X receives a reward in the form of an enhancement to his or her "action value". Individuals with high action values will have made good recommendations many times and other users will know that they can be relied upon. Preferably, the system allows the user the opportunity to sort the search results by the action values of the referring individuals thereby insuring that the recommendations from trusted individuals appear at the top of the list.

[0017] The invention may be carried into practice in a number of ways and one specific embodiment will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 shows an exemplary Personal Social Network (PSN):
Figure 2 illustrates schematically the main apparatus elements of an embodiment of the present invention:
Figure 3 illustrates the linking of the PERSON tables:
Figure 4 illustrates the action tables:
Figure 5 shows the method of registering and logging in referees in the database:
Figure 6 shows the method of modifying PERSON records:
Figure 7 shows the method of deleting PERSON records:
Figure 8 shows the method of storing relations:
Figure 9 shows the method of deleting relations:
Figure 10 shows the method of summarising and storing actions:
Figure 11 shows a method of processing incoming query parameters:
Figure 12 shows the method of building an in-memory view:
Figure 13 shows a digraph method of building in-memory views:
Figure 14 shows a digraph and binary tree method of building in-memory views, avoiding duplicates:
Figure 15 shows the method of filtering key word matches:
Figure 16 shows the method of ordering result items by rating:
Figure 17 shows how edge weightings may vary throughout a Personal Social Network:
Figure 18 shows the ordering of the in-memory view by relation:
Figure 19 shows the ordering of the results by relation:
Figure 20 shows the method of encoding and presenting the result list; and
Figure 21 shows the method of feeding back result actions.

1 Introduction

[0018] The preferred embodiment of the invention uses a system called Affinity Based Queries (ABQs) to improve

the relevance of search results to the person requesting them (known as the "search initiator"). In very general terms, the system works as follows:

1. The search initiator first does a standard keyword search e.g. to search for a particular book (a bookshop website).

2. The search returns a list of results or "hits" which are then filtered on the basis of information derived from a Personal Social Network (PSN) of which the search initiator is a member. Personal Social Networks will be described in more detail in section 2. The filter selects from the results list hits which are associated in some predefined way with others in the PSN, for example individuals who have rated the book. The other hits are discarded. This produces a list of filtered result items.

3. The result items are then ordered in some user-defined order for presentation to the search initiator. they could be ordered, for example, by the rating which other individuals have provided, or by the relationship between the search initiator and the other individuals (e.g. ratings by close friends take precedence), or they could be ordered by result "action" (e.g. putting at the top of the list recommendations by an individual whose previous recommendations have been good).

4. Finally, the system presents the sorted and filtered list of results to the search initiator along with an indication, in each case, of which individual within the PSN has recommended the book, and how that individual is related within the PSN to the search initiator.

[0019] It will be appreciated of course that Affinity Based Queries can be used for all types of information retrieval. We will refer to the information to be retrieved as a "document", it being understood that that term includes not only traditional documents in electronic form, but also internet web pages and, more generally, all types of searchable information whether held with in a database or otherwise. When we say, for example, that the search initiator is looking for "books" what we really mean is that he or she is searching one or more databases, typically held on the internet, which contain not books themselves but information about books.

[0020] A more formal description of the preferred embodiment will now be set out. The reader is referred to the glossary of terms, at the end of the specific description, which provides an explanation of certain expressions.

[0021] When an Affinity Based Query (ABQ) is run, it returns a result list which consists of a number of result items. Each result item has some association with the "referee" who is somebody within the search initiator's PSN who is either directly associated with the result item, or who is indirectly associated with the item through being related to someone else within the search initiator's PSN. Along with the result, each result item may also indicate a "relation" (showing how the referee is related to the search initiator through the PSN), along with a "rating" (an award which has been applied by a referee to a result).

[0022] Each result item in an ABQ result list is presented in the form:
<referee><relation><result><rating>.

[0023] Order is not important, nor is the frequency of in which terms are used.

[0024] In the above expression, the following terms are used:

- **referee** - mandatory term, a person associated with a result;
- **relation** - optional term, relating a referee to the search initiator through the PSN;
- **result** - mandatory term, the location (e.g. URL) of something returned from a query;
- **rating** - optional term, an award applied by a referee to a result.

[0025] The result of an ABQ is a result list in which keyword matches are "annotated" or "embellished" by additional information specifying not only what other referees thought of the result but how those referees are associated with or linked to the search initiator.

[0026] Taking the example of an on line bookstore, a result item might read:
*e.g. "Kate (**referee**), who knows Nick (**relation**), purchased 'Wuthering Heights' by Emily Bronte (**result**)"*

[0027] Or, consider a search engine. Rather than just keyword-match a URL query, ABQ's return results referencing other persons known by the search initiator who have visited/rated the URL:
e.g. "Dan (referee), recommended http://www.domain.com/ (result), awarding it 4/5 (rating)"

[0028] Or, consider skills searching in an organisation. Rather than just return persons possibly unknown to the user, ABQ's return results in relation to someone known by the search initiator:
e.g. "Andy (referee), who knows Nick (relation), knows Oracle (result), and rates himself as EXPERT (rating)"

[0029] Or finally, consider an article rating:

e.g. "Jim (referee), who Andy knows (relation), who knows Nick (relation), rated article X (result), as FAIR (rating)"

[0030] Of course, in the above examples the text in brackets would not actually be seen by the user: it is included merely to show how the expressions are made up from the basic terms which include referee, relation, result and rating.

[0031] By decorating standard database query results with respect to someone known by the search initiator, ABQ's allow users to make more informed decisions about which result to follow first, and a reference point to qualify that order with (i.e. the referee).

[0032] As mentioned above, the result list may be ordered in a variety of different ways, according to the needs of the search initiator. That will be described in more detail below.

2. Personal Social Networks (PSN)

[0033] Personal Social Networks are, per se, a known concept. Such a network connects an individual to a number of other individuals with in a community (real or virtual). A PSN can be described by means of a directed graph (a "digraph"). An example is shown in figure 1. Each node in the digraph represents an individual, and the directed edges between the nodes represent relationships. Relationships are
"reflexive" in nature and do not necessarily imply reciprocity: thus, a may know b, but b may not know of a.

[0034] In figure 1, the first order PSN represents those nodes that can be reached in one "hop" from the root node (in this case the search initiator), the second order network those nodes that can be reached in two "hops", and so on.

[0035] In addition to the nodes, representing a community of individuals, the PSN of this embodiment also stores information on "actions". These will be discussed in more detail in section 3, below, but, essentially, represent things done by individual nodes, which are relevant to the particular document being sought. Examples include visit URL, read article, purchase products, volunteer skill and so on. Information on actions is stored against each node of the PSN.

[0036] A search initiator may have a very large available PSN, in theory, out to order N. However, not all of that PSN may be useful for the search being carried out, and the search initiator is provided with means (such as a data entry field or a series of tick boxes) to restrict the search to a PSN of a user-selectable order. Typically the user may wish to restrict the search to a first, second or third order PSN.

[0037] A further option may be made available allowing the search initiator to exclude particular individuals or classes of individuals from that part of the PSN which is going to be used for the search. For example, the user may wish to exclude work-based contacts for a search which is being carried out in respect of a personal hobby.

[0038] Once the search initiator has specified the appropriate parameters, the system works through that individual's PSN, link by link, to construct the subset of the PSN that is going to be used in the search. This subset is then stored in memory as an "in-memory view". Storing information in memory ensures that that part of the PSN that is going to be used for searching does not need to be recreated if the user wishes to carry out additional searches using the same criteria.

[0039] Further details of the way in which the digraph of figure 1 is built up from a relational database, and traversed, will be described later.

[0040] The grouping of individuals making up the PSN, and the type of relations between them, is not of importance to the operation of the invention as a whole, but will depend upon the implementation and application. Typical PSNs might include individuals and their work contacts, groups of friends, groups of individuals with common interests or individuals conveniently grouped together for any other reason.

[0041] Typically (although not necessarily) each individual may choose his or her direct "buddies" (that is those individuals with which there will be a direct link) , with the second and higher order PSNs being generated by the system without direct involvement from the individual representing the root node. Alternatively, the entire PSN may be generated automatically, for example, by analysing things done by the individuals, without any individual needing to confirm explicitly whether another individual should or should not be added to the PSN.

[0042] If affinity ordering is to be carried out on the result items (see sections 3.7 and 4.12), the links or edges in the digraph need to be weighted. The weighting could be fixed (e.g. family members have a higher weighting than friends,) or it could be dynamic (for example on the basis of past actions affecting the relationship between connected individuals). Weightings could include such things as the time the linked individual has been known, the number of messages exchanged, the number of times a good recommendation has been received, and so on. To ensure that out of date information is not maintained within the database, the weighting could decay with time. Edge weights may be unidirectional or bi-directional. For the message used to weight the edges in the preferred embodiment, see section 4.12.

3. Apparatus

[0043] A schematic view of the apparatus used in the present embodiment is shown in figure 2. A DBMS 10 stored on an ABQ server 12 is used to generate an in-memory view 14 of the PSN. The in-memory view is used to generate an affinity filter 20 which filters the raw key word matches 16, 18 to provide a list of result items 22. On the basis of

information taken from the in-memory view 14, the result items 22 are ordered in an affinity order 24 to create a result list 26. Affinity ordering may be by relation, rating or result action. Finally the result list is encoded and presented to the user at 28 along with, where appropriate, a copy of the filtered keyword list 30.

[0044] Further details of the preferred apparatus of the present invention will be described in the following subsections.

3.1 Persistent storage of relations, actions and metadata.

[0045] The system requires some form of repository- preferably in persistent form- for the storage of relations, actions and metadata (such as date information and so on). Either a Relational Database Management System (RDBMS) or Object Oriented Database Management System (OODBMS) can be employed. In the specific embodiment to be described, an RDBMS is used, and code samples are given in generic SQL (Structured Query Language).

[0046] In order for persistence of relations, actions and metadata to be realised in an RDBMS, normalisation of the data into database tables needs to be done. Simplistically speaking, the database must hold information on a number of PERSONS, the relations between PERSONS, and data on their actions (e.g. purchase products, rate products, store documents, write articles etc.).

[0047] To that end, figure 3 illustrates the key PERSON and PERSON_TO_BUDDY tables. Keys within the table are indicated by the hash symbol.

[0048] In addition to information on individuals, the database also stores information on actions. (The exact nature of an action, and the information to be stored, will depend upon the application, but typically an "action" represents something that has been done by an individual, or statistics on a plurality of such events which may be of interest to the search initiator.)

[0049] For the purpose of illustration, we will describe an online bookstore example. However, it should be understood that the same principles apply whether actions are for an article rating, a URL search, for a skills match, etc. In each case, appropriate action and look up tables need to be added to the database in a similar manner to those described below. Such a task is within the capabilities of any database designer.

[0050] In an RDBMS, actions are stored within specific database tables such as those illustrated in figure 4. The actions here specifically relate to things being done by a person in relation to a specific book, such as purchasing it, reviewing it, rating it and so on. Metadata will typically be stored as well, for later retrieval, such as the date that the book was purchased, reviewed or rated.

3.2 Add/Modify/Delete Relations, Actions and Metadata

[0051] Once the basic schema has been defined in a database specific format, a mechanism to insert (add), update (modify) and delete relations, actions and metadata needs to be provided.

[0052] Because the bookstore example uses a relational database, SQL provides a process to achieve this requirement through INSERT, UPDATE and DELETE keywords. e.g. to create a relation from one buddy to another, the following SQL could be used:

```
INSERT INTO    PERSON_TO_BUDDY
               (ROOT_ID_FROM, ROOT_ID_TO, FRIEND_SINCE, NUM_MSGS, ...)
VALUES         (id, id2, DATE, 0, ...);
```

[0053] In the above, the elements id and id2 are provided as parameters.

3.3 Marshal In Memory Views

[0054] Bringing results to queries back from the database (i.e. as result items) requires an in-memory view of the PSN to be constructed in computer memory. This method is described in section 4.9. Suffice to say however that a working storage apparatus is needed for this process.

3.4 Publicly Volunteer Actions

[0055] Actions are classified by type and encoded in a client message header (or similar) from the client (search initiator) to the ABQ server. For example, ACTION_BUDDY_IM refers to a standard message from one buddy to another. ACTION_REFERRAL and ACTION_PURCHASE refer to a specific class of action.

[0056] Application specific action classifications can be derived according to the requirements of the implementation. In the bookstore example however, ACTION_PURCHASE_BOOK is used to distinguish this type of action from others,

e.g. ACTION_REFERRAL_URL for a URL referral, ACTION_REFERRAL_DOC for a document referral or ACTION_REFERRAL_ARTICLE for an article referral etc. Action types are embedded in a message header and handled in an ABQ server process. Thus an appropriate apparatus for encoding messages needs to be in place.

[0057] Preferably a community of users needs to consent to publicly volunteering actions. This has obvious privacy issues and, as such, implementers of the system may define their own policies and implementation rules. However, assuming an 'open' approach to information sharing is adopted (without exposing sensitive details such as account details etc.), the ABQ framework needs to capture and store publicly volunteered information on community actions.

[0058] Actions are volunteered either 'directly' or 'indirectly'. For example, a 'referral' from one PERSON to another is an example of a direct action. In a referral (e.g. book, article, product, service, URL etc.), a PERSON sends a specific type of action message to a PERSON ('buddy'), which then gets encoded in the message header (e.g. ACTION_REFERRAL_BOOK).

[0059] An example of an indirect action is a rating. A rating is either applied to an artefact or document (e.g. a book, product, article etc.) or to a PERSON (e.g. rating proficiency in a skill etc.). That is to say, indirect actions are applied to information or a PERSON without being passed to another PERSON. These are also encoded in the action field of the header as ACTION_RATING_BOOK,
ACTION_RATING_ARTICLE or ACTION_RATING_SKILL etc.

[0060] The derivation of action types is application specific, but the apparatus should typically classify actions by type and allow the community to volunteer actions publicly. Of course actions can be performed privately; however, the more that is shared, the more productive the system becomes on the whole.

3.5 An Apparatus to Create ABQ's

[0061] In compiling the result list, all actions need to pass through a central process (the ABQ server 12 in figure 1) in order for the system to act on them. Actions not involving any data storage or specific processing can simply be routed onto the PERSON recipient(s), e.g. ACTION_BUDDY_IM. However, other direct and indirect actions may require the system to respond. One example of this in the online bookstore system is the action of purchasing of a book (identified by the ACTION_PURCHASE_BOOK action type). This requires the system to make an association between the BOOK and PERSON. For example, following the server receiving and decoding this type of action, the following SQL would be executed, with book_id and root_id being provided as parameters:

```
INSERT INTO   BOOK_TO_PERSON
              (BOOK_ID, ROOT_ID, PURCHASE_DATE, ...)
VALUES        (book_id, root_id, DATE, ...);
```

[0062] Once an apparatus for sending, processing and storing data on actions is in place, further apparatus relating actions back to result items needs to be considered.

[0063] An apparatus to capture the initial search string (e.g. from a web form) is required. This can be a simple edit box on a web page with a server script to process the input. Information from client-side cookies (or similar) can extrapolate the search initiator's ROOT_ID and prepare the QUERY_STRING for the server CGI script. The server script will then parse the QUERY_STRING and construct a standard expression, transposed into SQL (in the RDBMS case).

[0064] Within the server scripts lie the rules for ABQ's as laid out in section 4, where details of the algorithms are contained. Suffice it to say at this stage that there should be:

- a server-based apparatus (database) to store information on persons, relations and actions;
- a server-based apparatus to process input keyword queries (e.g. CGI scripts) and;
- a process to present the result list of result items back to the initiator (e.g. CGI + DHTML).

3.6 Affinity Filtering of Result Items

[0065] Stored actions become result items once they are affinity 'filtered' and 'ordered'. Until then, they exist only in the database as actions. Once a keyword search has been composed by a search initiator, a standard database SELECT query (or similar) extracts a number of keyword matches from the action tables. For example, in our book system scenario, the search initiator could enter the word 'Bronte' as a search string. A simple keyword lookup in the database (using the BOOK, BOOK_TO_KEYWORD and KEYWORD action tables from our example) may then produce the following keyword matches:

- *Charlotte and Emily Bronte : The Complete Novels : Jane Eyre, Wuthering Heights, Shirley, Villette, the Professor -- Charlotte Bronte (Contributor), Emily Bronte; Hardcover*

- *Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback*
- *Wuthering Heights -- Emily Bronte; Paperback*

[0066] While providing valid keyword matches, this example has not has not yet taken into account any filtering with respect to 'affinities' in the PSN. Therefore, before the result items are ordered and presented back to the search initiator, keyword matches need to be filtered and then affinity ordered.

[0067] To achieve affinity filtering, the ROOT_ID of the search initiator needs to be provided (e.g. through cookies again) as does the 'distance' to search in the PSN (i.e. the ROOT_ORDER). If for example ROOT_ORDER = 3, the search initiator is prepared to extend the ABQ to a 'buddy of a buddy of a buddy' in their PSN.

[0068] Once the in memory view of the PSN has been constructed (see section 4.9), each of the keyword matches from the above list can be compared to each node in the in memory view. By extracting the ROOT_ID from each node in the PSN and repeatedly checking the return code from a SELECT query, a filtering of results matching PERSON's in the PSN to elements in the keyword match list can be made to produce filtered result items, e.g. (for each PSN in memory view node). In the following, book_id and root_id are provided as parameters:

```
SELECT    BOOK_ID
FROM      BOOK, BOOK_TO_PERSON, BOOK_REFERRAL
WHERE     BOOK.BOOK_ID=book_id
AND       BOOK.BOOK_ID=BOOK_TO_PERSON.BOOK_ID        // purchase action
AND       BOOK_TO_PERSON.ROOT_ID=root_id;
```

[0069] If the return code from this SELECT query indicates success, keyword matches can be categorised as being a filtered result item or; if the SELECT returns false, a filtered keyword match.

[0070] Filtered keyword matches can be ordered according to the matching rules applied by the default keyword ordering system, or whichever other system is employed.

[0071] From the book example, let us assume that the two 'Wuthering Height' books have been purchased by members in the search initiators PSN where ROOT_ORDER <= 3 from the search initiator. Then, following the affinity filtering process, the following lists are produced:

Filtered result items

[0072]

- *Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback*
- *Wuthering Heights -- Emily Bronte; Paperback*

Filtered keyword matches

[0073]

- *Charlotte and Emily Bronte : The Complete Novels : Jane Eyre, Wuthering Heights, Shirley, Villette, the Professor -- Charlotte Bronte (Contributor), Emily Bronte; Hardcover*

[0074] The next phase for the Filtered result items list is to affinity order them and form the result list.

3.7 Affinity Ordering of Result Items

[0075] Affinity ordering the result items depends on a user preference. Results can either be ordered by rating, relation or result action (see section 3.8). Rating ordering however, looks at the filtered result items and orders them according to the highest ratings applied by referees (see section 4.11), e.g.

Filtered and ordered result items (ordered by rating)

[0076]

- *Katie, who knows Nick purchased Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback, rating it 4.5/5*
- *Nick purchased Wuthering Heights -- Emily Bronte; Paperback, rating it 3/5*

[0077] Relation ordering orders the filtered result items by the strength of a relation, i.e. the 'edge weight' between referee nodes in the PSN (see section 4.12).

Filtered and ordered result items (ordered by relation)

[0078]

- *Nick purchased Wuthering Heights -- Emily Bronte; Paperback, rating it 3/5*
- *Katie, who knows Nick, purchased Wuthering Heights (Oxford World's Classics) -- Emily Bronte, Patsy Stoneman (Editor); Paperback, rating it 4.5/5.*

[0079] Result action ordering sorts the filtered result items according to how many result actions have been followed by the referee's past results (see sections 3.8 and 4.13 for more on result action ordering).

[0080] The now affinity 'filtered and ordered' result items are now referred to as the result list and can be encoded and presented back to the search initiator (see section 4.14).

3.8 Feeding Back Result Actions into the database.

[0081] Not all implementations of an ABQ system will necessarily include feedback. However, its usefulness is seen to be: (i) as another term by which to affinity order result items (see section 3.7) and; (ii) as a means of both rewarding and motivating a community to publicly volunteer actions (e.g. 'this week's most popular referee was **XXX'** etc.). Once the result list has been filtered, ordered, encoded and presented, and following a search initiator selecting a result action (i.e. following a result item), feedback can take place.

[0082] An apparatus for feeding back result actions into the database involves extracting the result identifier and referee identifier from the result item chosen by the search initiator. Once result and referee identifiers have been extracted, an SQL SELECT statement (or similar) can be formed to reward the referee by incrementing a reserved field in the database relating to the referee's node.

[0083] From figure 4, the BOOK_TO_PERSON table contains a RESULT_ACTIONS column: this is the reserved field used to count the number of result actions taken by search initiators of ABQ's.

[0084] For example, if the result action followed a result path in response to a buddy's action of purchasing a book, then the BOOK_TO_PERSON.RESULT_ACTIONS field would be updated following the result action, e.g.

```
UPDATE   BOOK_TO_PERSON
SET      BOOK_TO_PERSON.RESULT_ACTIONS=BOOK_TO_PERSON.RESULT ACTIONS+1
WHERE    BOOK_TO_PERSON.BOOK_ID=result_id
AND      BOOK_TO_PERSON.ROOT_ID=referee_id;
```

[0085] Where result_id and referee_id are provided as parameters.

[0086] Therefore, the ABQ apparatus should encode referee and result identifiers in the result list in order to reward a referee. Considering the bookstore system example, the result identifier is the BOOK_ID and the referee identifier is a ROOT_ID. The presentation of the results would contain elements (e.g. XML tags) identifying the result identifier (in this case, the BOOK_ID) and the referee identifier (e.g. the BOOK_TO_PERSON.ROOT_ID field) for each of the result items held in the result list, e.g.

```
<?xml version='1.0'?>

<!DOCTYPE ABQ SYSTEM "abq.dtd">
<results>
  <title>ABQ Results</title>
  <resultlist>
     <resultitem path="http://some.domain.com/books/bronte.html#wh">
        <resultaction>../cgi-bin/resultAction.pl</resultaction>
        <referee id="567321">Katie</referee> who knows
        <relation id="567274">Nick</relation> purchased
        <result id="984527">Wuthering Heights</result> rating it
        <rating>4.5</rating> out of 5
     </resultitem>
     ...
  </resultlist>

     </results>
```

[0087] A result item path (in the book example) is an XML tagged URL pointing to a web page about the book, and a result action would be the search initiator clicking on one of the result paths. After which, SQL such as that in section 3.8 can be run once the required tags have been extracted from the result item.

[0088] Apparatus to encode referee and result identifiers in the result should be provided (XML tags are one solution). Additionally, apparatus to capture result paths taken by search initiators (e.g. parameter passing back a CGI QUERY STRING on the server from a web browser). Finally, apparatus to process the result path (from the result action) needs to update the NUM_REFERRALS fields for the referee and present the result endpoint to the search initiator (e.g. a CGI server script). The result endpoint is considered (in the book example) a web page describing the book. Other examples of result endpoints are: (i) product pages; (ii) articles; (iii) email addresses; (iv) newsgroup articles; or (v) URL's and documents etc.

4. Methods

[0089] The following sub-sections describe the methods required to build the example book ABQ system. Details such as user interface have been omitted, as have references to specific Internet browsers or other software etc.

[0090] While some of these specific details of the methods described relate particularly to the example book ABQ system, the underlying principles are generic and it is well within the capabilities of a skilled person in the field to modify the methods shown for use in other applications. In effect, all that is required to implement another system is to capture information about referees, results, actions and metadata in any convenient way, and to modify the methods described below to handle that information in an appropriate manner.

4.1 A method to create the ABQ database schema

[0091] The first step is to create a database. Standard ANSI SQL can achieve this (assuming an RDBMS is being used). Once this is done, the ABQ schema needs to be defined. This is best illustrated through code. The following SQL illustrates schema creation for the online bookstore system (see figure 4)

```
USE ABQ;                                                    // assumes DB = 'ABQ'
CREATE TABLE PERSON                                         // Person table
(
     ROOT_ID           INT UNSIGNED PRIMARY KEY,            // primary key
     FIRST_NAME        VARCHAR,
     LAST_NAME         VARCHAR,
     EMAIL             VARCHAR,
     ...                                                    // other fields
};
CREATE TABLE PERSON_TO_BUDDY                                // Buddy table
(
     ROOT_ID_FROM      INT UNSIGNED,                        // foreign key
     ROOT_ID_TO        INT UNSIGNED,                        // foreign key
     FRIEND_SINCE      DATE,
     NUM_MSGS          INT UNSIGNED,                        // used for ordering
     ...                                                    // other fields
};
CREATE TABLE BOOK                                           // Book table
(
     BOOK_ID           INT UNSIGNED PRIMARY KEY,            // primary key
     TITLE             VARCHAR,
     AUTHOR            VARCHAR,
     PUBLISHER         VARCHAR,
     SUMMARY           LONG VARCHAR,                        // summarised précis
     ENTRY_DATE        DATE,
     ...                                                    // other fields
};
CREATE TABLE BOOK_TO_KEYWORD                                // Book keyword table
(
     BOOK_ID           INT UNSIGNED,                        // foreign key
     KEYWORD_ID        INT UNSIGNED,                        // foreign key (hash)
     ...                                                    // other fields
};
CREATE TABLE KEYWORD                                        // Keyword table
(
     KEYWORD_ID        INT UNSIGNED PRIMARY KEY,            // primary key (hash)
     KEYWORD           VARCHAR,
     ...                                                    // other fields
};
CREATE TABLE BOOK_TO_PERSON                                 // An 'action' table
(
     BOOK_ID           INT UNSIGNED,                        // foreign key
     ROOT_ID           INT UNSIGNED,                        // foreign key
     PURCHASE_DATE     DATE,
     REVIEW            LONG VARCHAR,                        // used for reviewing
     RATING            INT UNSIGNED,                        // used for filtering
     RESULT_ACTIONS    INT UNSIGNED,                        // used to reward
results
     ...                                                    // other fields
};
```

4.2 A method to store referees (PERSON's) in a database

[0092] Once the database and schema is constructed, the system waits for PERSON 'registrations' and 'logins',(see figure 5) This may be realised through an Instant Messaging client interface or extrapolated from information known to hold information on a Personal Social Network - e.g. through scanning closed community email address books - or through a more open online community (e.g. a newsgroup) where actions between PERSON's can be intercepted and processed by the ABQ server.

[0093] In the bookstore system example, an Instant Messaging client could be used to accompany the ABQ service and provide the raw relation data for the PSN.

[0094] The IM system works in conjunction with an Internet browser to monitor actions and act on them. Therefore, the following process outlines joint REGISTRATION and LOGIN methods, adding user details into the PERSON table in the database in the case of new registrations. For clarity, exception, primary key and handling processes etc. are omitted from figure 5.

4.3 A method to modify PERSON records

[0095] Should a PERSON wish to alter their details (e.g. their email address), a method allowing them to do so must be in place. This can be realised for example through the following SQL.

```
UPDATE    PERSON
SET       PERSON.EMAIL=email
WHERE     PERSON.ROOT_ID=root_id;
```

[0096] Figure 6 shows the process in flow chart form.

[0097] This assumes that the user has already registered with the ABQ service and has been allocated a unique ROOT_id. There after, the user is logged in and is known to the system (see section 4.2).

4.4 A method to delete PERSON records

[0098] Should a PERSON wish to delete their details completely, a method allowing them to do so must be in place. This can be realised for example through the following SQL.

```
DELETE FROM    PERSON
WHERE          PERSON.ROOT_ID=root_id;
```

[0099] Figure 7 shows the process in flow chart form. As with figure 6, this assumes that the user has already registered.

4.5 A method to store relations between PERSON's

[0100] Following a process of 'find buddy' (i.e. a method allowing users to input values to search the database for buddies), a selected buddy is inserted to the PERSON_TO_BUDDY table using SQL (e.g. see listing under section 3.2).

[0101] The process is shown in flow chart form in figure 8. At box 80, a number of different terms may be used to find buddies. In its simplest form the PERSON may know the ROOT_id of the buddy they wish to add. In a more complex example, the PERSON may know only the surname, or some other partial details.

[0102] For clarity, no reference has been made here to other tables which might be present in a practical example. These could include a PERSON_TO_IGNORE table which has similar properties to the PERSON_TO_BUDDY table and which provides a mean of ignoring actions from unwanted individuals within a community.

4.6 A method to delete relations

[0103] As well as adding buddies into a buddy list (or similar), a PERSON may want to remove a buddy from a buddy list. This could be done with the following SQL:

```
DELETE FROM    PERSON_TO_BUDDY
    WHERE          PERSON_TO_BUDDY.ROOT_ID_FROM=root_id_from
    AND            PERSON_TO_BUDDY.ROOT_ID_TO=root_id_to;
```

[0104] Figure 9 shows the process in a flow chart form.

4.7 A method to summarise and store actions

[0105] Actions from PERSON's are preferably summarised. This is so that keywords from the actions can be extracted and stored in a keyword table (for keyword matching). Summaries are stored along with action records, as are other data/metadata pertaining to the action.

[0106] The specifics of action summarisation are not part of the invention and various options are available. Summaries and keywords can be verbally agreed, provided (e.g. by a publisher in the example bookstore system) or extracted automatically (e.g. from a content or document summariser).

[0107] Following an action, a summarised version of the action is stored along with the data, metadata and keywords about it.

[0108] Figure 10 illustrates the method in flow chart form.

[0109] At box 100, an action key is generated or obtained. Keys can be generated in a number of ways (for example

by the use of hash values): any convenient approach can be used. At box 110, there is a check made of the last action summary date. This process allows for already-summarised actions to be re-summarised if they are out of date. This may be particularly necessary for volatile data, for example on the web. Box 120 illustrates obtaining the action summary and key words: this may either be provided manually or, preferably, through an automated content summariser that reads documents or texts and that provides summaries and key words. With particular reference to the bookstore system, at box 130 the action would be stored in the BOOK_TO_PERSON or the BOOK_REFERRAL tables.

4.8 A method to process incoming ABQ parameters

[0110] A generic approach to the processing of incoming ABQ parameters is shown in figure 11. At box 140, the ROOT_ID is obtained; this may be through the use of cookies (if a web browser is in use), or obtained from a registry setting, or from elsewhere. The encoding step shown at box 150 may use a bespoke TCP/UDP application header, or may encode the values in a QUERY_STRING. Further details of the items shown within the dotted line 160 may be found at section 4.10.

4.9 A method to build in memory views

[0111] In memory views require building the PSN digraph in computer memory. To do this, relations from the database need to be built with a number of abstract data types. There are numerous ways to do this and the following is but one example using a 'list of lists' (modelling a digraph).

[0112] However, the principle and methods of the flowchart (figure 12) remains the same irrespective of whichever implementation method of holding the in memory view is chosen.

[0113] At box 200 of figure 12, the next ROOT_ID is extracted. This is the ROOT_ID of the next node in the PSN to look up in the database. In the first loop it is the ROOT_ID of the service initiator.

[0114] Figure 13 below shows how by iteratively extracting PERSON 'buddies' from the database leads to the production of an in memory view of the PSN digraph (binary tree description follows later).

[0115] Nodes are internally represented by 'objects' and edges encapsulate pointers or object references into other objects. By implementing the digraph using a doubly linked list (for example), a path back to the root node from any other node in the digraph can be traced. This is used to determine the PSN relations (e.g. 'Kate, who knows Nick...' etc.)

[0116] Nodes are placed in the graph in sequential order as read from tuples returned from a SELECT query. Each level of indirection in the graph relates to another ROOT_ORDER from the root node (i.e. 'ME'). This structure is built in computer memory from repeated SELECT queries iterating out from the root node. Nodes appear in a number of places due to the nature of the PSN.

[0117] Of course, tracing directly into the digraph (and avoiding duplicates) requires another process. This can be implemented through the use of a binary tree (using ROOT_ID's of each PERSON as the tree node value). A fast method of getting into the digraph can be achieved by holding pointers to digraph objects inside a binary tree, as shown in figure 14.

[0118] In figure 14, the numbers before the names in the binary tree indicate ROOT_ID values from the database.

[0119] The use of the binary tree allows objects to enter the digraph once only. That is to say, relations retrieved from the database are first checked if they appear in the binary tree. If so, they are discarded (as a 'shortest path' to them has already been computed — apart from reference to the root node). If not, they are added to the digraph and their pointer is added into the corresponding ROOT_ID node location in the binary tree for fast access into the digraph.

4.10 A method to affinity filter ABQ keyword matches

[0120] Once the in memory view is built, the next phase is to filter the keyword matches into filtered result items and filtered keyword matches (see section 3.6). The process is shown in figure 15.

[0121] At box 210, we can see obtain keyword matches. This of course assumes that we have managed to find keyword matches to the search initiator's ABQ keyword list-see section 4.8.

[0122] Following this process, keyword matches have been processed into two lists: (i) filtered keyword matches (ordered by keyword) and (ii) filtered result items (also ordered by keyword). The next phase is to affinity order the filtered result items.

4.11 A method to affinity order result items by rating

[0123] Now that we have filtered the results into filtered keyword matches (no longer considered until encoding and presentation) and filtered result items, we need to order them.

[0124] Ordering by rating involves sorting the filtered result items according to ratings applied by PERSON's in the

PSN. Although filtering and ordering can be seen as two separate tasks, to reduce the number of database accesses, the in memory view can be prepared according to the method of ordering selected by the search initiator during filtering. For example, if the search initiator chose to order by rating, then whilst filtering the results (section 4.10), the database query could extract information on ratings applied at the same time as filtering, as in the following SQL:

```
SELECT   BOOK_ID, BOOK_TO_PERSON.RATING
FROM     BOOK, BOOK_TO_PERSON
WHERE    BOOK.BOOK_ID=book_id
AND      BOOK.BOOK_ID=BOOK_TO_PERSON.BOOK_ID
AND      BOOK_TO_PERSON.ROOT_ID=root_id;
```

[0125] Combining the filtering method with collecting information useful for ordering the result list can drastically speed up performance during the ordering phase. This is because one database read per PSN referee for filtering places 'ordering' information into the in memory view allowing traversal of memory to take place rather than repeated calls (and locks) to the database.

[0126] Considering that we already have the rating information in the in memory view, we can proceed rating order as shown in figure 16.

[0127] The ordered linked list in figure 16 is now in effect a linear version of the in memory view, ordered by PERSON's who have rated a result item. Thus, looping through each result item and repeatedly sorting the ordered linked list (according to each result item) and copying the results into another list produces a filtered and ordered list of result items. This list may contain duplicate references to a result item (where more than one person has rated a result item).

[0128] It will be understood of course that the sorting may be carried out using any convenient method, including quicksort, bubblesort, mergesort, insertsort, and so on.

4.12 A method to affinity order result items by relation

[0129] As mentioned in section 4.9, each node of the in memory view holds the ROOT_ID of a PERSON so that filtering and ordering of keyword matches can take place.

[0130] The other attribute held in the digraph - for relation ordering only - is the reciprocal edge weighting (Ew) value between PERSON's.

[0131] Ew can be computed in a number of ways and derived according to the application. However, the following method is sufficient for many needs and uses derivatives of action and time in its calculation. The first part of computing Ew between two PERSON's (A→B) is to calculate the total number of actions sent from A to all buddies (where ROOT_ORDER = 1) and is given as:

$$A = \sum_{i=1}^{j} a_i \qquad \text{(eq. 4.1)}$$

where: a = the number of actions sent from A to the i^{th} buddy, where A's root order=1

[0132] The next calculation is the number of actions received by A from all buddies and is given as:

$$B = \sum_{i=1}^{j} b_i \qquad \text{(eq. 4.2)}$$

where: b = the number of actions received by A from the i^{th} buddy, where A's root order=1

[0133] Next comes calculating Mt. Mt represents the maximum time anyone has been a buddy of A and is used in calculating Aw and Bw (below). Aw represents the edge weight from A to B and is given as:

$$Aw = \frac{a/A}{Bt/Mt} \qquad \text{(eq. 4.3)}$$

where:

a = the number of actions sent from A to B

A = the sum of all actions sent from A to all buddies, given in (eq. 4.1)
Bt = the time in which B has been a buddy of A
and
Mt = the maximum time anyone has been a buddy of A

[0134] Bw represents the edge weight to A from B and is given by

$$Bw = \frac{b/B}{Bt/Mt} \qquad \text{(eq. 44)}$$

where:

b = the number of actions sent to A from B
B = the sum of all actions sent to A from all buddies, given in (eq. 4.2)
Bt = the time in which B has been a buddy of A
and
Mt = the maximum time anyone has been a buddy of A

[0135] Finally, sum all Aw and Bw values for all buddies and compute Ew between A→B as follows:

$$Ew_{AB} = \frac{Aw + Bw}{\sum_{i=0}^{I} Aw_i + \sum_{i=0}^{I} Bw_i} \qquad \text{(eq. 4.5)}$$

[0136] Completing equation 4.5 for all buddies will produce a normalised array such that:

$$\sum_{m=1}^{n} Ew_m = 1 \qquad \text{(eq. 4.6)}$$

[0137] As an example, consider the following table showing the Ew values from A→B_1 .. B_5:

	A	B	C	D	E	F	G
1	From A to	*Bt*	*a*	*Aw#*	*b*	*Bw#*	*Ew#*
2	B_1	10	10	0.438	5	0.259	0.131
3	B_2	5	10	0.875	10	1.037	0.360
4	B_3	2	5	1.094	3	0.778	0.352
5	B_4	8	5	0.273	6	0.389	0.125
6	B_5	14	2	0.063	3	0.111	0.033
7							1.000
8							
9	*note. '#' indicates a calculated property*					*A#*	32
10	*...the rest are stored in the DBMS...*					*B#*	27
11						*Mt#*	14

[0138] Thus, by traversing the in memory view and calculating Ew values for each node, we can begin to build a picture of the PSN that (in our example) could resemble figure 17.

[0139] Once all Ew values have been calculated for the in memory view a process of sorting the relationships by root order and Ew value can produce the arrangement of figure 17. Once the in memory view is ordered, it is simply a matter for the algorithm to inspect each filtered result item and traverse the (ordered) in memory view as shown in figure 19.

[0140] In that figure, box 250 requires us to obtain the next ROOT_ID: this is the ROOT_ID of the next node in the memory view according to the current root order ("X" value). So, where X equals 1 the ROOT_ID order will be 1.1, 1.2, 1.3 (from figure 1),...etc.

4.13 A method to affinity order result items by result action

[0141] This method works in a similar manner to that presented in section 4.11. The in memory view does not require ordering (as in section 4.12) as ordering is done according to how 'valuable' PERSON's in the PSN are seen. Result actions (from section 3.8) are an incremental value and get updated after result actions have been taken by search initiators (section 4.15).

[0142] However, assuming the RESULT_ACTIONS field is modelled in the action tables of the database (see listing at section 4.1), affinity ordering the filtered result items by result action takes place in an identical fashion to that in section 4.11 with the exception that the final list is ordered on result action and not rating.

4.14 A method to encode and present result items

[0143] Once the filtered and ordered result items have been compiled, the penultimate phase involves encoding and presenting the result list back to the search initiator. An apparatus for this was presented in section 3.8. In this section, an example using XML was presented. However, the specifics of encoding are left open to the ABQ application developer. Suffice to say, the format of the (pre-encoded) result list will hold information on referee, result, relation and (where appropriate) rating. These may be fields within a software class held in an abstract data type such as a linked list. The result list is then used in conjunction with the binary tree and the in memory view to look up affinities with result items. These are then pushed through an encoding and presentation process and presented back to the search initiator. This is shown in figure 20. Note that the result list holds the result identifier, the binary tree the root identifier and the in-memory view other data (for example edge weights) and relational data with respect the root node.

4.15 A method to feedback result actions

[0144] As mentioned in section 3.8, a process to feedback results action made by search initiators is the final method to explain. This method assumes an apparatus tracking which result item the search initiator has selected can feedback values to the ABQ server (figure 2).

[0145] In the bookstore example, this apparatus is a hyperlink calling an ABQ 'resultAction' server script with the following three parameters:

- RESULT_PATH — the URL of a web page describing the product (book)
- REFEREE_ID — the referee identifier (see section 3.8 and listing at section 3.6)
- RESULT_ID — the result identifier the product (book).

[0146] This is shown in figure 21.

[0147] Once the result action have been sent back into the database, the ABQ process is complete, and the system then simply waits for the search initiator, or indeed any other members of the on line community, to start another search.

[0148] In a variant of the preferred embodiment (not shown) the system could allow the output to be ordered by more than one parameter. For instance, the output could first be ordered by relation, and then, within that, by rating. Options may be provided to allow the search initiator various possibilities.

Glossary of Terms

[0149]

ABQ	Affinity Based Query
Action	An activity performed by a member of a virtual or online community (e.g. visit URL, purchase book etc.). Actions can be direct, indirect, public and/or private
Affinity	The process of qualifying keyword matches against members of a
Filtering	Personal Social Network (PSN).
Affinity	The process of ordering affinity filtered result items with respect to
Ordering	rating, relation or result action

Digraph	Directed Graph. A way of presenting the Personal Social Network using nodes and directed edges.
Edge Weight	A relationship weighting between members of a Personal Social Network
In Memory	A computer memory model of the Personal Social Network held as a
View	collection of objects in an Abstract Data Type, examples are hash tables, linked lists and arrays etc.
Keyword	A matching between an entry in the ABQ database and a search
Match	initiators keywords. Keyword matches are pre affinity filtering and ordering
Metadata	Data about data. Stored data about actions not directly concerned with the action itself (e.g. the date the action was made etc.)
PSN	Personal Social Network. A way of describing relationships with respect to a root node
Rating	Part of a result item. A grade given either to information or to oneself
Referee	Part of a result item. A person within a Personal Social Network who is associated with information being sought by a search initiator.
Relation	Part of a result item. The method of describing the relationships between a search initiator and a result.
Result	Part of a result item. A matching of an action with a referee in an ABQ
Result Action	The activity taken by the search initiator following the presentation of the result list, e.g. following a hyperlink or emailing a relation etc.
Result Item	Part of a result list. An affinity filtered keyword matched action from the ABQ database
Result List	The product of keyword matches after they have been affinity filtered and ordered
Result Path	The location of a result, e.g. URL, directory path etc.
Root Node	The nucleus of a Personal Social Network.
Root Order	A way of describing 'distances' from a root node in a Personal Social Network to other members, i.e. 1st order relationships = 1 root order from the root node etc.
Search Initiator	An ABQ community user initiating a query.

Claims

1. A computer query method comprising receiving a query from a search initiator, carrying out a search for information on the basis of the query and filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

2. A computer query system as claimed in claim 1 in which the linked network of individuals is a personal social network.

3. A computer query system as claimed in claim 1 or claim 2 including the step of presenting the filtered results as a list.

4. A computer query system as claimed in claim 3 in which the hit is sorted according to the affinity of each result with individuals within the network.

5. A computer query system as claimed in any of the preceding claims including displaying each filtered search result in conjunction with information identifying the individual within the network that has an affinity with the result.

6. A computer query system as claimed in claim 5 further including displaying in conjunction with each filtered search result information identifying the type of affinity between the individual and the search result.

7. A computer query system as claimed in claim 6 in which the affinity consists of a rating applied by the individual to the result.

8. A computer query system as claimed in claim 5 further displaying in conjunction with each filtered search result information representative of a relation between the search initiator and the individual.

9. A computer query system as claimed in claim 4 in which the list is sorted according to ratings applied to the results by individuals within the network.

10. A computer query system as claimed in claim 4 in which the list is sorted according to a type or closeness of relation between the search initiator and the individuals which are associated with the filtered results.

11. A computer query system as claimed in claim 10 in which the list is sorted so that the first order relations appear before second order relations.

12. A computer query system as claimed in claim 10 in which the network includes weights associated with person to person links, the list being sorted in dependence upon the weights.

13. A computer query system as claimed in claim 12 in which the weights are time dependent.

14. A computer query system as claimed in claim 13 in which the weights are determined based on prior interactions between individuals within the network.

15. A computer query system as claimed in any one of the preceding claims including maintaining a database representative of the network of individuals.

16. A computer query system as claimed in claim 15 including constructing an in-memory view of a relevant part of the network, and filtering the search results according to information contained within the in-memory view.

17. A computer query system as claimed in claim 15 or claim 16 including maintaining within the database information on actions associated with or done by the individuals within the network.

18. A computer query system as claimed in claim 17 including maintaining information on a referral from one individual within the network to another.

19. A computer query system as claimed in claim 17 or claim 18 including maintaining information or messages between one individual within the network and another.

20. A computer query system as claimed in any one of claims 17 to 19 including maintaining information on the purchase of an item by an individual within the network.

21. A computer query system as claimed in any one of claims 17 to 20 including maintaining information on a rating by an individual of a document, a link or a product.

22. A computer query system as claimed in any one of claims 17 to 21 including maintaining information on a rating by an individual of another, or the same, individual within the network.

23. A computer query system as claimed in any one of the preceding claims in which the search is a keyword search.

24. A computer query system as claimed in any one of the preceding claims in which the search initiator selects the extent of filtering required by varying a minimum level of affinity which will return a result.

25. A computer query system as claimed in claim 15 including the step of selecting and carrying out a result action on one of the filtered results, and storing in the database information representative of the result action against the individual who was associated with the selected result.

26. A computer query system as claimed in claim 25 in which the said information stored against the individual comprises an action value.

27. A computer query system as claimed in claim 26 when dependent upon claim 3 in which the list is sorted according to the action values of the individuals associated with the filtered results.

28. A computer query system as claimed in claim 3 in which the search is a keyword search and in which the list is further sorted by keyword.

29. A computer program for carrying out a method as claimed in any one of the preceding claims.

30. A data-carrier carrying a computer program as claimed in claim 29.

31. A computer query system comprising means for receiving a query from a search initiator, means for carrying out a search for information on the basis of the query, and means for filtering the search results according to the affinity of each result with individuals within a linked network of individuals associated with the search initiator.

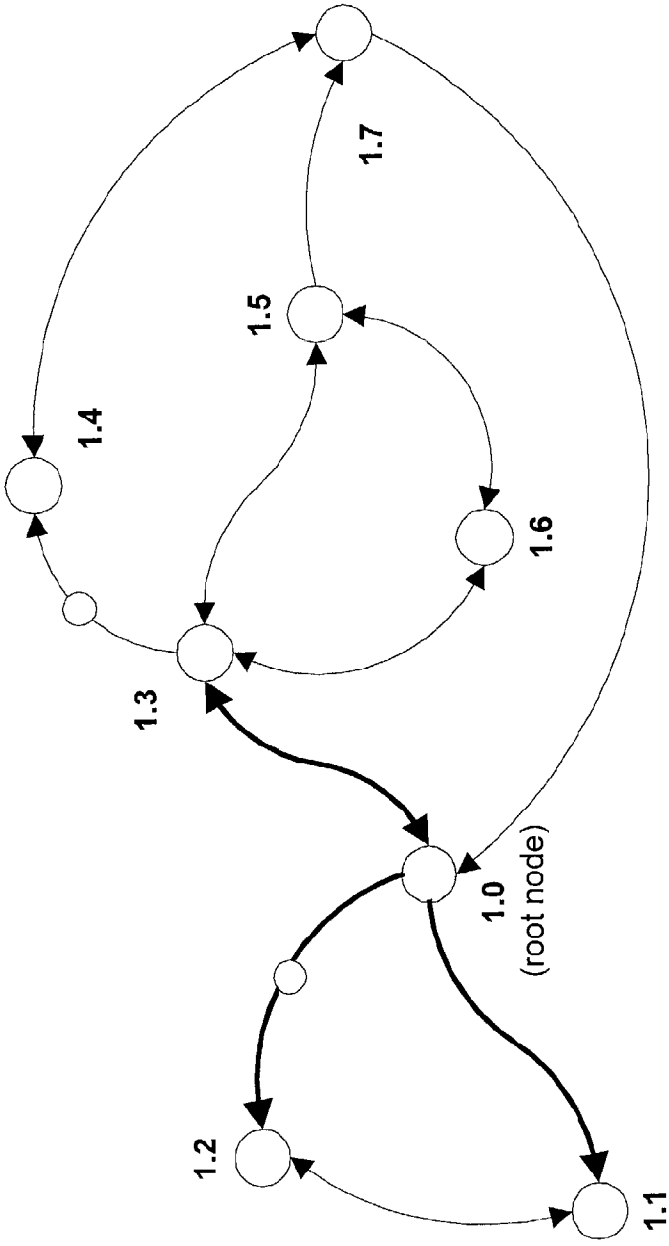


Figure 1

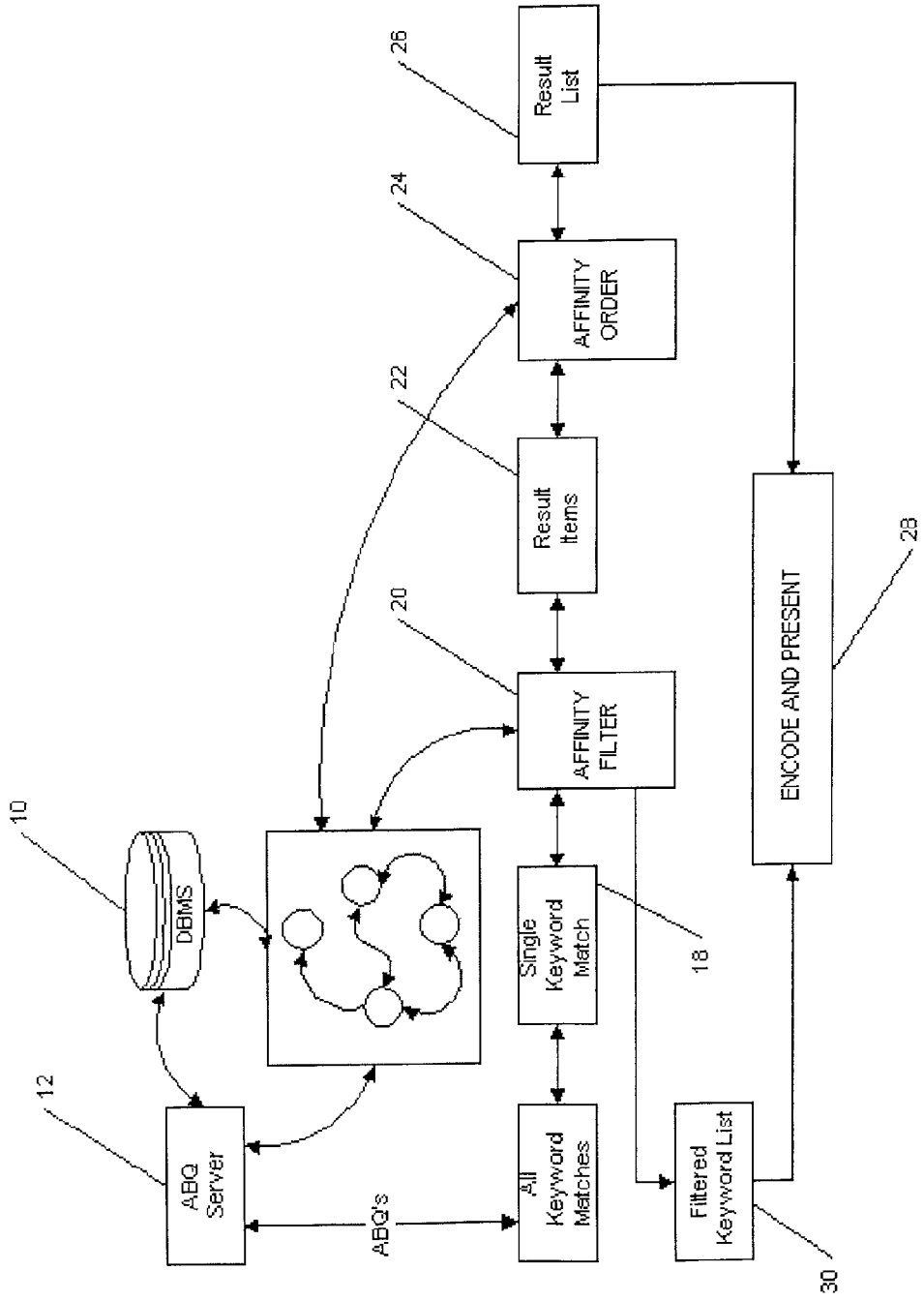


Figure 2

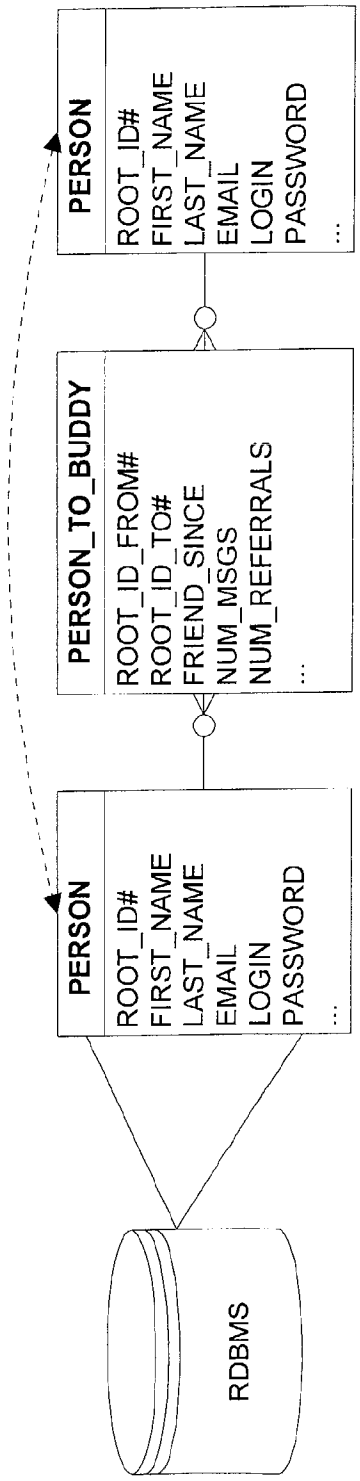

PERSON
ROOT_ID#
FIRST_NAME
LAST_NAME
EMAIL
LOGIN
PASSWORD
...
PERSON_TO_BUDDY
ROOT_ID_FROM#
ROOT_ID_TO#
FRIEND_SINCE
NUM_MSGS
NUM_REFERRALS
...
PERSON
ROOT_ID#
FIRST_NAME
LAST_NAME
EMAIL
LOGIN
PASSWORD
...
RDBMS

Figure 3

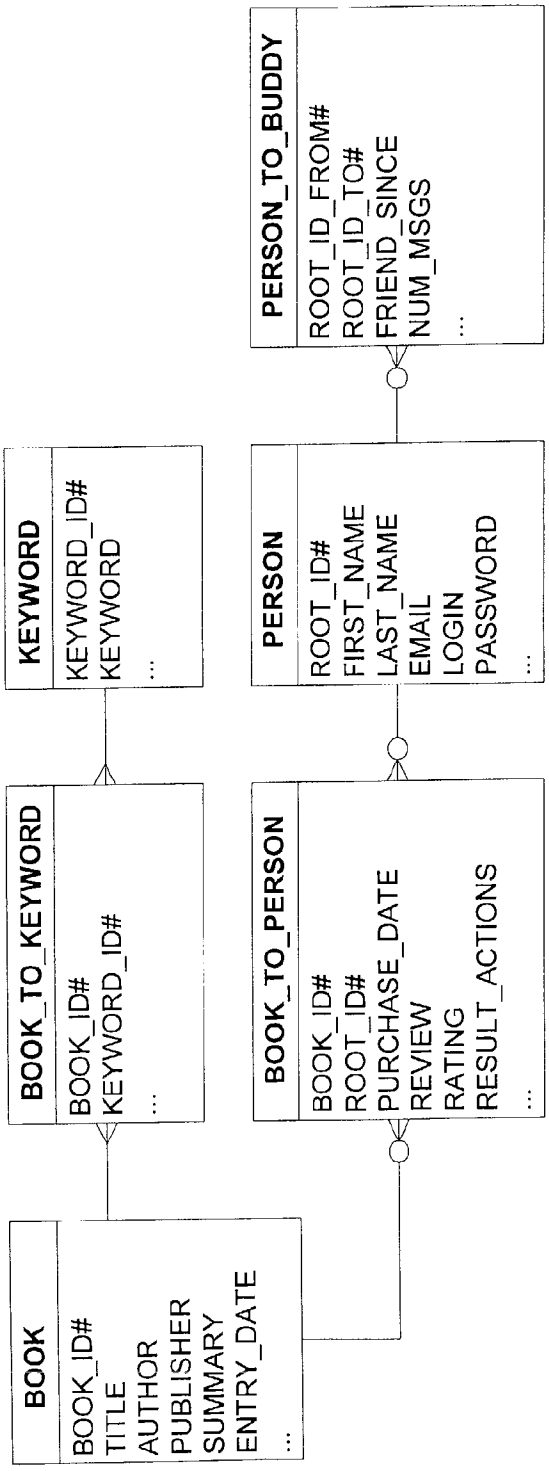


Figure 4

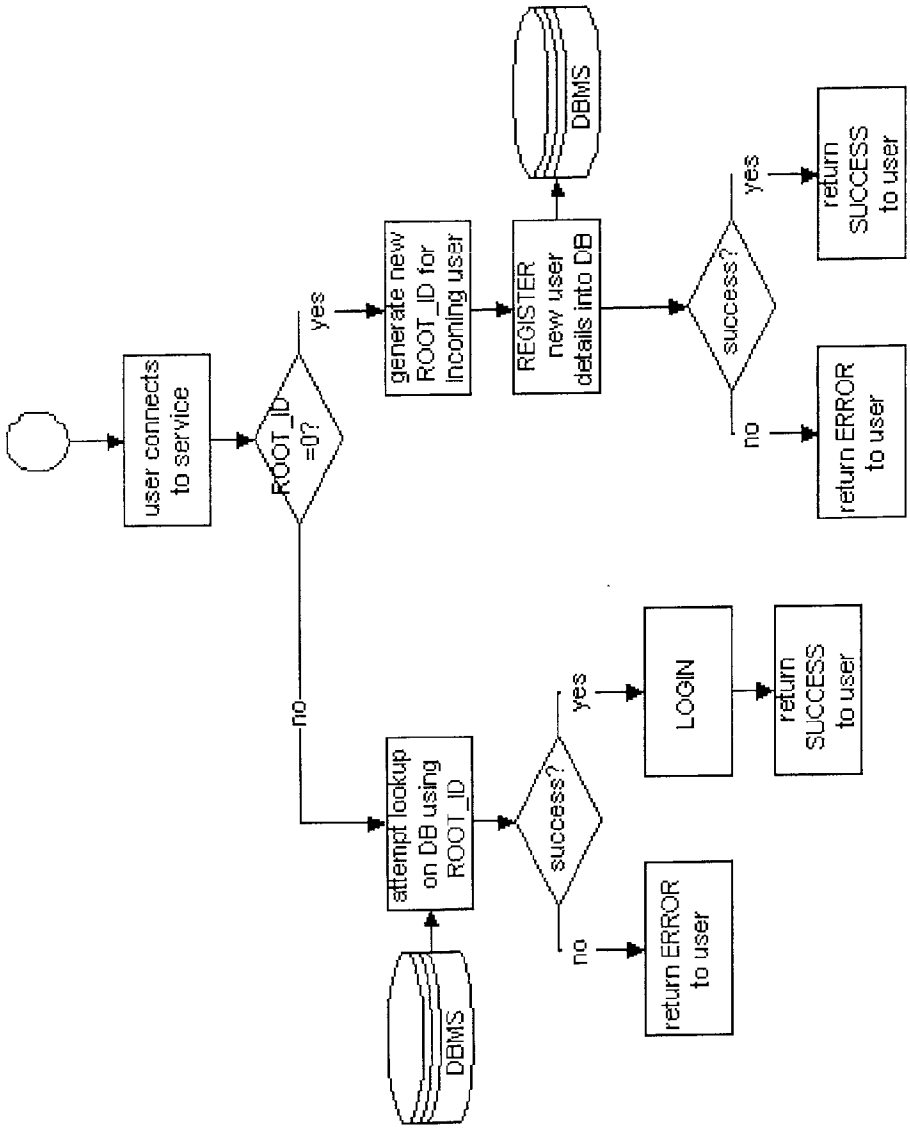


Figure 5

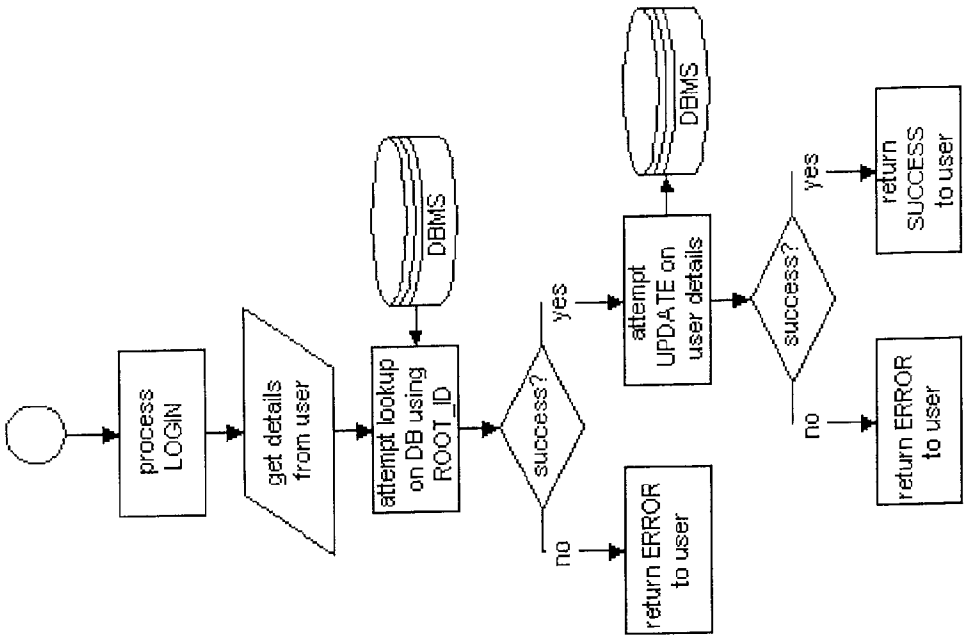

process LOGIN
get details from user
attempt lookup on DB using ROOT_ID
DBMS
success?
yes
no
return ERROR to user
attempt UPDATE on user details
DBMS
success?
yes
no
return SUCCESS to user
return ERROR to user

Figure 6

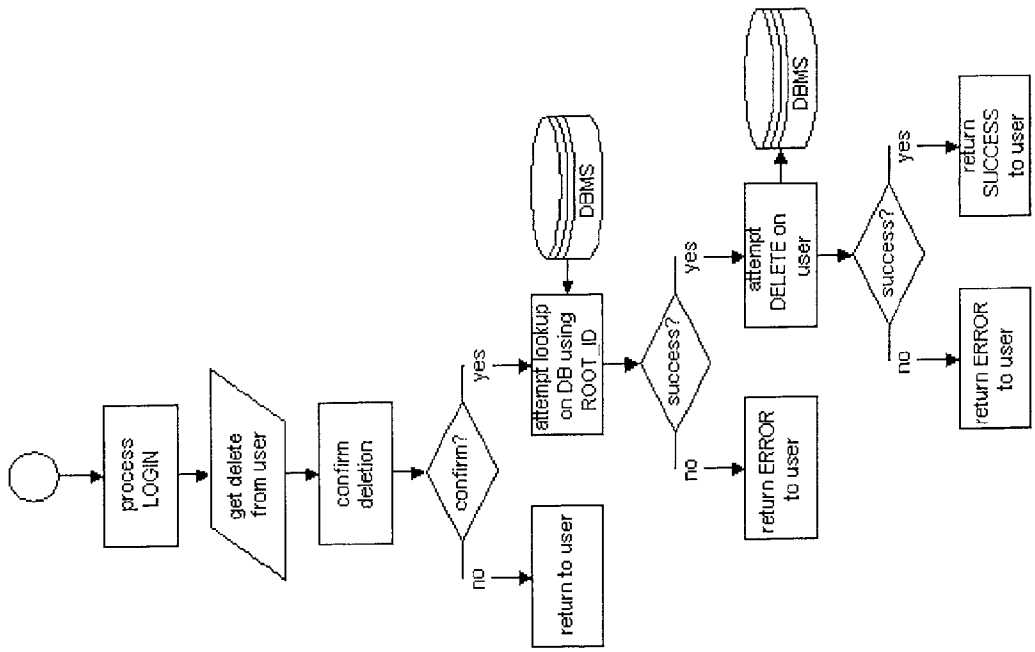


Figure 7

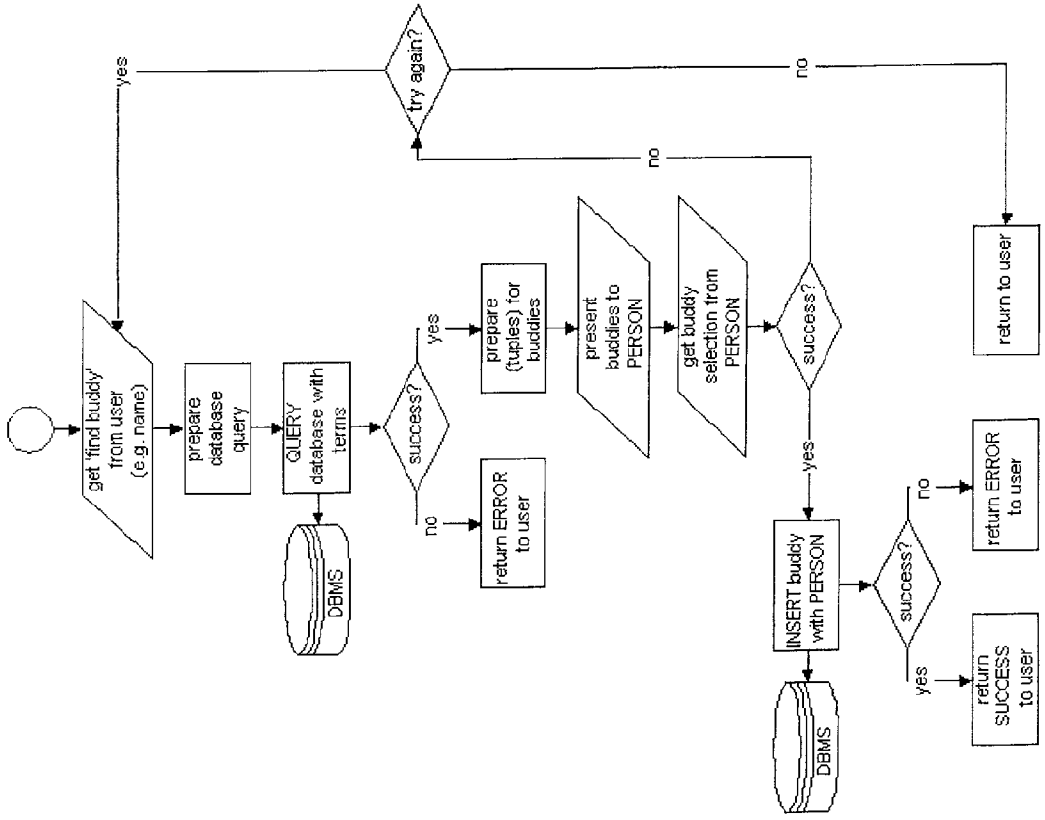


Figure 8

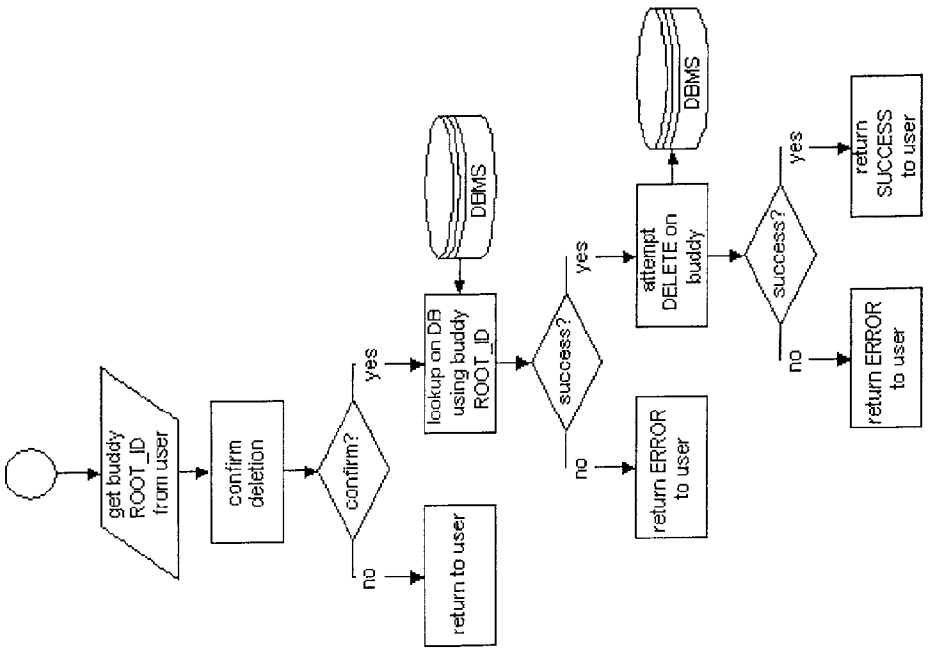


Figure 9

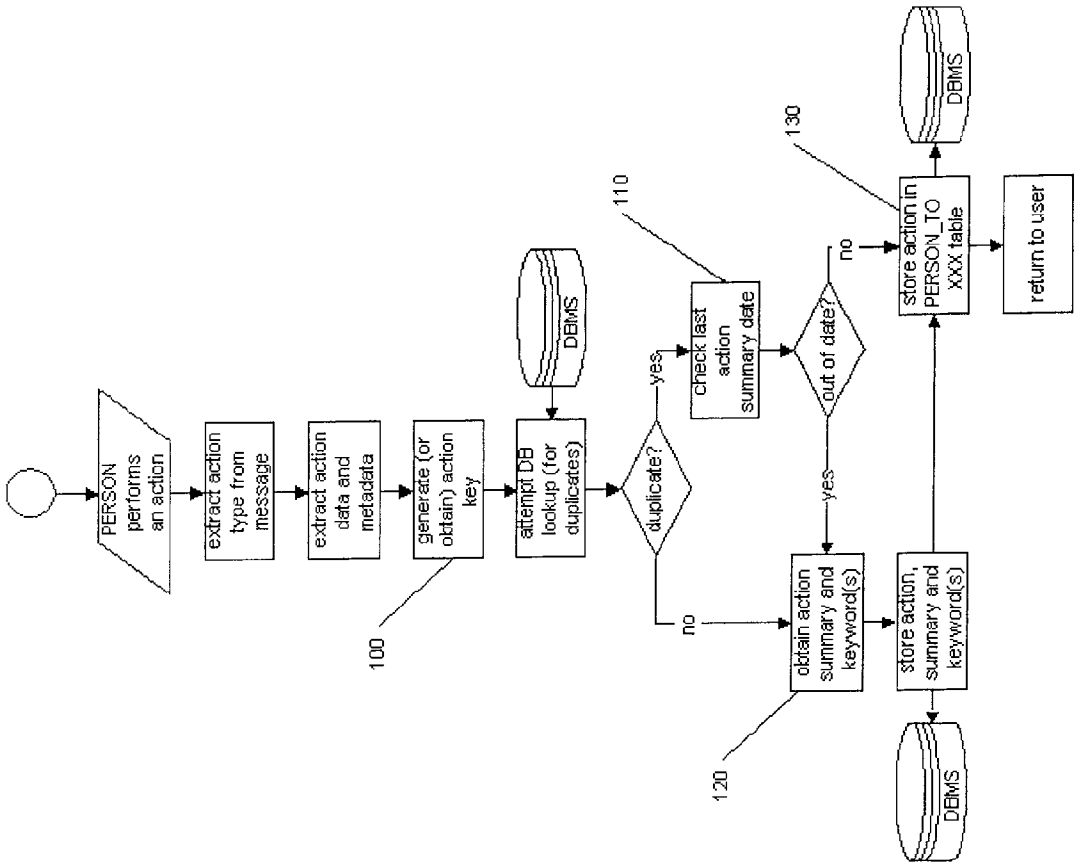


Figure 10

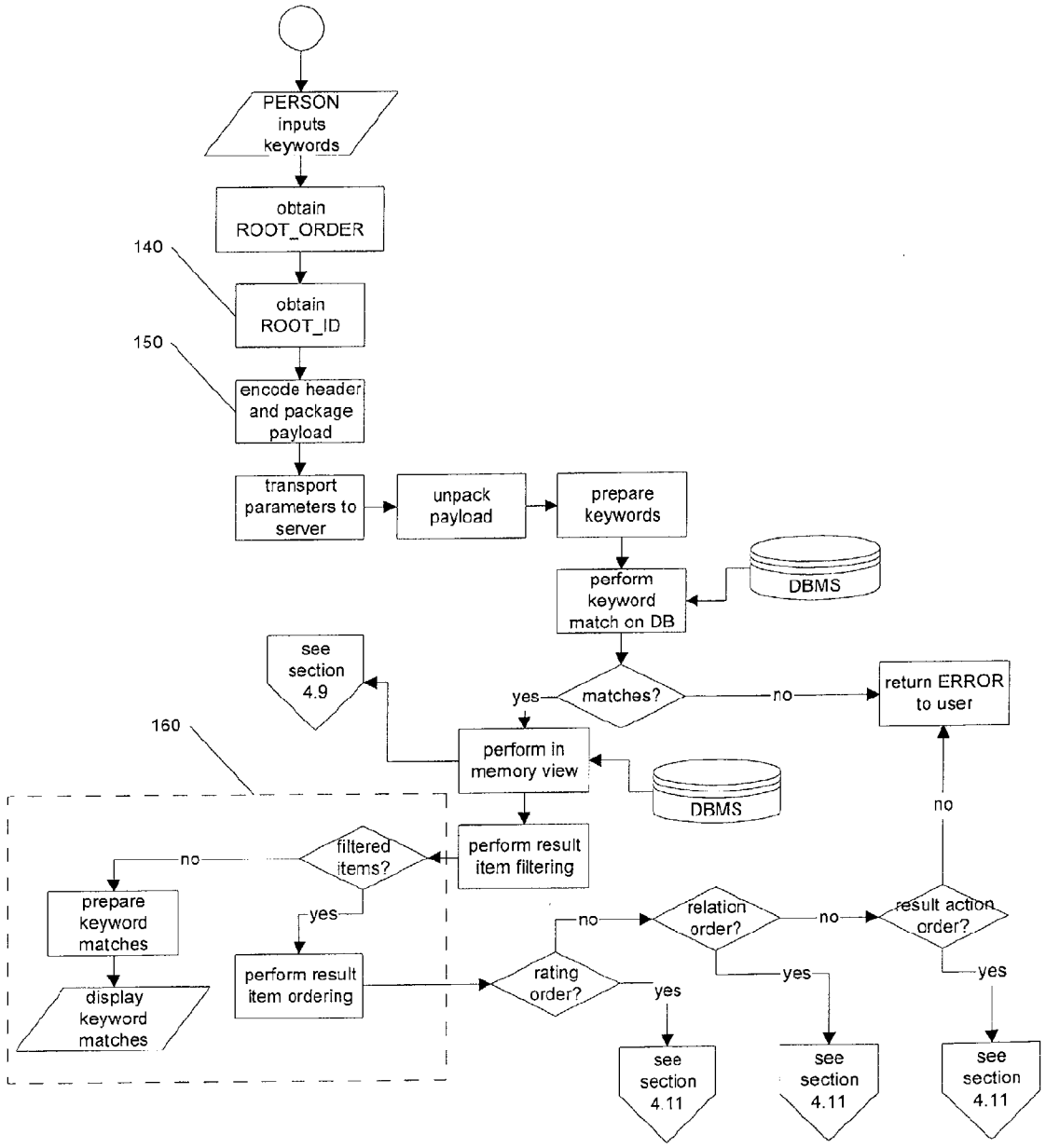


Figure 11

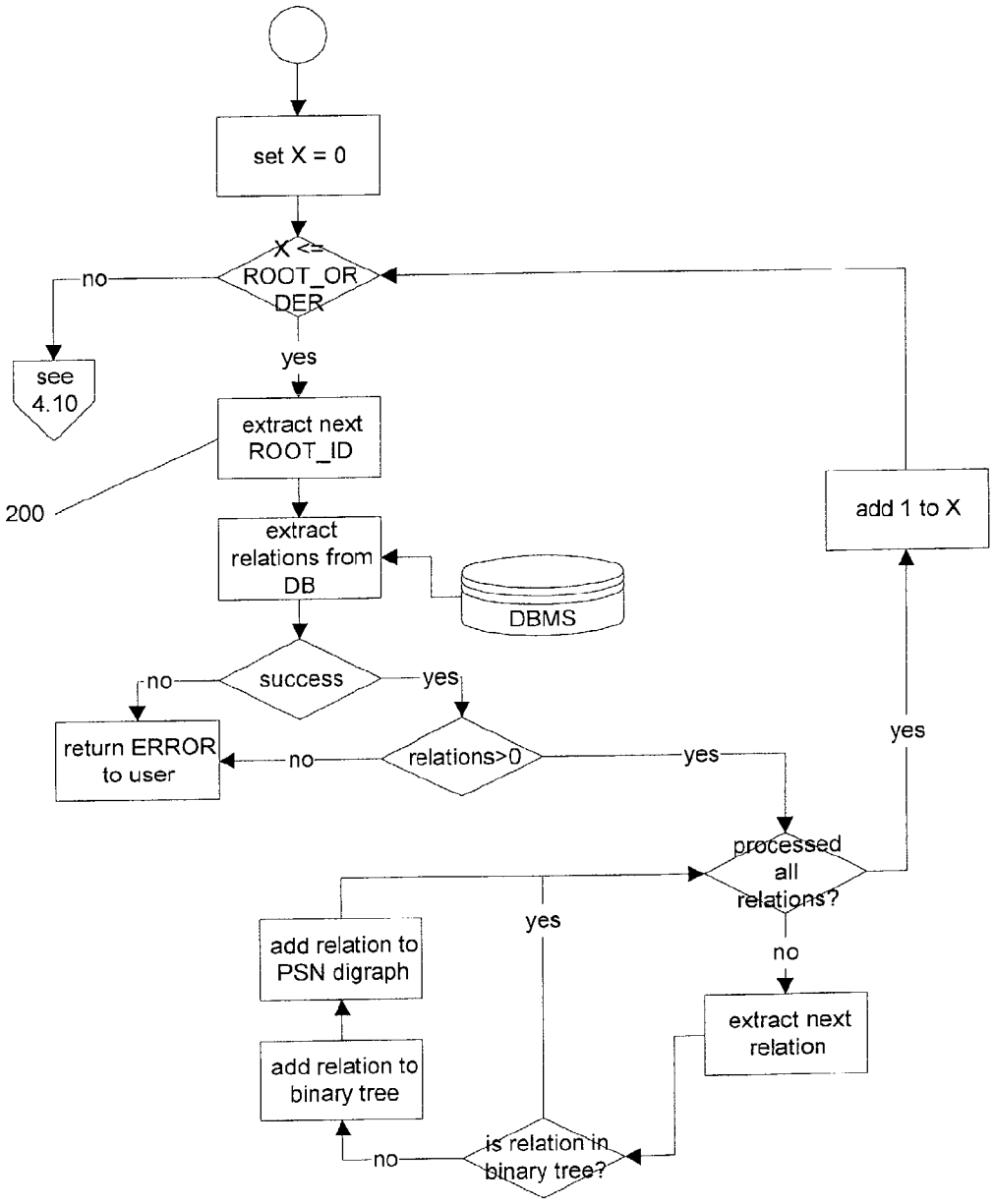


Figure 12

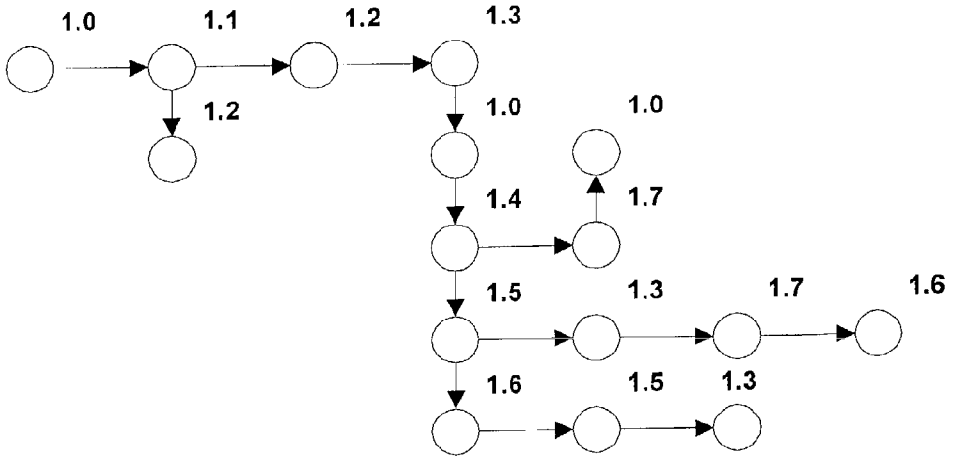


Figure 13

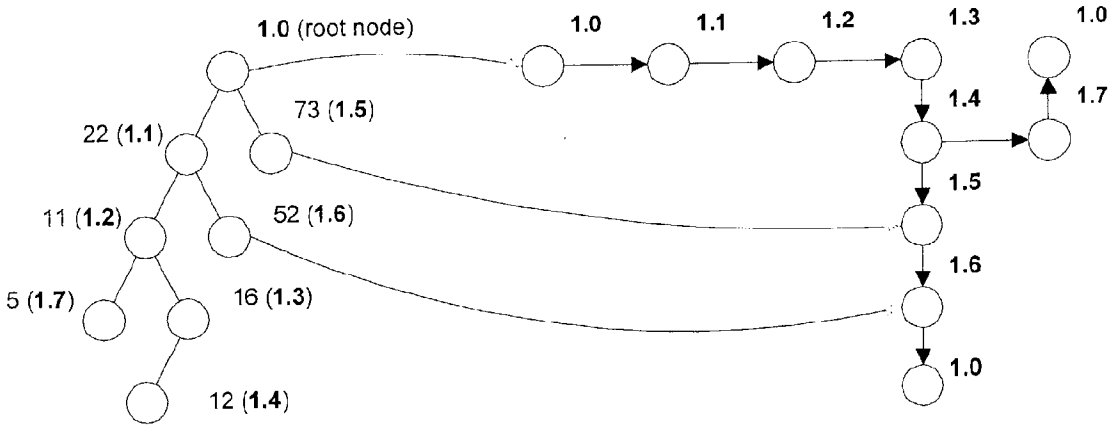


Figure 14

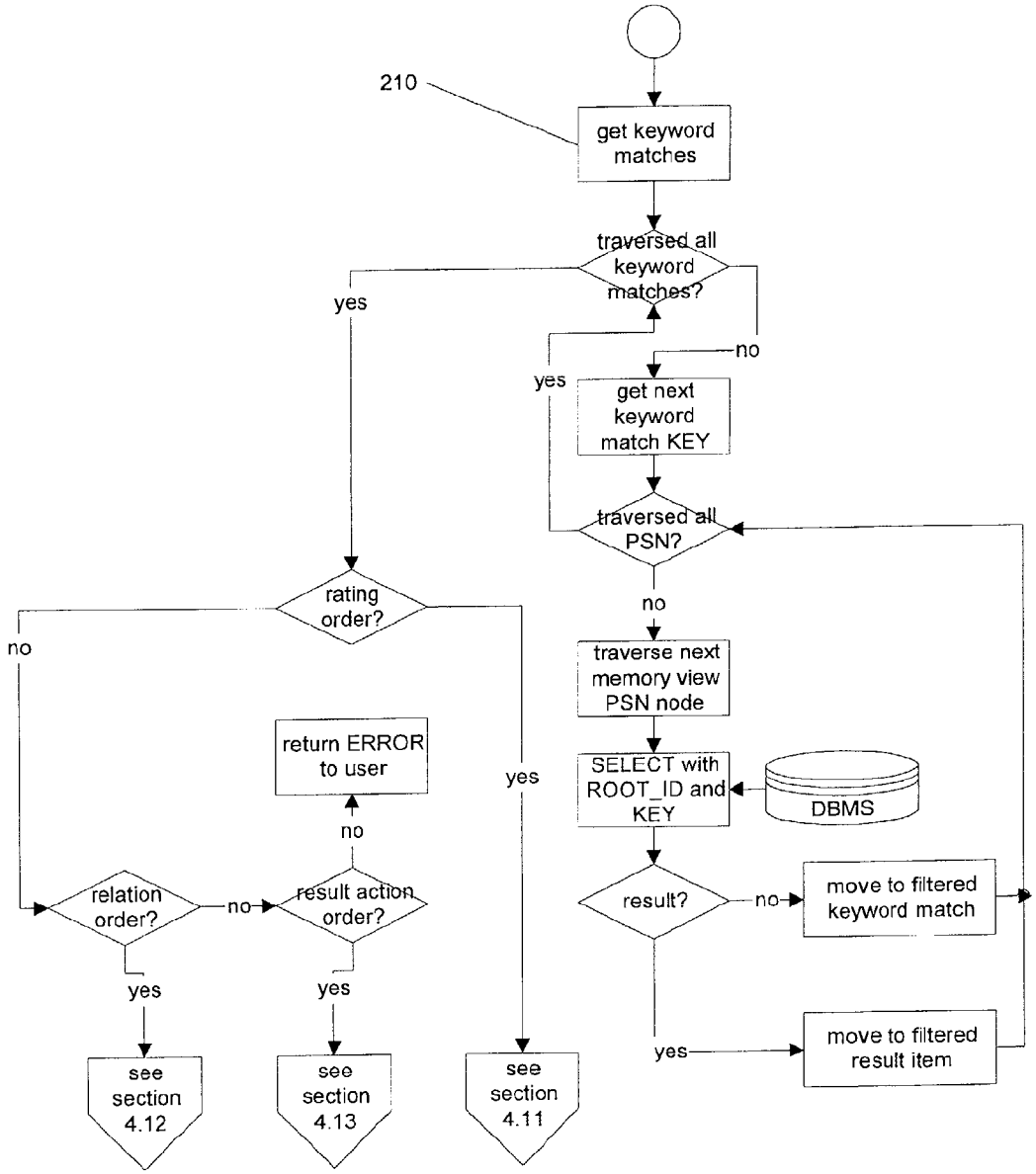
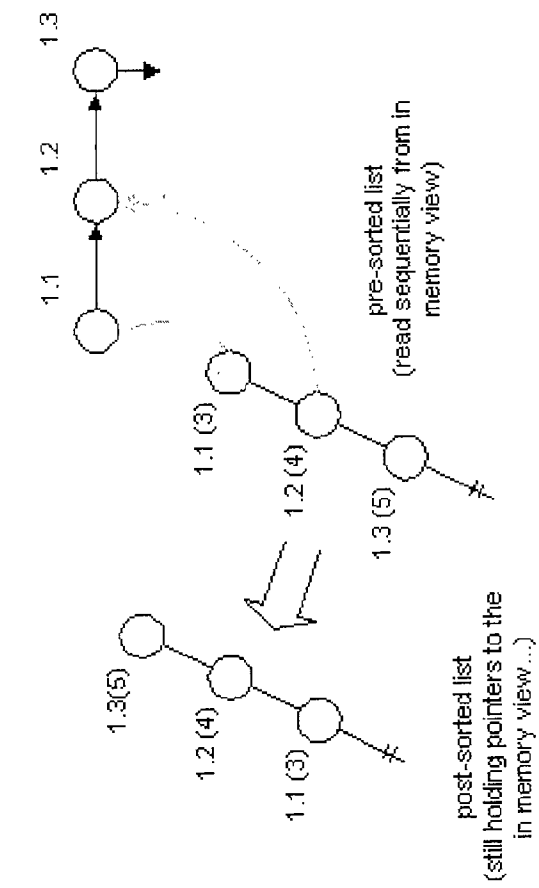


Figure 15



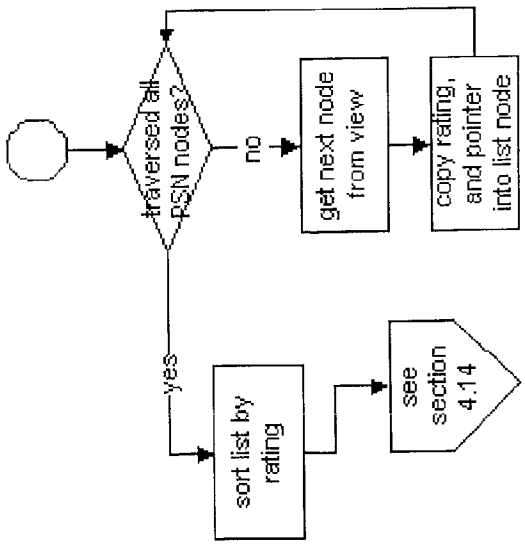


Figure 16

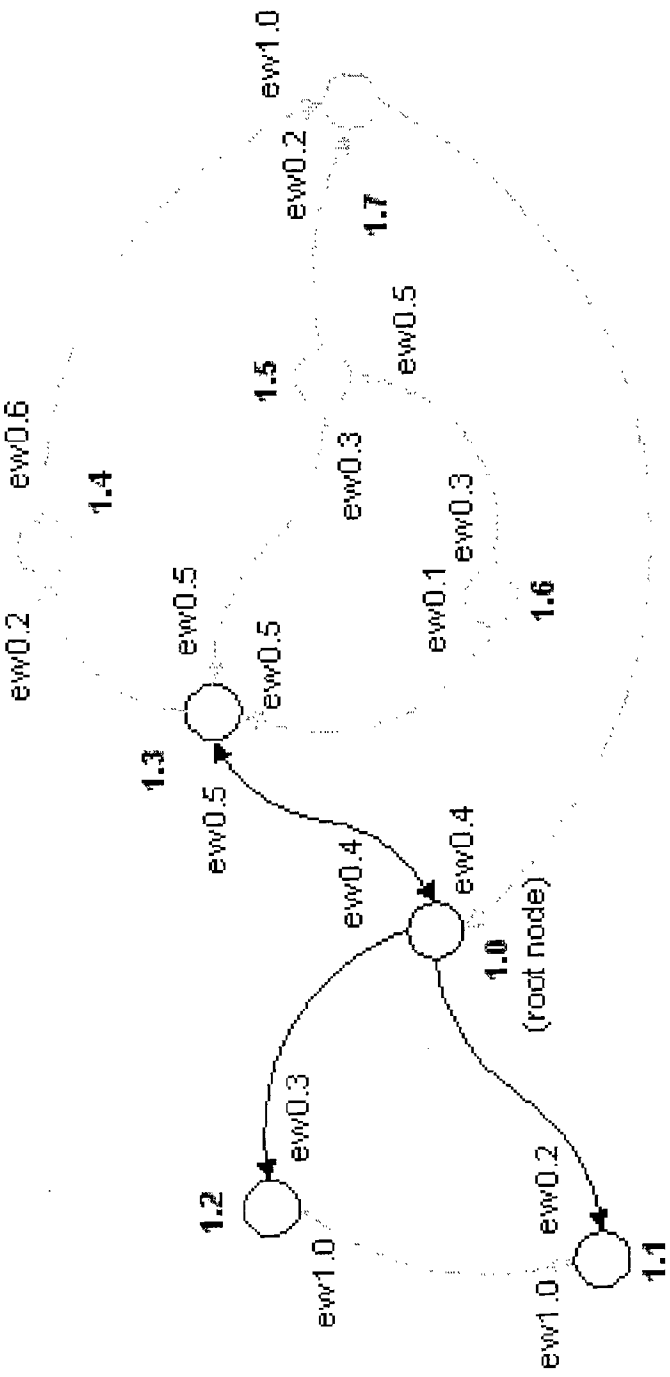


Figure 17

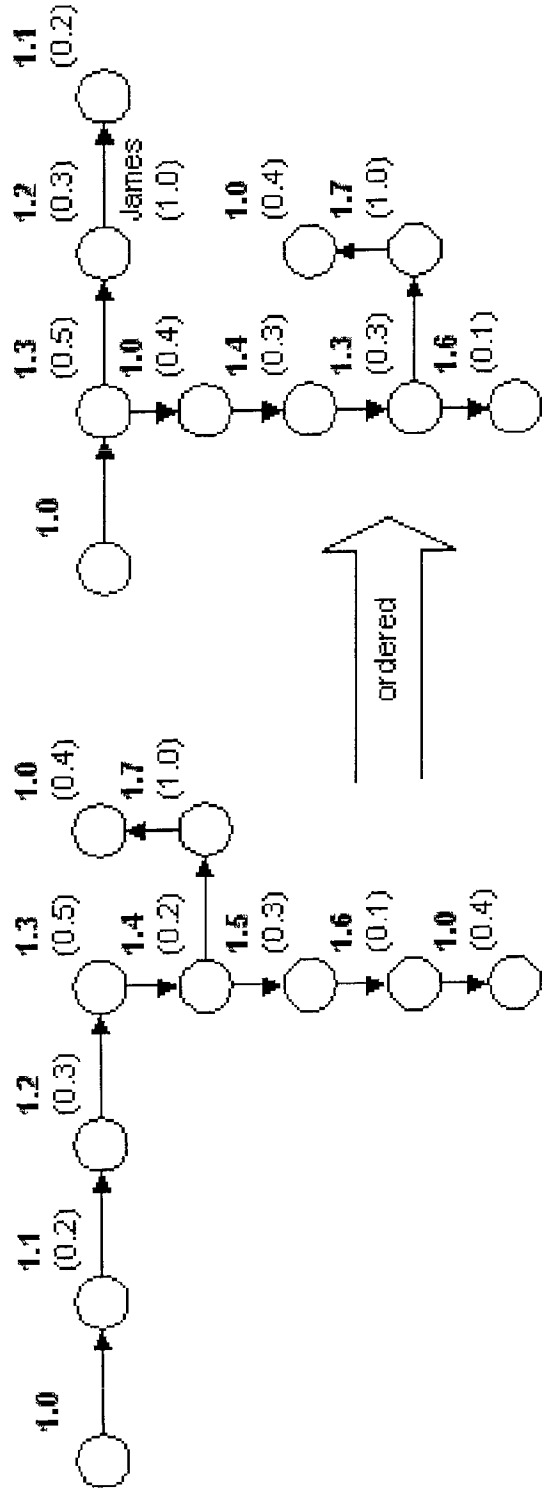


Figure 18

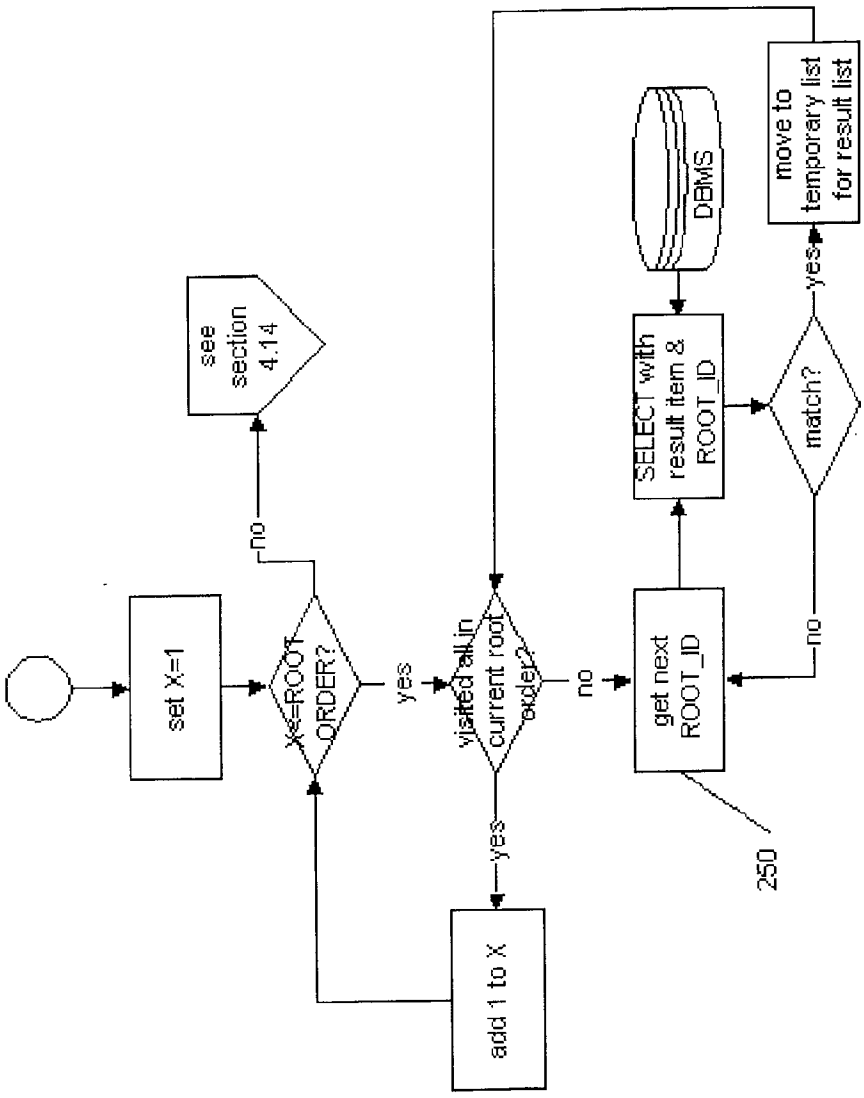

set X=1
X<=ROOT ORDER?
no
see section 4.14
yes
visited all in current root order?
yes
add 1 to X
no
get next ROOT_ID
250
SELECT with result item & ROOT_ID
DBMS
match?
yes
no
move to temporary list for result list

Figure 19

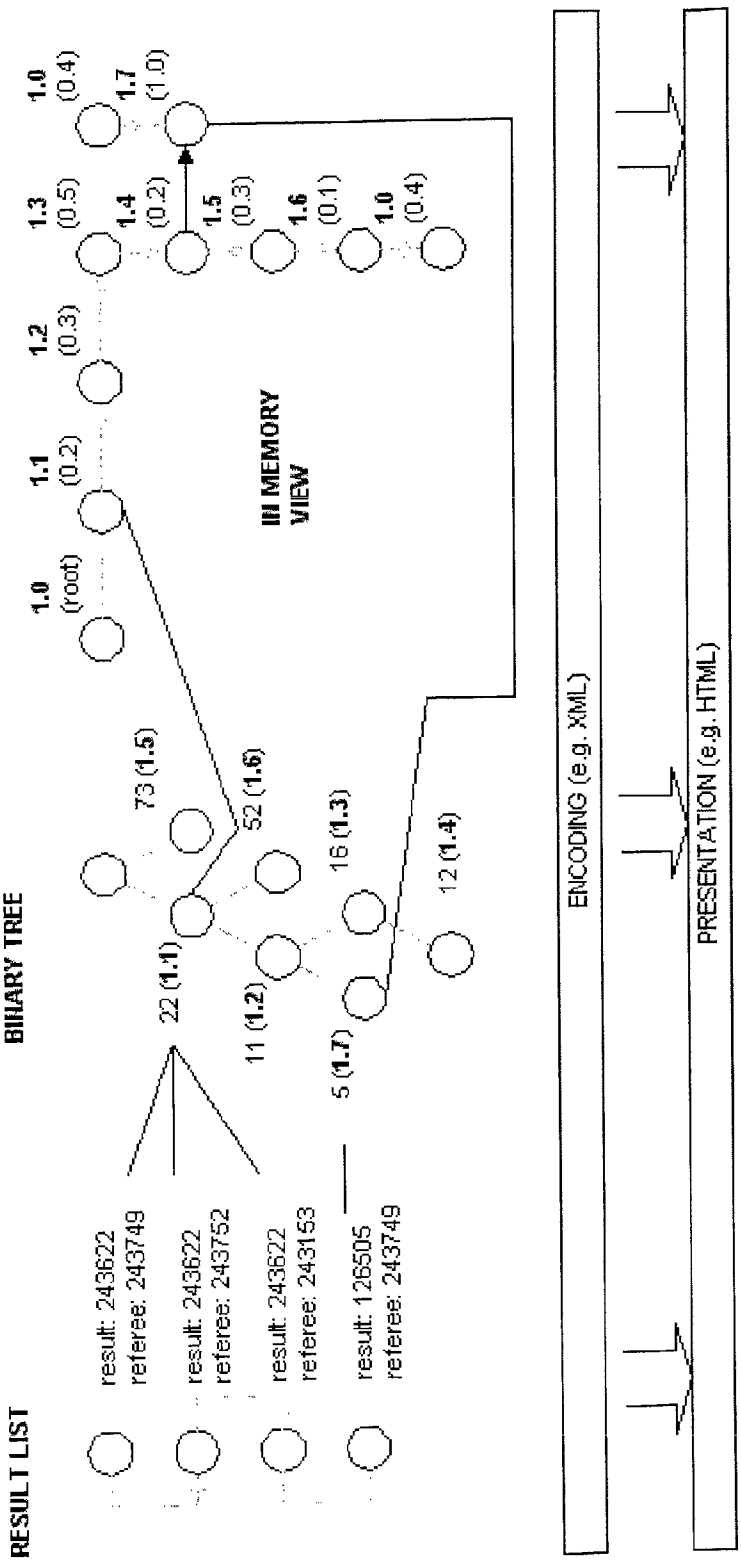


Figure 20

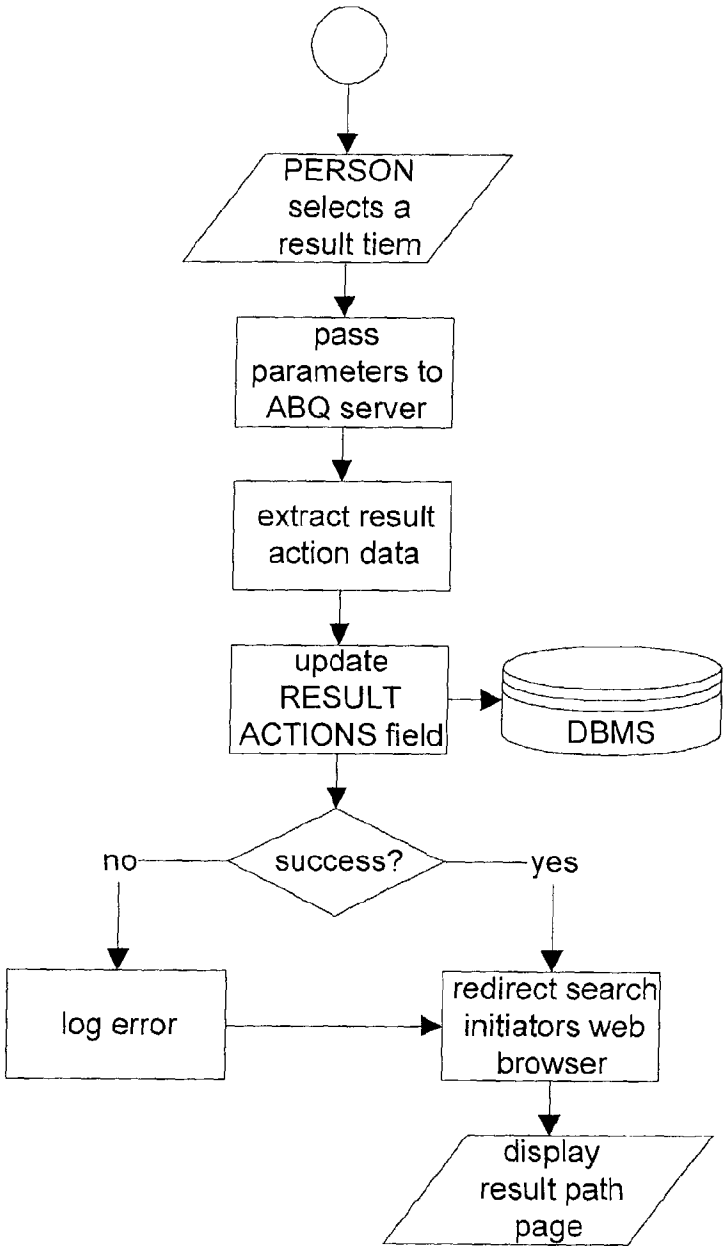

PERSON selects a result tiem
pass parameters to ABQ server
extract result action data
update RESULT ACTIONS field
DBMS
success?
no
yes
log error
redirect search initiators web browser
display result path page

Figure 21

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 01 30 7210

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	KAUTZ H ET AL: "The hidden web" AI MAGAZINE, SUMMER 1997, AMERICAN ASSOC. ARTIFICIAL INTELLIGENCE, USA, vol. 18, no. 2, pages 27-36, XP002202174 ISSN: 0738-4602 * abstract * * page 28, right-hand column, line 47 - page 29, left-hand column, line 14 * * page 31, right-hand column, line 4 - page 32, left-hand column, line 52 * * page 32, right-hand column, line 48 - page 33, left-hand column, line 3 *	1-23, 28-31	G06F17/30
Y	WO 00 04474 A (NET PERCEPTIONS INC) 27 January 2000 (2000-01-27) * abstract * * page 7, line 28 - page 9, line 15 * * page 10, line 1 - page 10, line 15 * * page 17, line 2 - page 17, line 23 * * page 25, line 8 - page 25, line 29 *	1-23, 28-31	
A	KUMAR R ET AL: "Recommendation systems: a probabilistic analysis" FOUNDATIONS OF COMPUTER SCIENCE, 1998. PROCEEDINGS. 39TH ANNUAL SYMPOSIUM ON PALO ALTO, CA, USA 8-11 NOV. 1998, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, & November 1998 (1998-11-08), pages 664-673, XP010318861 ISBN: 0-8186-9172-7 * page 664, left-hand column, line 18 - page 664, right-hand column, line 34 * * page 665, left-hand column, line 16 - page 665, left-hand column, line 50 * * page 666, right-hand column, line 11 - page 666, right-hand column, line 40 *	1-31	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G06F

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	13 June 2002	Boyadzhiev, Y

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 01 30 7210

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	EP 0 751 471 A (MASSACHUSETTS INST TECHNOLOGY) 2 January 1997 (1997-01-02) * abstract * * page 2, line 1 - page 2, line 42 * * page 3, line 32 - page 4, line 17 * * page 5, line 43 - page 6, line 21 *	1-31	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	13 June 2002	Boyadzhiev, Y

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 01 30 7210

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

13-06-2002

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0004474 A	27-01-2000	US 6334127 B1	25-12-2001
		AU 4973999 A	07-02-2000
		EP 1099177 A1	16-05-2001
		WO 0004474 A1	27-01-2000
EP 0751471 A	02-01-1997	AU 6282596 A	05-02-1997
		CA 2225790 A1	23-01-1997
		EP 0751471 A1	02-01-1997
		JP 11509019 T	03-08-1999
		US 6041311 A	21-03-2000
		US 6049777 A	11-04-2000
		US 6092049 A	18-07-2000
		US 6112186 A	29-08-2000
		WO 9702537 A1	23-01-1997
		US 5872850 A	16-02-1999

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

Electronic Patent Application Fee Transmittal

Application Number:	10685749
Filing Date:	14-Oct-2003
Title of Invention:	A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER--
First Named Inventor:	Edmund J. Fish
Filer:	Michael Glenn/kristianne serrano
Attorney Docket Number:	AOL0142

Filed as Large Entity

Utility Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
Processing Fee, except for Provis. apps	1808	1	130	130
Submission- Information Disclosure Stmt	1806	1	180	180
		Total in USD ($)		**310**

Electronic Acknowledgement Receipt

EFS ID:	1030095
Application Number:	10685749
Confirmation Number:	6161
Title of Invention:	A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER--
First Named Inventor:	Edmund J. Fish
Customer Number:	22862
Filer:	Michael Glenn/kristianne serrano
Filer Authorized By:	Michael Glenn
Attorney Docket Number:	AOL0142
Receipt Date:	24-APR-2006
Filing Date:	14-OCT-2003
Time Stamp:	19:38:50
Application Type:	Utility
International Application Number:	

Payment information:

Submitted with Payment	yes
Payment was successfully received in RAM	$310.0
RAM confirmation Number	360
Deposit Account	071445
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: Charge any Additional Fees required under 37 C.F.R. Section 1.16 and 1.17	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)	Multi Part	Pages
1		AOL0142IDS.pdf	293463	yes	8

Multipart Description

Doc Desc	Start	End
Transmittal to TC	1	1
Information Disclosure Statement (IDS) Filed	2	7
NPL Documents	8	8

Warnings:					
Information:					
2	NPL Documents	baudisch97profile1.pdf	2042630	no	7
Warnings:					
Information:					
3	NPL Documents	budzik00user1.pdf	2341489	no	8
Warnings:					
Information:					
4	NPL Documents	budzik99watson1.pdf	2722966	no	14
Warnings:					
Information:					
5	NPL Documents	freund98efficient1.pdf	3036260	no	18
Warnings:					
Information:					
6	NPL Documents	improvingcategoryspB396C1.pdf	2269316	no	9
Warnings:					
Information:					
7	NPL Documents	pazzani99framework1.pdf	2521909	no	16
Warnings:					
Information:					

8	NPL Documents	realtimeusercontext.pdf	842068	no	3
Warnings:					
Information:					
9	NPL Documents	shavlik98building1.pdf	2045521	no	6
Warnings:					
Information:					
10	NPL Documents	shavlik98intelligent1.pdf	2602851	no	8
Warnings:					
Information:					
11	NPL Documents	shavlik99instructable1.pdf	1347593	no	4
Warnings:					
Information:					
12	NPL Documents	watsonaninfrastructB3951.pdf	104530	no	6
Warnings:					
Information:					
13	Foreign Reference	ep1288795A1.pdf	1536029	no	42
Warnings:					
Information:					
14	Fee Worksheet (PTO-875)	fee-info.pdf	8429	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	23715054		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

E-FILING COVERSHEET

Application Serial No. 10/685,749 Attorney Docket No. AOL0142

I hereby certify that this correspondence is being ELECTRONICALLY FILED to the United States Patent and Trademark Office

From: GLENN PATENT GROUP
Customer No.: 22,862
Tel: (650) 474-8400
Fax: (650) 474-8401

on 24 April 2006
Date

Signature

Kristianne Serrano
Typed or printed name of person signing Certificate

Note: Each paper must have its own certificate of transmission, or this certificate must identify each submitted paper.

Attached to this cover-sheet please find the following documents:

- Cover-sheet (1 page);
- Information Disclosure Statement (2 pages in duplicate);
- SB/08B (3 pages); and
- Cited References.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Edmund J. Fish **Docket No.:** AOL0142

Serial No.: 10/685,749 **Art Unit:** 2141

Filed: 14 October 2003 **Examiner:** Nguyen, Quang N.

Title: A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER

23 May 2006

Commissioner for Patents
Mail Stop – Issue Fee
P.O. Box 1450
Alexandria, VA 22313-1450

PETITION UNDER 37 CFR § 1.48(a)

Applicant respectfully petitions the Commissioner to amend the inventive entity for the application indicated above to name the following additional inventor:

Name:	Bradley Chase Harrison
Residence:	275 Greenwich Street, Apt. 5A
	New York, NY 10007
Postal Address:	Same as residence
Citizenship:	United States of America

This petition is accompanied by:

1. A statement from the inventor Bradley Chase Harrison that the error by which he was not named as an inventor occurred without deceptive intention;
2. A Declaration under 37 CFR § 1.63 signed by the inventor Bradley Chase Harrison; and
3. Consent of the Assignee

05/24/2006 CNGUYEN1 00000001 071445 10685749
03 FC:1464 130.00 DA

The Commissioner is authorized to charge the fee of $130.00 for the Petition, any additional fees that may be due, and credit any overpayments to Deposit Account 07-1445, Glenn Patent Group.

Respectfully submitted,

Michael A. Glenn
Reg. No. 30,176

Customer No. 22862

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Edmund J. Fish
Serial No.: 10/685,749
Filed: 14 October 2003
Title: Search Enhancement System Having Personal Search Parameters

Docket No.: AOL0142
Art Unit: 2141
Examiner: Nguyen, Q. N.

19 May 2006

Commissioner for Patents
Mail Stop – Issue Fee
P. O. Box 1450
Alexandria, VA 22313-1450

STATEMENT OF BRADLEY CHASE HARRISON UNDER 37 CFR § 1.48(a)

I, BRADLEY CHASE HARRISON, residing at 275 Greenwich Street, Apt. 5A, New York, NY 10007, state the following:

1. That I am a citizen of United States of America;
2. That I am an inventor in the above-named application; and
3. That the error by which I was not named as an inventor occurred without any deceptive intention on my part.

BRADLEY CHASE HARRISON

5/20/2006
Date

Attorney Docket No. AOL0142

DECLARATION FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name;

I believe I am the original, first, and sole inventor (if only one name is listed below) or an original, first, and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

SEARCH ENHANCEMENT SYSTEM HAVING

PERSONAL SEARCH PARAMETERS

the specification of which (check one) ___ is attached hereto, or _X_ was filed on 10/14/2003 as Application Serial No. 10/685,749 and was amended on 01/24/2006 (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the patentability of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

==========

I hereby claim foreign priority benefits under Title 35, United Sates Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)			Priority Claimed Yes	No
Number	Country	Day/Month/Year Filed	___	___
Number	Country	Day/Month/Year Filed	___	___

==========

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith:

MICHAEL A. GLENN, Reg. No. 30,176
DONALD M. HENDRICKS, Reg. No. 40,355
CHRISTOPHER PEIL, Reg. No. 45,005
JEFFREY BRILL, 51,198
JULIA A. THOMAS, Reg. No. 52,283

SEND CORRESPONDENCE TO:

GLENN PATENT GROUP, 3475 Edison Way, Suite L, Menlo Park, CA 94025

==========

Attorney Docket No. AOL0142

I hereby claim the benefit under Title 35, United States code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Ser. No.	Filing Date	Status: Patented, Pending, Abandoned

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor: EDMUND J. FISH

Inventor's signature _____ Date _____

Residence 22265 Pacific Blvd. 63A:B07, Dulles, Virginia 20166

Post Office Address Same

Citizenship United States of America

Full name of second inventor: BRADLEY CRASE HARRISON

Inventor's signature _____ Date 5/20/2006

Residence 275 Greenwich Street, Apt. 5A, New York, NY, 10007

Post Office Address Same

Citizenship United States of America

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Edmund J. Fish **Docket No.:** AOL0142
Serial No.: 10/685,749 **Art Unit:** 2141
Filed: 14 October 2003 **Examiner:** Nguyen, Q. N.
Title: Search Enhancement System Having Personal Search Parameters

22 May 2006

Commissioner for Patents
Mail Stop – Issue Fee
P.O. Box 1450
Alexandria, VA 22313-1450

CONSENT OF ASSIGNEE TO AMEND INVENTIVE ENTITY UNDER 37 CFR § 1.48(a)

On behalf of AOL LLC ("the Company"), I hereby give consent to amend the inventive entity in the above-named patent application ("the Application") by naming Bradley Chase Harrison as an additional inventor.

I affirm that, as an officer of the Company, I am authorized to take such action on the Company's behalf.

An Assignment, by which the inventor, Edmund J. Fish, transferred one hundred per cent of his right, title, and interest in the Application to the Company, is provided herewith to establish the Company's ownership of the Application

RANDALL BOE
Executive Vice President and General Counsel

5-22-06
Date

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

22862 7390 02/23/2006

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission
I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

Della Revecho	(Depositor's name)
	(Signature)
May 23, 2006	(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

TITLE OF INVENTION: A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER—

APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	NO	$1400	$300	$1700	05/23/2006

EXAMINER	ART UNIT	CLASS-SUBCLASS
NGUYEN, QUANG N	2141	709-246000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).
☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. Use of a Customer Number is required.

2. For printing on the patent front page, list
(1) the names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.
1 Michael A. Glenn
2 Glenn Patent Group
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)
PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: America Online, Inc.
(B) RESIDENCE: (CITY and STATE OR COUNTRY): Dulles, Virginia

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. The following fee(s) are enclosed:
☒ Issue Fee
☒ Publication Fee (No small entity discount permitted)
☐ Advance Order - # of Copies _____

4b. Payment of Fee(s):
☐ A check in the amount of the fee(s) is enclosed.
☐ Payment by credit card. Form PTO-2038 is attached.
☒ The Director is hereby authorized by charge the required fee(s), or credit any overpayment, to Deposit Account Number 07-1445 (Glenn Patent Group) (enclose an extra copy of this form).

5. Change in Entity Status (from status indicated above)
☐ a. Applicant claims SMALL ENTITY status. See 37 CFR 1.27.
☐ b. Applicant is no longer claiming SMALL ENTITY status. See 37 CFR 1.27(g)(2).

The Director of the USPTO is requested to apply the Issue Fee and Publication Fee (if any) or to re-apply any previously paid issue fee to the application identified above.
NOTE: The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

Authorized Signature _____ Date May 23, 2006

Typed or printed name Michael A. Glenn Registration No. 30,176

This collection of information is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Edmund J. Fish **Docket No.:** AOL0142
Serial No.: 10/685,749 **Art Unit:** 2141
Filed: 14 October 2003 **Examiner:** Nguyen, Quang N.
Title: A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER

23 May 2006

Commissioner of Patents
Mail Stop - Issue Fee
P.O. Box 1450
Alexandria, VA 22313-1450

ISSUE FEE AUTHORIZATION

Examiner:

Enclosed are the following documents:

1. Certificate of Fax Transmission (1 sheet);
2. Issue Fee Authorization (1 page in duplicate); and
3. Part B – Issue Fee Transmittal (1 page in duplicate)

The Commissioner is authorized to charge the Issue Fee of $ 1,400.00, Publication Fee of $ 300.00, any additional fees that may be due, and credit any overpayments to Deposit Account No. 07-1445 (Order No. AOL0142). Attached is a duplicate copy of this sheet for accounting purposes.

Respectfully submitted,

MICHAEL A. GLENN
Reg. No. 30,176

Customer No. 22862

The Commissioner is authorized to charge the fee of $130.00 for the Petition, any additional fees that may be due, and credit any overpayments to Deposit Account 07-1445, Glenn Patent Group.

Respectfully submitted,

Michael A. Glenn
Reg. No. 30,176

Customer No. 22862

PTO/SB/97 (08-03)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Certificate of Transmission under 37 CFR 1.8

Application Serial No. 10/685,749 Attorney Docket No. AOL0142

I hereby certify that this correspondence is being facsimile transmitted to the United States Patent and Trademark Office

From: GLENN PATENT GROUP
Customer No.: 22,862
Tel: (650) 474-8400
Fax: (650) 474-8401

on 23 May 2006
Date

Signature

Della Revecho
Typed or printed name of person signing Certificate

Note: Each paper must have its own certificate of transmission, or this certificate must identify each submitted paper.

Attached to this cover-sheet please find the following documents:

1. Certificate of Fax Transmission (1 sheet);
2. Issue Fee Authorization (1 page in duplicate);
3. Part B – Issue Fee Transmittal (1 page in duplicate);
4. Petition Under 37 CFR 1.48 (2 pages in duplicate);
5. Statement from Inventor Harrison (1 page);
6. Declaration-Oath from Inventor Harrison (2 pages); and
7. Consent of Assignee (1 page)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

22862 7590 06/08/2006

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

EXAMINER	
NGUYEN, QUANG N	

ART UNIT	PAPER NUMBER
2141	

DATE MAILED: 06/08/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

UNITED STATES DEPARTMENT OF COMMERCE
U.S. Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450

APPLICATION NO./ CONTROL NO.	FILING DATE	FIRST NAMED INVENTOR / PATENT IN REEXAMINATION	ATTORNEY DOCKET NO.

EXAMINER	
ART UNIT	PAPER
	20060531

DATE MAILED:

Please find below and/or attached an Office communication concerning this application or proceeding.

Commissioner for Patents

Detailed Action

The information disclosure statements (IDS) submitted on 04/24/2006 was filed after the mailing date of the Notice of Allowance on 02/23/2006. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Quang N. Nguyen whose telephone number is (571) 272-3886.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's SPE, Rupal Dharia, can be reached at (571) 272-3880. The fax phone number for the organization is (571) 273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

RUPAL DHARIA
SUPERVISORY PATENT EXAMINER

PTO-90C (Rev.04-03)

Attorney Docket No. AOL0142

Form 1449 (Modified) **Information Disclosure Statement By Applicant** (Use Several Sheets if Necessary)	Atty. Docket No. AOL0142 Applicant: Fish Filing Date: October 14, 2003	Serial No.: 10/685,749 Group: 2141

U.S. Patent Documents

Examiner Initial	No.	Patent No.	Date	Patentee	Class	Sub-class	Filing Date
QN	A	5,862,325	1/19/1999	Reed, et al	395	200.31	9/27/1996
	B	5.974 412	10/26/1999	Hazlehurst, et al	707	3	9/24/1997
	C	6,018,738	1/25/2000	Breese, et al	700	100	1/22/1998
	D	6,144,964	11/07/2000	Breese, et al	707	10	1/22/1998
	E	6,256,664B1	7/03/2001	Donoho, et al	709	204	3/19/1999
	F	6,263,362B1	7/17/2001	Donoho, et al	709	207	7/09/1999
	G	6.289,353B1	9/11/2001	Hazlehurst, et al	707	102	6/10/1999
	H	6,314,420B1	11/06/2001	Lang, et al	707	3	12/03/1998
	I	6,345,264B1	2/05/2002	Breese, et al	706	21	1/22/1998
	J	6,345,288B1	2/05/2002	Reed, et al	709	201	5/15/2000
	K	6,353,813B1	3/05/2002	Breese, et al	706	12	1/22/1998
	L	6,356,936B1	3/12/2002	Donoho, et al	709	206	5/20/1999
	M	6,493,637B1	12/10/2002	Steeg	702	19	9/24/1999
	N	6,510,458B1	1/21/2003	Berstis, et al	709	219	7/15/1999
QN	O	6,539,392B1	3/25/2003	Rebane	707	101	3/29/2000

Foreign Patent or Published Foreign Patent Application

Examiner Initial	No.	Document No.	Publication Date	Country or Patent Office	Class	Sub-class	Translation Yes	Translation No

Other Documents

Examiner Initial	No.	Author, Title, Date, Place (e.g. Journal) of Publication
QN	P	FREUND, Y. et al.; *An Efficient Boosting Algorithm for Combining Preferences*; AT&T Labs, MIT Laboratory for Computer Science;
QN	Q	SHAVLIK, J. et al.; *Building Intelligent Agents for Web-Based Tasks: A Theory-Refinement Approach*; University of Wisconsin-Madison;
QN	R	SHAVLIK, J. et al.; *Intelligent Agents for Web-based Tasks: An Advice-Taking Approach*; University of Wisconsin-Madison;

/Quang N. Nguyen/ (05/31/2006)

[redacted]

QN	S	BAUDISCH, P.; *The Profile Editor: Designing a Direct Manipulative Tool for Assembling Profiles*; Institute for Integrated Information and Publication Systems IPSI, German National Research Center for Information Technology GMD, Germany;
	T	BUDZIK, J. et al.; *User Interactions with Everyday Applications as Context for Just-in-time Information Access*; Intelligent Information Laboratory, Northwestern University;
	U	BUDZIK, J. et al.; *Watson: Anticipating and Contextualizing Information Needs*; Northwestern University;
	V	GLOVER, E.J. et al.; *Improving Category Specific Web Search by Learning Query Modifications*; NEC Research Institute, Princeton, NJ, EECS Department, University of Michigan, Ann Arbor, MI, Information Sciences and Technology, Pennsylvania State University;
	W	PAZZANI, M.J. et al.; *A Framework for Collaborative, Content-Based and Demographic Filtering*; Department of Information and Computer Science, University of California, Irvine;
	X	BAUER, T. et al.; *Real Time User Context Modeling for Information Retrieval Agents*; Computer Science Department, Indiana University;
	Y	SHAVLIK, J. et al.; *An Instructable, Adaptive Interface for Discovery and Monitoring Information on the World-Wide Web*; University of Wisconsin-Madison;
	Z	BUDZIK, J. et al.; *Watson: An Infrastructure for Providing Task-Relevant, Just-In-Time Information*; Department of Computer Science, Northwestern University;
QN	AA	NAHL, D.; *Ethnography Of Novices' First Use Of Web Search Engines: Affective Control In Cognitive Processing*; Internet Reference Services Quarterly, vol. 3, no. 2, p. 51-72; 1998;

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

/Quang N. Nguyen/ (05/31/2006)

[redacted]

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	10/14/2003	Edmund J. Fish	AOL0142	6161

22862 7590 12/12/2006

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

EXAMINER	
NGUYEN, QUANG N	

ART UNIT	PAPER NUMBER
2141	

DATE MAILED: 12/12/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

UNITED STATES DEPARTMENT OF COMMERCE
U.S. Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450

APPLICATION NO./ CONTROL NO.	FILING DATE	FIRST NAMED INVENTOR / PATENT IN REEXAMINATION	ATTORNEY DOCKET NO.
10-685749			

EXAMINER	
ART UNIT	PAPER
	20061205

DATE MAILED:

Please find below and/or attached an Office communication concerning this application or proceeding.

Commissioner for Patents

1. This Office Action is in response to the Petition Under 37 CFR 1.48(a) filed on 05/23/2006.

Inventorship

2. In view of the papers filed 05/23/2006, it has been found that this nonprovisional application, as filed, through error and without deceptive intent, improperly set forth the inventorship, and accordingly, this application has been corrected in compliance with 37 CFR 1.48(a). The inventorship of this application has been changed by including Bradley Chase Harrison as an additional inventor.

The application will be forwarded to the Office of Initial Patent Examination (OIPE) for issuance of a corrected filing receipt, and correction of Office records to reflect the inventorship as corrected.

3. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Quang N. Nguyen whose telephone number is (571) 272-3886.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's SPE, Rupal Dharia, can be reached at (571) 272-3880. The fax phone number for the organization is (571) 273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Q.N.

Quang N. Nguyen
Patent Examiner
AU – 2141

RUPAL DHARIA
SUPERVISORY PATENT EXAMINER

PTO-90C (Rev.04-03)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,749	01/16/2007	7165119	AOL0142	6161

22862 7590 12/27/2006

GLENN PATENT GROUP
3475 EDISON WAY, SUITE L
MENLO PARK, CA 94025

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 280 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at (571)-272-4200.

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

Edmund J. Fish, Dulles, VA;

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Fish

Docket No.: AOL0142

Serial No: 10/685,749

Filed: October 14, 2003

U.S. Patent Number: 7,165,119

Date Issued: January 16, 2007

Examiner: Rupal Dharia

Art Unit: 2141

Title: Search Enhancement System and Method Having Rankings, Explicitly Specified by the User, Based Upon Applicability and Validity of Search Parameters in Regard to a Subject Matter

April 13, 2007

Assistant Commissioner for Patents
Mail Stop Certificate of Corrections
P.O. Box 1450
Alexandria, VA 22313-1450

Request for Certificate of Correction in Patent under 35 USC §254

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 322.

The correction requested involves mistakes made by the Patent Office. The Office has omitted one of the inventors listed Front Page, item *[75]* Inventor, "Bradley Chase Harrison." Please correct to include both inventors: Edmund J. Fish and Bradley Chase Harrison. A Petition under 37 CFR 1.48(a) was filed on May 23, 2006 to include Bradley Chase Harrison and was approved. The inventorship of this application has been corrected in compliance with 37 CFR 1.48(a) to include Bradley Chase Harrison as an additional inventor.

The patentee is entitled to correction of good-faith transcription of a clerical error where "the correction does not involve such changes in the patent as would constitute new matter or would require reexamination." 35 U.S.C. § 255. Therefore, no new matter is provided with this Certificate of Correction.

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 323, in duplicate, with at least one copy being suitable for printing.

Please send the Certificate to the undersigned at the address shown below.

The Commissioner is hereby authorized to charge the fee of $100 to Deposit Account 07-1445 (Order No. AOL0142). This paper is provided in duplicate.

Respectfully Submitted,

Michael A. Glenn
Reg. No. 30,176

Customer No. 22862

COPY

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Fish
Docket No.: AOL0142
Serial No: 10/685,749
Filed: October 14, 2003
U.S. Patent Number: 7,165,119
Date Issued: January 16, 2007
Examiner: Rupal Dharia
Art Unit: 2141

Title: Search Enhancement System and Method Having Rankings, Explicitly Specified by the User, Based Upon Applicability and Validity of Search Parameters in Regard to a Subject Matter

April 13, 2007

Assistant Commissioner for Patents
Mail Stop Certificate of Corrections
P.O. Box 1450
Alexandria, VA 22313-1450

Request for Certificate of Correction in Patent under 35 USC §254

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 322.

The correction requested involves mistakes made by the Patent Office. The Office has omitted one of the inventors listed Front Page, item *[75]* Inventor, "Bradley Chase Harrison." Please correct to include both inventors: Edmund J. Fish and Bradley Chase Harrison. A Petition under 37 CFR 1.48(a) was filed on May 23, 2006 to include Bradley Chase Harrison and was approved. The inventorship of this application has been corrected in compliance with 37 CFR 1.48(a) to include Bradley Chase Harrison as an additional inventor.

The patentee is entitled to correction of good-faith transcription of a clerical error where "the correction does not involve such changes in the patent as would constitute new matter or would require reexamination." 35 U.S.C. § 255. Therefore, no new matter is provided with this Certificate of Correction.

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 323, in duplicate, with at least one copy being suitable for printing.

Please send the Certificate to the undersigned at the address shown below.

COPY

The Commissioner is hereby authorized to charge the fee of $100 to Deposit Account 07-1445 (Order No. AOL0142). This paper is provided in duplicate.

Respectfully Submitted,

Michael A. Glenn
Reg. No. 30,176

Customer No. 22862

PTO/SB/44 (04-05)
Approved for use through 04/30/2007. OMB 0651-0033
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

(Also Form PTO-1050)

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 7,165,119 B2

APPLICATION NO.: 10/685,749

ISSUE DATE : January 16,2007

INVENTOR(S) : Fish

It is certified that an error appears or errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Front Page, Item [75] Inventor, replace, "Edmund J. Fish"

with: Edmund J. Fish, Dulles, VA (US)
Bradley Chase Harrison, New York, NY (US)

MAILING ADDRESS OF SENDER (Please do not use customer number below):

PTO/SB/44 (04-05)
Approved for use through 04/30/2007. OMB 0651-0033
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

(Also Form PTO-1050)

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 7,165,119 B2

APPLICATION NO.: 10/685,749

ISSUE DATE : January 16,2007

INVENTOR(S) : Fish

It is certified that an error appears or errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Front Page, Item [75] Inventor, replace, "Edmund J. Fish"

with: Edmund J. Fish, Dulles, VA (US)
Bradley Chase Harrison, New York, NY (US)

MAILING ADDRESS OF SENDER (Please do not use customer number below):

US007165119B2

(12) **United States Patent**
Fish

(10) **Patent No.: US 7,165,119 B2**
(45) **Date of Patent: Jan. 16, 2007**

(54) **SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER**

(75) Inventor: **Edmund J. Fish**, Dulles, VA (US)
BRADLEY CHASE HARRISON, NEW YORK, NY (US)
(73) Assignee: **America Online, Inc.**, Dulles, VA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 280 days.

(21) Appl. No.: **10/685,749**

(22) Filed: **Oct. 14, 2003**

(65) **Prior Publication Data**

US 2005/0097188 A1 May 5, 2005

(51) **Int. Cl.**
G06F 15/16 (2006.01)

(52) **U.S. Cl.** **709/246**; 709/203; 709/219; 709/228

(58) **Field of Classification Search** 709/203, 709/218, 219, 225; 707/3, 225
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,642,502	A	6/1997	Driscoll	395/606
5,848,396	A	12/1998	Gerace	705/10
5,862,325	A	1/1999	Reed et al.	395/200.31
5,974,412	A	10/1999	Hazlehurst et al.	707/3
5,983,214	A	11/1999	Lang et al.	707/1
6,018,738	A	1/2000	Breese et al.	700/100
6,029,161	A	2/2000	Lang et al.	707/1
6,029,195	A	2/2000	Herz	709/219
6,047,327	A	4/2000	Tso et al.	709/232
6,144,964	A	11/2000	Breese et al.	707/10
6,182,068	B1	1/2001	Culliss	707/5
6,256,639	B1 *	7/2001	Himmel et al.	709/219

(Continued)

FOREIGN PATENT DOCUMENTS

EP 1 072 982 1/2001

(Continued)

OTHER PUBLICATIONS

Pitkow, J. et al.; *Personalized Search;* Communications of the ACM, vol. 45, No. 9, p. 50-5; Sep. 2002.

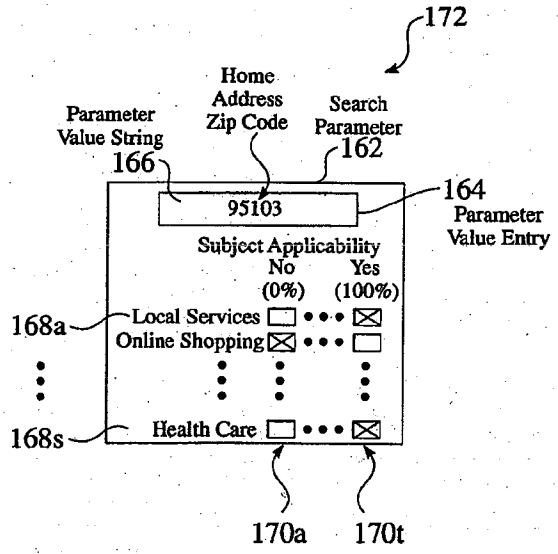
(Continued)

Primary Examiner—Rupal Dharia
Assistant Examiner—Quang N. Nguyen
(74) *Attorney, Agent, or Firm*—Michael A. Glenn; Glenn Patent Group

(57) **ABSTRACT**

The system and method comprises enhancement of results for a search engine, wherein the results from the search engine are refined or reorganized, based upon information from an identified secondary source. The results obtained using a conventional search are compared against the identified secondary source, e.g. a ratings service, and are filtered and/or sorted appropriately. In some embodiments, identification of the secondary source, such as a ratings service comprising information which may supplement the subject of a search query, is based upon information entered by the user. In alternate embodiments, the secondary source is associated with a user, as part of general user-specified search parameters, wherein one or more parameters are consulted automatically for searches for appropriate subject matter.

51 Claims, 15 Drawing Sheets



Electronic Patent Application Fee Transmittal

Application Number:	10685749
Filing Date:	14-Oct-2003
Title of Invention:	A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER--
First Named Inventor/Applicant Name:	Edmund J. Fish
Filer:	Michael Glenn/Vicky Smith
Attorney Docket Number:	AOL0142

Filed as Large Entity

Utility Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Certificate of correction	1811	1	100	100
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
			Total in USD ($)	**100**

Electronic Acknowledgement Receipt

EFS ID:	1683081
Application Number:	10685749
International Application Number:	
Confirmation Number:	6161
Title of Invention:	A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER--
First Named Inventor/Applicant Name:	Edmund J. Fish
Customer Number:	22862
Filer:	Michael Glenn/Vicky Smith
Filer Authorized By:	Michael Glenn
Attorney Docket Number:	AOL0142
Receipt Date:	13-APR-2007
Filing Date:	14-OCT-2003
Time Stamp:	18:26:18
Application Type:	Utility

Payment information:

Submitted with Payment	yes
Payment was successfully received in RAM	$100
RAM confirmation Number	1057
Deposit Account	071445
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: Charge any Additional Fees required under 37 C.F.R. Section 1.16 and 1.17	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)	Multi Part /.zip	Pages (if appl.)
1		certificateofcorrection41307AOL0142.pdf	305936	yes	8

	Multipart Description/PDF files in .zip description		
	Document Description	Start	End
	Miscellaneous Incoming Letter	1	1
	Request for Certificate of Correction	2	8

Warnings:

Information:

2	Fee Worksheet (PTO-06)	fee-info.pdf	8269	no	2

Warnings:

Information:

Total Files Size (in bytes):	314205

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

E-FILING COVERSHEET

Patent No. 7,165,119 B2

Attorney Docket No. AOL0142

I hereby certify that this correspondence is being ELECTRONICALLY FILED to the United States Patent and Trademark Office

From: GLENN PATENT GROUP
Customer No.: 22,862
Tel: (650) 474-8400
Fax: (650) 474-8401

on April 13, 2007.
Date

Vicky S. Smith
Signature

Vicky S. Smith
Typed or printed name of person signing Certificate

Note: Each paper must have its own certificate of transmission, or this certificate must identify each submitted paper.

Included in this Electronic Filing please find the following documents:

- E-filing Cover-sheet (1 page);
- Request for Certificate of Correction (2 pages, in duplicate);
- Certificate of Correction (1 page, in duplicate); and
- Page one of the Patent - (1 page).

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Fish

Docket No.: AOL0142

Serial No: 10/685,749

Filed: October 14, 2003

U.S. Patent Number: 7,165,119

Date Issued: January 16, 2007

Examiner: Rupal Dharia

Art Unit: 2141

Title: Search Enhancement System and Method Having Rankings, Explicitly Specified by the User, Based Upon Applicability and Validity of Search Parameters in Regard to a Subject Matter

April 13, 2007

Assistant Commissioner for Patents
Mail Stop Certificate of Corrections
P.O. Box 1450
Alexandria, VA 22313-1450

Request for Certificate of Correction in Patent under 35 USC §254

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 322.

The correction requested involves mistakes made by the Patent Office. The Office has omitted one of the inventors listed Front Page, item *[75]* Inventor, "Bradley Chase Harrison." Please correct to include both inventors: Edmund J. Fish and Bradley Chase Harrison. A Petition under 37 CFR 1.48(a) was filed on May 23, 2006 to include Bradley Chase Harrison and was approved. The inventorship of this application has been corrected in compliance with 37 CFR 1.48(a) to include Bradley Chase Harrison as an additional inventor.

The patentee is entitled to correction of good-faith transcription of a clerical error where "the correction does not involve such changes in the patent as would constitute new matter or would require reexamination." 35 U.S.C. § 255. Therefore, no new matter is provided with this Certificate of Correction.

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 323, in duplicate, with at least one copy being suitable for printing.

Please send the Certificate to the undersigned at the address shown below.

The Commissioner is hereby authorized to charge the fee of $100 to Deposit Account 07-1445 (Order No. AOL0142). This paper is provided in duplicate.

Respectfully Submitted,

Michael A. Glenn
Reg. No. 30,176

Customer No. 22862

COPY

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Fish
Docket No.: AOL0142
Serial No: 10/685,749
Filed: October 14, 2003
U.S. Patent Number: 7,165,119
Date Issued: January 16, 2007
Examiner: Rupal Dharia
Art Unit: 2141

Title: Search Enhancement System and Method Having Rankings, Explicitly Specified by the User, Based Upon Applicability and Validity of Search Parameters in Regard to a Subject Matter

April 13, 2007

Assistant Commissioner for Patents
Mail Stop Certificate of Corrections
P.O. Box 1450
Alexandria, VA 22313-1450

Request for Certificate of Correction in Patent under 35 USC §254

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 322.

The correction requested involves mistakes made by the Patent Office. The Office has omitted one of the inventors listed Front Page, item *[75]* Inventor, "Bradley Chase Harrison." Please correct to include both inventors: Edmund J. Fish and Bradley Chase Harrison. A Petition under 37 CFR 1.48(a) was filed on May 23, 2006 to include Bradley Chase Harrison and was approved. The inventorship of this application has been corrected in compliance with 37 CFR 1.48(a) to include Bradley Chase Harrison as an additional inventor.

The patentee is entitled to correction of good-faith transcription of a clerical error where "the correction does not involve such changes in the patent as would constitute new matter or would require reexamination." 35 U.S.C. § 255. Therefore, no new matter is provided with this Certificate of Correction.

The enclosed Certificate of Correction (PTO Form 1050) for the above-identified patent is submitted under Rule 323, in duplicate, with at least one copy being suitable for printing.

Please send the Certificate to the undersigned at the address shown below.

COPY

The Commissioner is hereby authorized to charge the fee of $100 to Deposit Account 07-1445 (Order No. AOL0142). This paper is provided in duplicate.

Respectfully Submitted,

Michael A. Glenn
Reg. No. 30,176

Customer No. 22862

PTO/SB/44 (04-05)
Approved for use through 04/30/2007. OMB 0651-0033
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

(Also Form PTO-1050)

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 7,165,119 B2

APPLICATION NO.: 10/685,749

ISSUE DATE : January 16,2007

INVENTOR(S) : Fish

It is certified that an error appears or errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Front Page, Item [75] Inventor, replace, "Edmund J. Fish"

with: Edmund J. Fish, Dulles, VA (US)
Bradley Chase Harrison, New York, NY (US)

MAILING ADDRESS OF SENDER (Please do not use customer number below):

Glenn Patent Group
3475 Edison Way, Suite L
Menlo Park, CA 94025

PTO/SB/44 (04-05)
Approved for use through 04/30/2007. OMB 0651-0033
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

(Also Form PTO-1050)

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 7,165,119 B2

APPLICATION NO.: 10/685,749

ISSUE DATE : January 16,2007

INVENTOR(S) : Fish

It is certified that an error appears or errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Front Page, Item [75] Inventor, replace, "Edmund J. Fish"

with: Edmund J. Fish, Dulles, VA (US)
Bradley Chase Harrison, New York, NY (US)

MAILING ADDRESS OF SENDER (Please do not use customer number below):

Glenn Patent Group
3475 Edison Way, Suite L
Menlo Park, CA 94025

US007165119B2

(12) **United States Patent**
Fish

(10) **Patent No.: US 7,165,119 B2**
(45) **Date of Patent: Jan. 16, 2007**

(54) **SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER**

(75) Inventor: **Edmund J. Fish**, Dulles, VA (US)
BRADLEY CHASE HARRISON, NEW YORK, NY (US)
(73) Assignee: **America Online, Inc.**, Dulles, VA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 280 days.

(21) Appl. No.: **10/685,749**

(22) Filed: **Oct. 14, 2003**

(65) **Prior Publication Data**

US 2005/0097188 A1 May 5, 2005

(51) **Int. Cl.**
G06F 15/16 (2006.01)

(52) **U.S. Cl.** **709/246**; 709/203; 709/219; 709/228

(58) **Field of Classification Search** 709/203, 709/218, 219, 225; 707/3, 225
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,642,502	A	6/1997	Driscoll	395/606
5,848,396	A	12/1998	Gerace	705/10
5,862,325	A	1/1999	Reed et al.	395/200.31
5,974,412	A	10/1999	Hazlehurst et al.	707/3
5,983,214	A	11/1999	Lang et al.	707/1
6,018,738	A	1/2000	Breese et al.	700/100
6,029,161	A	2/2000	Lang et al.	707/1
6,029,195	A	2/2000	Herz	709/219
6,047,327	A	4/2000	Tso et al.	709/232
6,144,964	A	11/2000	Breese et al.	707/10
6,182,068	B1	1/2001	Culliss	707/5
6,256,639	B1 *	7/2001	Himmel et al.	709/219

(Continued)

FOREIGN PATENT DOCUMENTS

EP 1 072 982 1/2001

(Continued)

OTHER PUBLICATIONS

Pitkow, J. et al.; *Personalized Search;* Communications of the ACM, vol. 45, No. 9, p. 50-5; Sep. 2002.

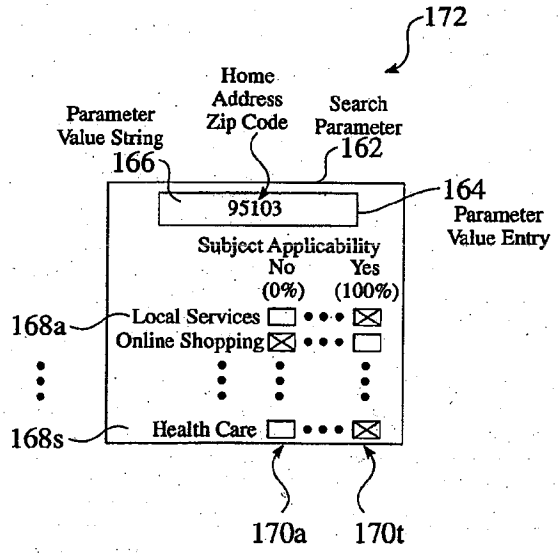
(Continued)

Primary Examiner—Rupal Dharia
Assistant Examiner—Quang N. Nguyen
(74) *Attorney, Agent, or Firm*—Michael A. Glenn; Glenn Patent Group

(57) **ABSTRACT**

The system and method comprises enhancement of results for a search engine, wherein the results from the search engine are refined or reorganized, based upon information from an identified secondary source. The results obtained using a conventional search are compared against the identified secondary source, e.g. a ratings service, and are filtered and/or sorted appropriately. In some embodiments, identification of the secondary source, such as a ratings service comprising information which may supplement the subject of a search query, is based upon information entered by the user. In alternate embodiments, the secondary source is associated with a user, as part of general user-specified search parameters, wherein one or more parameters are consulted automatically for searches for appropriate subject matter.

51 Claims, 15 Drawing Sheets



Electronic Acknowledgement Receipt

EFS ID:	1683180
Application Number:	10685749
International Application Number:	
Confirmation Number:	6161
Title of Invention:	A SEARCH ENHANCEMENT SYSTEM AND METHOD HAVING RANKINGS, EXPLICITLY SPECIFIED BY THE USER, BASED UPON APPLICABILITY AND VALIDITY OF SEARCH PARAMETERS IN REGARD TO A SUBJECT MATTER--
First Named Inventor/Applicant Name:	Edmund J. Fish
Customer Number:	22862
Filer:	Michael Glenn/Vicky Smith
Filer Authorized By:	Michael Glenn
Attorney Docket Number:	AOL0142
Receipt Date:	13-APR-2007
Filing Date:	14-OCT-2003
Time Stamp:	18:44:37
Application Type:	Utility

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)	Multi Part /.zip	Pages (if appl.)
1		certificateofcorrection41307AOL0142.pdf	306504	yes	8

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Miscellaneous Incoming Letter	1	1
	Request for Certificate of Correction	2	8
Warnings:			
Information:			
	Total Files Size (in bytes):	306504	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

E-FILING COVERSHEET

Patent No. 7,165,119 B2

Attorney Docket No. AOL0142

I hereby certify that this correspondence is being ELECTRONICALLY FILED to the United States Patent and Trademark Office

From: GLENN PATENT GROUP
Customer No.: 22,862
Tel: (650) 474-8400
Fax: (650) 474-8401

on April 13, 2007.
Date

Vicky S. Smith
Signature

Vicky S. Smith
Typed or printed name of person signing Certificate

Note: Each paper must have its own certificate of transmission, or this certificate must identify each submitted paper.

Included in this Electronic Filing please find the following documents:

- E-filing Cover-sheet (1 page);
- Request for Certificate of Correction (2 pages, in duplicate);
- Certificate of Correction (1 page, in duplicate); and
- Page one of the Patent - (1 page).

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,165,119 B2
APPLICATION NO. : 10/685749
DATED : January 16, 2007
INVENTOR(S) : Fish

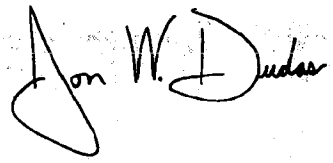
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Front Page, Item [75] Inventor, replace, "Edmund J. Fish"

with: --Edmund J. Fish, Dulles, VA (US)
Bradley Chase Harrison, New York, NY (US)--

Signed and Sealed this

Fifteenth Day of May, 2007



JON W. DUDAS
Director of the United States Patent and Trademark Office