IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

PERSONALIZED USER MODEL, L.L.P., Plaintiff, v. GOOGLE, INC., Defendant.	C.A. No. 09-525 (LPS)
GOOGLE, INC. Counterclaimant, v. PERSONALIZED USER MODEL, L.L.P. and YOCHAI KONIG, Counterclaim-Defendants.	**PUBLIC VERSION**

VOLUME 2 OF 3 - DECLARATION OF JAIME G. CARBONELL IN SUPPORT OF PLAINTIFF PERSONALIZED USER MODEL, L.L.P.'S ANSWERING BRIEF IN OPPOSITION TO GOOGLE'S MOTION FOR SUMMARY JUDGMENT ON INVALIDITY

OF COUNSEL:

Marc S. Friedman
SNR Denton US LLP
1221 Avenue of the Americas
New York, NY 10020-1089
(212) 768-6700

Mark C. Nelson
SNR Denton US LLP
2000 McKinney Avenue, Ste. 1900
Dallas, TX 75201
(214) 259-0901

Jennifer D. Bennett
SNR Denton US LLP
1530 Page Mill Road, Ste. 200
Palo Alto, CA 94304-1125
(650) 798-0300

MORRIS, NICHOLS, ARSHT & TUNNELL LLP
Karen Jacobs Louden (#2881)
Jeremy A. Tigan (#5239)
1201 N. Market Street
P.O. Box 1347
Wilmington, DE 19899-1347
(302) 658-9200
klouden@mnat.com

Attorneys for Personalized User Model, L.L.P.

Original Filing Date: January 14, 2013

Redacted Filing Date: January 23, 2013

Exhibit 29

Please type a plus (+) inside this box [+]

PTO/SB/05 (03-01)
Approved for use through 10/31/2002. OMB 0651-0032
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.	11378-0014-999	*Total Pages*	
First Named Inventor or Application Identifier			
Stephen R. Lawrence			
Express Mail Label No.	EV 313 841 855 US		

093003 U.S. PTO 10/067611 093003

APPLICATION ELEMENTS

See MPEP chapters 600 concerning utility patent application contents.

Address to: **Commissioner for Patents**
P.O. Box 1450
Mail Stop PATENT APPLICATION
Alexandria, VA 22313-1450

1. ☒ Fee Transmittal Form
(Submit an original, and a duplicate for fee processing)

2. ☐ Applicant claims Small Entity status, see 37 C.F.R. § 1.27

3. ☒ Specification [Total Pages 37]
(preferred arrangement set forth below, MPEP 1503.01)
- Descriptive title of the Invention
- Cross Reference to Related Applications
- Statement Regarding Fed sponsored R&D
- Reference to Microfiche Appendix
- Background of the Invention
- Brief Summary of the Invention
- Brief Description of the Drawings (if filed)
- Detailed Description of the Invention (including drawings, if filed)
- Claim(s)
- Abstract of the Disclosure

4. ☒ Drawing(s) *(35 USC 113)* [Total Sheets 12]

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) named 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 Readable Form (CRF)
b. ☐ Specification Sequence Listing on:
i. ☐ CD-ROM or CD-R (2 copies); or
ii. ☐ paper
c. ☐ Statement 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)* ☐ Power of Attorney
11. ☐ English Translation Document (if applicable)
12. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ 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)(i). Applicant must attach form PTO/SB/35 or it's equivalent
17. ☐ Other:

18. If a CONTINUING APPLICATION, *check appropriate box and supply the requisite information below and in a preliminary amendment, or in an Application Data Sheet under 37 CFR 1.76:*

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) | of prior application No.: filed .

Prior application information: Examiner: Group Art Unit:

For CONTINUATION OR 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

☒ Customer Number **24341** *Insert Customer No.* or ☐ Correspondence address below

NAME					
ADDRESS					
CITY		*STATE*		*ZIP CODE*	
COUNTRY		*TELEPHONE*		*FAX*	

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

Express Mail No. EV 313 841 855 US

PENNIE & EDMONDS LLP
COUNSELORS AT LAW
3300 Hillview Avenue
Palo Alto, California 94304

ATTORNEY DOCKET NO. 11378-0014-999 Date: September 30, 2003

Commissioner for Patents
P.O. Box 1450
Mail Stop PATENT APPLICATION
Alexandria, VA 22313-1450

Sir:

The following utility patent application is enclosed for filing:

Applicant(s): Stephen R. Lawrence Executed on: September 30, 2003

Title of Invention: PERSONALIZATION OF WEB SEARCH

PATENT APPLICATION FEE VALUE

TYPE	NO. FILED	LESS	EXTRA	EXTRA RATE	FEE
Total Claims	58	- 20	38	$18.00 each	$ 684
Independent	4	- 3	1	$84.00 each	$ 84
				Minimum Fee	$ 750
				Multiple Dependency Fee If Applicable ($280.00)	$ 0
				Total	$ 1518
				Applicant qualifies for the 50% Reduction for Independent Inventor, Nonprofit Organization or Small Business Concern	$ 0
				Total Filing Fee	$ 1518

[x] **DO NOT PUBLISH**. I hereby certify that the invention disclosed in the attached application **has not and will not be** the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen months after filing. I hereby request that the attached application not be published under 35 U.S.C. 122(b).

[ ] Priority of application no. filed on is claimed under 35 U.S.C. § 119.

[ ] The certified copy of the priority application has been filed in application no. filed on .

[ ] Amend the specification by inserting before the first line the following sentence: This is a continuation-in-part of application no. filed on .

Please charge the required fee to Pennie & Edmonds LLP Deposit Account 16-1150. A copy of this sheet is enclosed.

Respectfully submitted,

Gary S. Williams 31,066
PENNIE & EDMONDS LLP Reg No.

This form is not for use with continuation, divisional, re-issue, design or plant patent applications.

Personalization of Web Search

FIELD OF THE INVENTION

[0001] The present invention relates generally to the field of a search engine in a computer network system, in particular to system and method of creating a user profile for a user of a search engine and using the user profile to customize search results in response to search queries submitted by the user.

BACKGROUND OF THE INVENTION

[0002] Search engines provide a powerful source of indexed documents from the Internet (or an intranet) that can be rapidly scanned in response to a search query submitted by a user. Such a query is usually very short (on average about two to three words). As the number of documents accessible via the Internet grows, the number of documents that match the query may also increase. However, not every document matching the query is equally important from the user's perspective. As a result, a user is easily overwhelmed by an enormous number of documents returned by a search engine, if the engine does not order the search results based on their relevance to the user's query.

[0003] One approach to improving the relevance of search results to a search query is to use the link structure of different web pages to compute global "importance" scores that can be used to influence the ranking of search results. This is sometimes referred to as the PageRank algorithm. A more detailed description of the PageRank algorithm can be found in the article "The Anatomy of a Large-Scale Hypertextual Search Engine" by S. Brin and L. Page, 7th International World Wide Web Conference, Brisbane, Australia and U.S. Pat. No. 6,285,999, both of which are hereby incorporated by reference as background information.

[0004] An important assumption in the PageRank algorithm is that there is a "random surfer" who starts his web surfing journey at a randomly picked web page and keeps clicking on the links embedded in the web pages, never hitting the "back" button. Eventually, when this random surfer gets bored of the journey, he may re-start a new journey by randomly picking another web page. The probability that the random surfer visits (i.e., views or downloads) a web page depends on the web page's page rank.

[0005] From an end user's perspective, a search engine using the PageRank algorithm treats a search query the same way no matter who submits the query, because the search engine does not ask the user to provide any information that can uniquely identify the user. The only factor that affects the search results is the search query itself, e.g., how many terms are in the query and in what order. The search results are a best fit for the interest of an abstract user, the "random surfer", and they are not be adjusted to fit a specific user's preferences or interests.

[0006] In reality, a user like the random surfer never exists. Every user has his own preferences when he submits a query to a search engine. The quality of the search results returned by the engine has to be evaluated by its users' satisfaction. When a user's preferences can be well defined by the query itself, or when the user's preference is similar to the random surfer's preference with respect to a specific query, the user is more likely to be satisfied with the search results. However, if the user's preference is significantly biased by some personal factors that are not clearly reflected in a search query itself, or if the user's preference is quite different from the random user's preference, the search results from the same search engine may be less useful to the user, if not useless.

[0007] As suggested above, the journey of the random surfer tends to be random and neutral, without any obvious inclination towards a particular direction. When a search engine returns only a handful of search results that match a query, the order of the returned results is less significant because the requesting user may be able to afford the time to browse each of them to discover the items most relevant to himself. However, with billions of web pages connected to the Internet, a search engine often returns hundreds or even thousands of documents that match a search query. In this case, the ordering of the search results is very important. A user who has a preference different from that of the random surfer may not find what he is looking for in the first five to ten documents listed in the search results. When that happens, the user is usually left with two options: (1) either spending the time required to review more of the listed documents so as to locate the relevant documents; or (2) refining the search query so as to reduce the number of documents that match the query. Query refinement is often a non-trivial task, sometimes requiring more knowledge of the subject or more expertise with search engines than the user possesses, and sometimes requiring more time and effort than the user is willing to expend.

[0008] For example, assume that a user submits to a search engine a search query having only one term "blackberry". Without any other context, on the top of a list of documents returned by a PageRank-based search engine may be a link to www.blackberry.net, because this web page has the highest page rank. However, if the query requester is a person with interests in foods and cooking, it would be more useful to order the search results so as to include at the top of the returned results web pages with recipes or other food related text, pictures or the like. It would be desirable to have a search engine that is able to reorder its search results, or to otherwise customize the search results, so as to emphasize web pages that are most likely to be of interest to the person submitting the search query. Further, it would be desirable for such a system to require minimal input from individual users, operating largely or completely without explicit input from the user with regard to the user's preferences and interests. Finally, it would be desirable for such a system to meet users' requirements with respect to security and privacy.

SUMMARY

[0009] A search engine utilizes user profiles to customize search results. A user profile comprises multiple items that characterize a user's search preference. These items are extracted from various information sources, including previous search queries submitted by the user, links from or to the documents identified by the previous queries, sampled content from the identified documents as well as personal information implicitly or explicitly provided by the user.

[0010] When the search engine receives a search query from a user, it first identifies a set of documents that match the search query. Each document is associated with a generic rank based on the document's page rank, the text associated with the document, and the search query. The search engine also identifies the user's profile and correlates the user profile with each of the identified documents. The correlation between a document and the user profile produces a profile rank for the document, indicating the relevance of the document to the user. The search engine then combines the document's generic rank and profile rank into a personalized rank. Finally, the documents are ordered according to their personalized ranks.

[0011] In one embodiment, a user profile may comprise a plurality of sub-profiles, each sub-profile characterizing the user's interest from a different perspective. A term-based profile comprises a plurality of terms, each term carrying a weight indicative of its importance relative to other terms. A category-based profile comprises multiple categories, optionally organized into a hierarchical map. The user's search preferences may be associated with at least a subset of the multiple categories, each category having an associated weight indicating the user's interest in the documents falling into this category. There may be multiple category-based profiles for a user. In some embodiments, the sub-profiles include a link-based profile, which includes a plurality of links that are, directly or indirectly, related to identified documents, each link having a weight indicating the importance of the link. Links in the link-based profile may be further organized with respect to different hosts and domains.

[0012] The present invention, including user profile construction and search results re-ordering and/or scoring, can be implemented on either the client side or the server side of a client-server network environment.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The aforementioned features and advantages of the invention as well as additional features and advantages thereof will be more clearly understood hereinafter as a result of a detailed description of preferred embodiments of the invention when taken in conjunction with the drawings.

[0014] Fig. 1 illustrates a client-server network environment.

[0015] Fig. 2 illustrates multiple sources of user information and their relationship to a user profile.

[0016] Fig. 3 is an exemplary data structure that may be used for storing term-based profiles for a plurality of users.

[0017] Fig. 4A is an exemplary category map that may be used for classifying a user's past search experience.

[0018] Fig. 4B is an exemplary data structure that may be used for storing category-based profiles for a plurality of users.

[0019] Fig. 5 is an exemplary data structure that may be used for storing link-based profiles for a plurality of users.

[0020] Fig. 6 is a flowchart illustrating paragraph sampling.

[0021] Fig. 7A is a flowchart illustrating context analysis.

[0022] Fig. 7B depicts a process of identifying important terms using context analysis.

[0023] Fig. 8 illustrates a plurality of exemplary data structures that may be used for storing information about documents after term-based, category-based and/or link-based analyses, respectively.

[0024] Fig. 9A is a flowchart illustrating a personalized web search process according to one embodiment.

[0025] Fig. 9B is a flowchart illustrating a personalized web search process according to another embodiment.

[0026] Fig. 10 is a block diagram of a personalized search engine.

[0027] Like reference numerals refer to corresponding parts throughout the several views of the drawings.

DESCRIPTION OF EMBODIMENTS

[0028] The embodiments discussed below include systems and methods that create a user profile based a user's past experience with a search engine and then use the user profile to rank search results in response to search queries provided by the user.

[0029] Fig. 1 provides an overview of a typical client-server network environment 100 in which the present invention may be implemented. A plurality of clients 102 are connected to a search engine system 107 through a network 105, e.g., the Internet. Search

engine system 107 comprises one or more search engines 104. A search engine 104 is responsible for processing a search query submitted by a client 102, generating search results in accordance with the search query and returning the results to the client. Search engine system 107 may also comprise one or more content servers 106 and one or more user profile servers 108. A content server 106 stores a large number of indexed documents retrieved from different websites. Alternately, or in addition, the content server 106 stores an index of documents stored on various websites. In one embodiment, each indexed document is assigned a page rank according to the document's link structure. The page rank serves as a query independent measure of the document's importance. A search engine 104 communicates with one or more content servers 106 to select a plurality of documents in response to a specific search query. The search engine assigns a score to each document based on the document's page rank, the text associated with the document, and the search query.

[0030] A user profile server 108 stores a plurality of user profiles. Each profile includes information that uniquely identifies a user as well as his previous search experience and personal information, which can be used to refine search results in response to the search queries submitted by this user. Different approaches are available for user profile construction. For example, a user profile can be created by requiring a first-time user to fill in a form or answer a survey. This approach may be useful in certain applications such as opening a bank account. But it is hardly a favorable one in the context of a search engine. First, a user's interaction with a search engine is usually a dynamic process. As time goes on, the user's interests may change. This change may be reflected by the search queries submitted by the user, or by the user's handling of the search results, or both. The user's answers to questions on a form tend to become less useful over time, unless the user chooses to update his answers periodically. Unlike an occasional update of phone number in the case of an on-line bank account, frequent updates of a user profile in the case of a search engine significantly affect its user friendliness, which is an important consideration when a user chooses among the search engines currently available. Further, it is known that users are reluctant to provide explicit feedback, such as filling out of a form, as many users find it too burdensome. Thus, while some users may provide explicit feedback on their interests, it is desirable to have a procedure for implicitly obtaining information about the user's interests without requiring any explicit or new actions by the user.

[0031] It is has been observed that a search engine user's past search activities provide useful hints about the user's personal search preferences. Fig. 2 provides a list of sources of user information that are beneficial for user profile construction. For example, previously submitted search queries 201 are very helpful in profiling a user's interests. If a user has submitted multiple search queries related to diabetes, it is more likely than not that this is a topic of interest to the user. If the user subsequently submits a query including the term "organic food", it can be reasonably inferred that he may be more interested in those organic foods that are helpful in fighting diabetes. Similarly, the universal resource locators (URL) 203 associated with the search results in response to the previous search queries and their corresponding anchor texts 205, especially for search result items that have been selected or "visited" by the user (e.g., downloaded or otherwise viewed by the user), are helpful in determining the user's preferences. When a first page contains a link to a second page, and the link has text associated with it (e.g., text neighboring the link), the text associated with the link is called "anchor text" with respect to the second page. Anchor text establishes a relationship between the text associated with a URL link in a document and another document to which the URL link points. The advantages of anchor text include that it often provides an accurate description of the document to which the URL link points, and it can be used to index documents that cannot be indexed by a text-based search engine, such as images or databases.

[0032] After receiving search results, the user may click on some of the URL links, thereby downloading the documents referenced by those links, so as to learn more details about those documents. Certain types of general information 207 can be associated with a set of user selected or use identified documents. For purposes of forming a user profile, the identified documents from which information is derived for inclusion in the user profile may include: documents identified by search results from the search engine, documents accessed (e.g., viewed or downloaded, for example using a browser application) by the user (including documents not identified in prior search results), documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user, or any subset of such documents.

[0033] The general information 207 about the identified documents may answer questions such as, what is the format of the document? Is it in hypertext markup language

(HTML), plain text, portable document format (PDF), or Microsoft Word? What is the topic of the document? Is it about science, health or business? This information is also helpful in profiling the user's interests. In addition, information about a user's activities 209 with respect to the user selected documents (sometimes herein call the identified documents), such as how long the user spent viewing the document, the amount of scrolling activity on the document, and whether the user has printed, saved or bookmarked the document, also suggests the importance of the document to the user as well as the user's preferences. In some embodiments, information about user activities 209 is used both when weighting the importance of information extracted or derived from the user identified documents. In some embodiments, information about user activities 209 is used to determine which of the user identified documents to use as the basis for deriving the user profile. For example, information 209 may be used to select only documents that received significant user activity (in accordance with predefined criteria) for generating the user profile, or information 209 may be used to exclude from the profiling process documents that the user viewed for less than a predefined threshold amount of time.

[0034] Finally, the content of the identified documents from previous search activities is a rich source of information about a user's interests and preferences. Key terms appearing in the identified documents and their frequencies with which they appear in the identified documents are not only useful for indexing the document, but are also a strong indication of the user's personal interests, especially when they are combined with other types of user information discussed above. In one embodiment, instead of the whole documents, sampled content 211 from the identified documents is extracted for the purpose of user profile construction, to save storage space and computational cost. In another embodiment, various information related to the identified documents may be classified to constitute category information 213 about the identified documents. More discussion about content sampling, the process of identifying key terms in an identified document and the usage of the category information is provided below.

[0035] Optionally, a user may choose to offer personal information 215, including demographic and geographic information associated with the user, such as the user's age or age range, educational level or range, income level or range, language preferences, marital status, geographic location (e.g., the city, state and country in which the user resides, and

possibly also including additional information such as street address, zip code, and telephone area code), cultural background or preferences, or any subset of these. Compared with other types of personal information such as a user's favorite sports or movies that are often time varying, this personal information is more static and more difficult to infer from the user's search queries and search results, but maybe crucial in correctly interpreting certain queries submitted by the user. For example, if a user submits a query containing "Japanese restaurant", it is very likely that he may be searching for a local Japanese restaurant for dinner. Without knowing the user's geographical location, it is hard to order the search results so as to bring to the top those items that are most relevant to the user's true intention. In certain cases, however, it is possible to infer this information. For example, users often select results associated with a specific region corresponding to where they live.

[0036] Creating a user profile 230 from the various sources of user information is a dynamic and complex process. In some embodiments, the process is divided into sub-processes. Each sub-process produces one type of user profile characterizing a user's interests or preferences from a particular perspective. They are:

- a term-based profile 231 — this profile represents a user's search preferences with a plurality of terms, where each term is given a weight indicating the importance of the term to the user;

- a category-based profile 233 — this profile correlates a user's search preferences with a set of categories, which may be organized in a hierarchal fashion, with each category being given a weight indicating the extent of correlation between the user's search preferences and the category; and

- a link-based profile 235 — this profile identifies a plurality of links that are directly or indirectly related to the user's search preferences, with each link being given a weight indicating the relevance between the user's search preferences and the link.

[0037] In some embodiments, the user profile 230 includes only a subset of these profiles 231, 233, 235, for example just one or two of these profiles. In one embodiment, the user profile 230 includes a term-based profile 231 and a category-based profile 233, but not a link-based profile 235.

[0038] In one embodiment, a user profile is created and stored on a server (e.g., user profile server 108) associated with a search engine. The advantage of such deployment is that the user profile can be easily accessed by multiple computers, and that since the profile is stored on a server associated with (or part of) the search engine 104, it can be easily used by the search engine 104 to personalize the search results. In another embodiment, the user profile can be created and stored on the user's computer, sometimes called the client in a network environment. Creating and storing a user profile on a user's computer not only reduces the computational and storage cost for the search engine's servers, but also satisfies some users' privacy requirements. In yet another embodiment, the user profile may be created and updated on the client, but stored on a search engine server. Such embodiment combines some of the benefits illustrated in the other two embodiments. A disadvantage of this arrangement is that it may increase the network traffic between clients and the search engine servers. It is understood by a person of ordinary skill in the art that the user profiles of the present invention can be implemented using client computers, server computers, or both.

[0039] Fig. 3 illustrates an exemplary data structure, a term-based profile table 300, that may be used for storing term-based profiles for a plurality of users. Table 300 includes a plurality of records 310, each record corresponding to a user's term-based profile. A term-based profile record 310 includes a plurality of columns including a USER_ID column 320 and multiple columns of (TERM, WEIGHT) pairs 340. The USER_ID column stores a value that uniquely identifies a user or a group of users sharing the same set of (TERM, WEIGHT) pairs, and each (TERM, WEIGHT) pair 340 includes a term, typically 1-3 words long, that is usually important to the user or the group of users and a weight associated with the term that quantifies the importance of the term. In one embodiment, the term may be represented as one or more *n*-grams. An *n*-gram is defined as a sequence of *n* tokens, where the tokens may be words. For example, the phrase "search engine" is an *n*-gram of length 2, and the word "search" is an *n*-gram of length 1.

[0040] *N*-grams can be used to represent textual objects as vectors. This makes it possible to apply geometric, statistical and other mathematical techniques, which are well defined for vectors, but not for objects in general. In the present invention, *n*-grams can be used to define a similarity measure between two terms based on the application of a mathematical function to the vector representations of the terms.

[0041] The weight of a term is not necessarily a positive value. If a term has a negative weight, it may suggest that the user prefers that his search results should not include this term and the magnitude of the negative weight indicates the strength of the user's preference for avoiding this term in the search results. By way of example, for a group of surfing fans at Santa Cruz, California, the term-based profile may include terms like "surfing club", "surfing event" and "Santa Cruz" with positive weights. The terms like "Internet surfing" or "web surfing" may also be included in the profile. However, these terms are more likely to receive a negative weight since they are irrelevant and confusing with the authentic preference of the users sharing this term-based profile.

[0042] A term-based profile itemizes a user's preference using specific terms, each term having certain weight. If a document matches a term in a user's term-based profile, i.e., its content includes exactly this term, the term's weight will be assigned to the document; however, if a document does not match a term exactly, it will not receive any weight associated with this term. Such a requirement of relevance between a document and a user profile sometimes may be less flexible when dealing with various scenarios in which a fuzzy relevance between a user's preference and a document exists. For example, if a user's term-based profile includes terms like "Mozilla" and "browser", a document containing no such terms, but other terms like "Galeon" or "Opera" will not receive any weight because they do not match any existing term in the profile, even though they are actually Internet browsers. To address the need for matching a user's interests without exact term matching, a user's profile may include a category-based profile.

[0043] Fig. 4A illustrates a hierarchal category map 400 according to the Open Directory Project (http://dmoz.org/). Starting from the root level of map 400, documents are organized under several major topics, such as "Art", "News", "Sports", etc. These major topics are often too broad to delineate a user's specific interest. Therefore, they are further divided into sub- topics that are more specific. For example, topic "Art" may comprise sub-topics like "Movie", "Music" and "Literature" and the sub-topic "Music" may further comprise sub-sub-topics like "Lyrics", "News" and "Reviews". Note that each topic is associated with a unique CATEGORY_ID like 1.1 for "Art", 1.4.2.3 for "Talk Show" and 1.6.1 for "Basketball".

[0044] A user's specific interests may be associated with multiple categories at various levels, each of which may have a weight indicating the degree of relevance between the category and the user's interest. In one embodiment, a category-based profile may be implemented using a Hash table data structure as shown in Fig. 4B. A category-based profile table 450 includes a table 455 that comprises a plurality of records 460, each record including a USER_ID and a pointer pointing to another data structure, such as table 460-1. Table 460-1 may include two columns, CATEGORY_ID column 470 and WEIGHT column 480. CATEGORY_ID column 470 contains a category's identification number as shown in Fig. 4A, suggesting that this category is relevant to the user's interests and the value in the WEIGHT column 480 indicates the degree of relevance of the category to the user's interests.

[0045] A user profile based upon the category map 400 is a topic-oriented implementation. The items in a category-based profile can also be organized in other ways. In one embodiment, a user's preference can be categorized based on the formats of the documents identified by the user, such as HTML, plain text, PDF, Microsoft Word, etc. Different formats may have different weights. In another embodiment, a user's preference can be categorized according to the types of the identified documents, e.g., an organization's homepage, a person's homepage, a research paper, or a news group posting, each type having an associated weight. Another type category that can be used to characterize a user's search preferences is document origin, for instance the country associated with each document's host. In yet another embodiment, the above-identified category-based profiles may co-exist, with each one reflecting one aspect of a user's preferences.

[0046] Besides term-based and category-based profiles, another type of user profile is referred to as a link-based profile. As discussed above, the PageRank algorithm is based on the link structure that connects various documents over the Internet. A document that has more links pointing to it is often assigned a higher page rank and therefore attracts more attention from a search engine. Link information related to a document identified by a user can also be used to infer the user's preferences. In one embodiment, a list of preferred URLs are identified for a user by analyzing the frequency of his access to those URLs. Each preferred URL may be further weighted according to the time spent by the user and the user's scrolling activity at the URL, and/or other user activities (209, Fig. 2) when visiting the document at the URL. In another embodiment, a list of preferred hosts are identified for a

user by analyzing the user's frequency of accessing web pages of different hosts. When two preferred URLs are related to the same host the weights of the two URLs may be combined to determine a weight for the host. In another embodiment, a list of preferred domains are identified for a user by analyzing the user's frequency of accessing web pages of different domains. For example, for finance.yahoo.com, the host is "finance.yahoo.com" while the domain is "yahoo.com".

[0047] Fig. 5 illustrates a link-based profile using a Hash table data structure. A link-based profile table 500 includes a table 510 that includes a plurality of records 520, each record including a USER_ID and a pointer pointing to another data structure, such as table 510-1. Table 510-1 may include two columns, LINK_ID column 530 and WEIGHT column 540. The identification number stored in the LINK_ID column 530 may be associated with a preferred URL or host. The actual URL/host/domain may be stored in the table instead of the LINK_ID, however it is preferable to store the LINK_ID to save storage space.

[0048] A preferred list of URLs and/or hosts includes URLs and/or hosts that have been directly identified by the user. The preferred list of URLs and/or host may furthermore extend to URLs and/or hosts indirectly identified by using methods such as collaborative filtering or bibliometric analysis, which are known to persons of ordinary skill in the art. In one embodiment, the indirectly identified URLs and/or host include URLs or hosts that have links to/from the directly identified URLs and/or hosts. These indirectly identified URLs and/or hosts are weighted by the distance between them and the associated URLs or hosts that are directly identified by the user. For example, when a directly identified URL or host has a weight of 1, URLs or hosts that are one link away may have a weight of 0.5, URLs or hosts that are two links away may have a weight of 0.25, etc. This procedure can be further refined by reducing the weight of links that are not related to the topic of the original URL or host, e.g., links to copyright pages or web browser software that can be used to view the documents associated with the user selected URL or host. Irrelevant Links can be identified based on their context or their distribution. For example, copyright links often use specific terms (e.g., copyright or "All rights reserved" are commonly used terms in the anchor text of a copyright link); and links to a website from many unrelated websites may suggest that this website is not topically related (e.g., links to the Internet Explorer website are often included

in unrelated websites). The indirect links can also be classified according to a set of topics and links with very different topics may be excluded or be assigned a low weight.

[0049] The three types of user profiles discussed above are generally complimentary to one another since different profiles delineate a user's interests and preferences from different vantage points. However, this does not mean that one type of user profile, e.g., category-based profile, is incapable of playing a role that is typically played by another type of user profile. By way of example, a preferred URL or host in a link-based profile is often associated with a specific topic, e.g., finance.yahoo.com is a URL focusing on financial news. Therefore, what is achieved by a link-based profile that comprises a list of preferred URLs or hosts to characterize a user's preference may also be achievable, at least in part, by a category-based profile that has a set of categories that cover the same topics covered by preferred URLs or hosts.

[0050] It is a non-trivial operation to construct various types of user profiles that can be stored in the data structures shown in Figs. 3-5 based on the user information listed in Fig. 2. Given a document identified (e.g., viewed) by a user, different terms in the document may have different importance in revealing the topic of the document. Some terms, e.g., the document's title, may be extremely important, while other terms may have little importance. For example, many documents contain navigational links, copyright statements, disclaimers and other text that may not be related to the topic of the document. How to efficiently select appropriate documents, content from those documents and terms from within the content is a challenging topic in computational linguistics. Additionally, it is preferred to minimize the volume of user information processed, so as make the process of user profile construction computationally efficient. Skipping less important terms in a document helps in accurately matching a document with a user's interest.

[0051] Paragraph sampling (described below with reference to Fig. 6) is a procedure for automatically extracting content from a document that may be relevant to a user. An important observation behind this procedure is that less relevant content in a document, such as navigational links, copyright statements, disclaimer, etc., tend to be relatively short segments of text. In one embodiment, paragraph sampling looks for the paragraphs of greatest length in a document, processing the paragraphs in order of decreasing length until

the length of a paragraph is below a predefined threshold. The paragraph sampling procedure optionally selects up to a certain maximum amount of content from each processed paragraph. If few paragraphs of suitable length are found in a document, the procedure falls back to extracting text from other parts of the document, such as anchor text and ALT tags.

[0052] Fig. 6 is a flowchart illustrating the major steps of paragraph sampling. Paragraph sampling begins with the step 610 of removing predefined items, such as comments, JavaScript and style sheets, etc., from a document. These items are removed because they are usually related to visual aspects of the document when rendered on a browser and are unlikely to be relevant to the document's topic. Following that, the procedure may select the first N words (or M sentences) at step 620 from each paragraph whose length is greater than a threshold value, MinParagraphLength, as sampled content. In one embodiment, the values of N and M are chosen to be 100 and 5, respectively. Other values may be used in other embodiments.

[0053] In order to reduce the computational and storage load associated with the paragraph sampling procedure, the procedure may impose a maximum limit, e.g., 1000 words, on the sampled content from each document. In one embodiment, the paragraph sampling procedure first organizes all the paragraphs in a document in length decreasing order, and then starts the sampling process with a paragraph of maximum length. It is noted that the beginning and end of a paragraph depend on the appearance of the paragraph in a browser, not on the presence of uninterrupted a text string in the HTML representation of the paragraph. For this reason, certain HTML commands, such as commands for inline links and for bold text, are ignored when determining paragraph boundaries. In some embodiments, the paragraph sampling procedure screens the first N words (or M sentences) so as to filter out those sentences including boilerplate terms like "Terms of Service" or "Best viewed", because such sentences are usually deemed irrelevant to the document's topic.

[0054] Before sampling a paragraph whose length is above the threshold value, the procedure may stop sampling content from the document if the number of words in the sampled content has reached the maximum word limit. If the maximum word limit has not been reached after processing all paragraphs of length greater than the threshold, optional steps 630, 640, 650 and 670 are performed. In particular, the procedure adds the document

title (630), the non-inline HREF links (640), the ALT tags (650) and the meta tags (670) to the sampled content until it reaches the maximum word limit.

[0055] Once the documents identified by a user have been scanned, the sampled content can be used for identifying a list of most important (or unimportant) terms through context analysis. Context analysis attempts to learn context terms that predict the most important (or unimportant) terms in a set of identified documents. Specifically, it looks for prefix patterns, postfix patterns, and a combination of both. For example, an expression "x's home page" may identify the term "x" as an important term for a user and therefore the postfix pattern "* home page" can be used to predict the location of an important term in a document, where the asterisk "*" represents any term that fits this postfix pattern. In general, the patterns identified by context analysis usually consist of m terms before an important (or unimportant) term and n terms after the important (or unimportant) term, where both m and n are greater than or equal to 0 and at least one of them is greater than 0. Typically, m and n are less than 5, and when non-zero are preferably between 1 and 3. Depending on its appearance frequency, a pattern may have an associated weight that indicates how important (or unimportant) the term recognized by the pattern is expected to be.

[0056] According to one embodiment of the present invention (Fig. 7A), context analysis has two distinct phases, a training phase 701 and an operational phase 703. The training phase 701 receives and utilizes a list of predefined important terms 712, an optional list of predefined unimportant terms 714, and a set of training documents (step 710). In some embodiments, the list of predefined unimportant terms is not used. The source of the lists 712, 714 is not critical. In some embodiments, these lists 712, 714 are generated by extracting words or terms from a set of documents (e.g., a set of several thousand web pages of high page rank) in accordance with a set of rules, and then editing them to remove terms that in the opinion of the editor do not belong in the lists. The source of the training documents is also not critical. In some embodiments, the training documents comprise a randomly or pseudo-randomly selected set of documents already known to the search engine. In other embodiments, the training documents are selected from a database of documents in the search engine in accordance with predefined criteria.

[0057] During the training phase 701, the training documents are processed (step 720), using the lists of predefined important and unimportant terms, so as to identify a plurality of context patterns (e.g., prefix patterns, postfix patterns, and prefix-postfix patterns) and to associate a weight with each identified context pattern. During the operational phase 703, the context patterns are applied to documents identified by the user (step 730) to identify a set of important terms (step 740) that characterize the user's specific interests and preferences. Learning and delineating a user's interests and preferences is usually an ongoing process. Therefore, the operational phase 703 may be repeated to update the set of important terms that have been captured previously. This may be done each time a user accesses a document, according to a predetermined schedule, at times determined in accordance with specified criteria, or otherwise from time to time. Similarly, the training phase 701 may also be repeated to discover new sets of context patterns and to recalibrate the weights associated with the identified context patterns.

[0058] Below is a segment of pseudo code that exemplifies the training phase:

```
For each document in the set {
   For each important term in the document {
      For m = 0 to MaxPrefix {
         For n = 0 to MaxPostfix {
            Extract the m words before the important
            term and the n words after the important
            term as s;
            Add 1 to ImportantContext(m,n,s);
         }
      }
   }

   For each unimportant term in the document {
      For m = 0 to MaxPrefix {
         For n = 0 to MaxPostfix {
            Extract the m words before the
            unimportant term and the n words after
            the unimportant term as s;
            Add 1 to UnimportantContext(m,n,s);
         }
      }
   }
}

For m = 0 to MaxPrefix {
   For n = 0 to MaxPostfix {
```

```
            For each value of s {
               Set the weight for s to a function of
               ImportantContext(m,n,s), and
               UnimportantContext(m,n,s);
            }
         }
      }
```

[0059] In the pseudo code above, the expression *s* refers to a prefix pattern (n=0), a postfix pattern (m=0) or a combination of both (m>0 & n>0). Each occurrence of a specific pattern is registered at one of the two multi-dimensional arrays, ImportantContext(m,n,s) or UnimportantContext(m,n,s). The weight of a prefix, postfix or combination pattern is set higher if this pattern identifies more important terms and fewer unimportant terms and vice versa. Note that it is possible that a same pattern may be associated with both important and unimportant terms. For example, the postfix expression "* operating system" may be used in the training documents 716 in conjunction with terms in the list of predefined important terms 712 and also used in conjunction with terms in the list of predefined unimportant terms 714. In this situation, the weight associated with the postfix pattern "* operating system" (represented by the expression Weight(1,0,"operating system")) will take into account the number of times the postfix expression is used in conjunction with terms in the list of predefined important terms as well as the number of times the postfix expression is used in conjunction with terms in the list of predefined unimportant terms. One possible formula to determine the weight of a context pattern *s* is:

Weight(m,n,s) = Log(ImportantContext(m,n,s)+1)- Log(UnimportantContext(m,n,s)+1).

Other weight determination formulas may be used in other embodiments.

[0060] In the second phase of the context analysis process, the weighted context patterns are used to identify important terms in one or more documents identified by the user. Referring to Fig. 7B, in the first phase a computer system receives training data 750 and creates a set of context patterns 760, each context pattern having an associated weight. The computer system then applies the set of context patterns 760 to a document 780. In Fig. 7B, previously identified context patterns found within the document 780 are highlighted. Terms 790 associated with the context patterns are identified and each such term receives a weight based on the weights associated with the context patterns. For example, the term "Foobar"

appears in the document twice, in association with two different patterns, the prefix pattern "Welcome to *" and the postfix pattern "* builds", and the weight 1.2 assigned to "Foobar" is the sum of the two patterns' weights, 0.7 and 0.5. The other identified term "cars" has a weight of 0.8 because the matching prefix pattern "world's best *" has a weight of 0.8. In some embodiments the weight for each term is computed using a log transform, where the final weight is equal to log(initial weight +1). It is possible that the two terms "Foobar" and "cars" may not be in the training data 750 and may have never been encountered by the user before. Nevertheless, the context analysis method described above identifies these terms and adds them to the user's term-based profile. Thus, context analysis can be used to discover terms associated with a user's interests and preferences even when those terms are not included in a predefined database of terms.

[0061] As noted, the output of context analysis can be used directly in constructing a user's term-based profile. Additionally, it may be useful in building other types of user profiles, such as a user's category-based profile. For example, a set of weighted terms can be analyzed and classified into a plurality of categories covering different topics, and those categories can be added to a user's category-based profile.

[0062] After executing the context analysis on a set of documents identified by or for a user, the resulting set of terms and weights may occupy a larger amount of storage than allocated for each user's term-based profile. Also, the set of terms and corresponding weights may include some terms with weights much, much smaller than other terms within the set. Therefore, in some embodiments, at the conclusion of the context analysis, the set of terms and weights is pruned by removing terms having the lowest weights (A) so that the total amount of storage occupied by the term-based profile meets predefined limits, and/or (B) so as to remove terms whose weights are so low, or terms that correspond to older items, as defined by predefined criteria, that the terms are deemed to be not indicative of the user's search preferences and interests. In some embodiments, similar pruning criteria and techniques are also applied to the category-based profile and/or the link-based profile.

[0063] In some embodiments, a user's profile is updated each time the user performs a search and selects at least one document from the search results to download or view. In some embodiments, the search engine builds a list of documents identified by the user (e.g.,

by selecting the documents from search results) over time, and at predefined times (e.g., when the list reaches a predefined length, or a predefined amount of time has elapsed), performs a profile update. When performing an update, new profile data is generated, and the new profile data is merged with the previously generated profile data for the user. In some embodiments, the new profile data is assigned higher importance than the previously generated profile data, thereby enabling the system to quickly adjust a user's profile in accordance with changes in the user's search preferences and interests. For example, the weights of items in the previously generated profile data may be automatically scaled downward prior to merging with the new profile data. In one embodiment, there is a date associated with each item in the profile, and the information in the profile is weighted based on its age, with older items receiving a lower weight than when they were new. In other embodiments, the new profile data is not assigned high importance than the previously generated profile data.

[0064] The paragraph sampling and context analysis methods may be used independently or in combination. When used in combination, the output of the paragraph sampling is used as input to the context analysis method.

[0065] It is further noted that the above-described methods used for creating user profiles, e.g., paragraph sampling and context analysis, may be also leveraged for determining the relevance of a candidate document to a user's preference. Indeed, the primary mission of a search engine is to identify a series of documents that are most relevant to a user's preference based on the search queries submitted by the user as well as the user's user profile. Fig. 8 illustrates several exemplary data structures that can be used to store information about a document's relevance to a user profile from multiple perspectives. For each candidate document, each identified by a respective DOC_ID, term-based document information table 810 includes multiple pairs of terms and their weights, category-based document information table 830 includes a plurality of categories and associated weights, and link-based document information table 850 includes a set of links and corresponding weights.

[0066] The rightmost column of each of the three tables (810, 830 and 850) stores the rank (i.e., a computed score) of a document when the document is evaluated using one specific type of user profile. A user profile rank can be determined by combining the weights

of the items associated with a document. For instance, a category-based or topic-based profile rank may be computed as follows. A user may prefer documents about science with a weight of 0.6, while he dislikes documents about business with a weight of -0.2. Thus, when a science document matches a search query, it will be weighted higher than a business document. In general, the document topic classification may not be exclusive. A candidate document may be classified as being a science document with probability of 0.8 and a business document with probability of 0.4. A link-based profile rank may be computed based on the relative weights allocated to a user's URL, host, domain, etc., preferences in the link-based profile. In one embodiment, term-based profile rank can be determined using known techniques, such as the term frequency-inverse document frequency (TF-IDF). The term frequency of a term is a function of the number of times the term appears in a document. The inverse document frequency is an inverse function of the number of documents in which the term appears within a collection of documents. For example, very common terms like "the" occur in many documents and consequently as assigned a relatively low inverse document frequency.

[0067] When a search engine generates search results in response to a search query, a candidate document D that satisfies the query is assigned a query score, QueryScore, in accordance with the search query. This query score is then modulated by document D's page rank, PageRank, to generate a generic score, GenericScore, that is expressed as

$$\text{GenericScore} = \text{QueryScore}*\text{PageRank}.$$

[0068] This generic score may not appropriately reflect document D's importance to a particular user U if the user's interests or preferences are dramatically different from that of the random surfer. The relevance of document D to user U can be accurately characterized by a set of profile ranks, based on the correlation between document D's content and user U's term-based profile, herein called the TermScore, the correlation between one or more categories associated with document D and user U's category-based profile, herein called the CategoryScore, and the correlation between the URL and/or host of document D and user U's link-based profile, herein called the LinkScore. Therefore, document D may be assigned a personalized rank that is a function of both the document's generic score and the user profile scores. In one embodiment, this personalized score can be expressed as:

PersonalizedScore = GenericScore*(TermScore+CategoryScore+LinkScore).

[0069] Figs. 9A and 9B represent two embodiments, both implemented in a client-server network environment such as the network environment 100 shown in Fig. 1. In the embodiment shown in Fig. 9A, the search engine 104 receives a search query from a client 102 at step 910 that is submitted by a particular user. In response, the search engine 104 may optionally generate a query strategy at step 915 (e.g., the search query is normalized so as to be in proper form for further processing, and/or the search query may be modified in accordance with predefined criteria so as to automatically broaden or narrow the scope of the search query). At step 920, the search engine 104 submits the search query (or the query strategy, if one is generated) to the content server 106. The content server identifies a list of documents that match the search query at step 920, each document having a generic score that depends on the document's page rank and the search query. In general, all the three operations (steps 910, 915 and 920) are conducted by the search engine system 107, which is on the server side of the network environment 100. There are two options on where to implement the operations following these first three steps.

[0070] In some embodiments that employ a server-side implementation, the user's identification number is embedded in the search query. Based on the user's identification number, the user profile server 108 identifies the user's user profile at step 925. Starting from step 930, the user profile server 108 or the search engine 104 analyzes each document identified at step 920 to determine its relevance to the user's profile, creates a profile score for the identified document at step 935 and then assigns the document a personalized score that is a function of the document's generic and profile scores at step 940. At step 942, the user profile server 108 or the search engine 104 checks whether this the last one in the list of identified documents. If no, the system processes the next document in the list. Otherwise, the list of documents are re-ordered according to their personalized scores and then sent to the corresponding client from which the user submitted the search query.

[0071] Embodiments using a client-side implementation are similar to the server-side implementation, except that after step 920, the identified documents are sent to the corresponding client from which the user submitted the query. This client stores the user's user profile and it is responsible for re-ordering the documents based upon the user profile.

Therefore, this client-side implementation may reduce the server's workload. Further, since there is no privacy concern with the client-side implementation, a user may be more willing to provide private information to customize the search results. However, a significant limitation to the client-side implementation is that only a limited number of documents, e.g., the top 50 documents (as determined using the generic rank), may be sent to a client for re-ordering due to limited network bandwidth. In contrast, the server-side implementation may be able to apply a user's profile to a much larger number of documents, e.g., 1000, that match the search query. Therefore, the client-side implementation may deprive a user access to those documents having relatively low generic ranks, but significantly high personalized ranks.

[0072] Fig. 9B illustrates another embodiment. Unlike the embodiment depicted in Fig. 9A, where the search query is not personalized before submitting the search query to the search engine 104, a generic query strategy is adjusted (step 965) according to the user's user profile to create a personalized query strategy. For example, relevant terms from the user profile may be added to the search query with associated weights. The creation of the personalized query strategy can be performed either on the client side or on the server side of the system. This embodiment avoids the network bandwidth restriction facing the previous embodiment. Finally, the search engine 104 submits the personalized query strategy to the content server 106 (step 970), and therefore the search results returned by the content server have already been ordered by the documents' personalized ranks (step 975).

[0073] The profiles of a group of users with related interests may be combined together to form a group profile, or a single profile may be formed based on the documents identified by the users in the group. For instance, several family members may use the same computer to submit search queries to a search engine. If the computer is tagged with a single user identifier by the search engine, the "user" will be the entire family of users, and the user profile will be represent a combination or mixture of the search preferences of the various family members. An individual user in the group may optionally have a separate user profile that differentiates this user from other group members. In operation, the search results for a user in the group are ranked according to the group profile, or according to the group profile and the user's user profile when the user also has a separate user profile.

[0074] It is possible that a user may switch his interests so dramatically that his new interests and preferences bear little resemblance to his user profile, or a user may be temporarily interested in a new topic. In this case, personalized search results produced according to the embodiments depicted in Figs. 9A and 9B may be less favorable than search results ranked in accordance with the generic ranks of the documents in the search results. Additionally, the search results provided to a user may not include new websites among the top listed documents because the user's profile tends to increase the weight of older websites which the user has visited (i.e., older websites from which the user has viewed or downloaded web pages) in the past.

[0075] To reduce the impact caused by a change in a user's preferences and interests, the personalized search results may be merged with the generic search results. In one embodiment, the generic search results and personalized search results are interleaved, with the odd positions (e.g., 1, 3, 5, etc.) of a search results list reserved for generic search results and the even positions (e.g., 2, 4, 6, etc.) reserved for personalized search results, or vice versa. Preferably, the items in the generic search results will not duplicate the items listed in the personalized search results, and vice versa. More generally, generic search results are intermixed or interleaved with personalized search results, so that the items in the search results presented to the user include both generic and personalized search results.

[0076] In another embodiment, the personalized ranks and generic ranks are further weighted by a user profile's confidence level. The confidence level takes into account factors such as how much information has been acquired about the user, how close the current search query matches the user's profile, how old the user profile is, etc. If only a very short history of the user is available, the user's profile may be assigned a correspondingly low confidence value. The final score of an identified document can be determined as:

FinalScore = ProfileScore*ProfileConfidence + GenericScore*(1- ProfileConfidence).

When intermixing generic and personalized results, the fraction of personalized results may be adjusted based on the profile confidence, for example using only one personalized result when the confidence is low.

[0077] Sometimes, multiple users may share a machine, e.g., in a public library. These users may have different interests and preferences. In one embodiment, a user may explicitly login to the service so the system knows his identity. Alternatively, different users can be automatically recognized based on the items they access or other characteristics of their access patterns. For example, different users may move the mouse in different ways, type differently, and use different applications and features of those applications. Based on a corpus of events on a client and/or server, it is possible to create a model for identifying users, and for then using that identification to select an appropriate "user" profile. In such circumstances, the "user" may actually be a group of people having somewhat similar computer usage patterns, interests and the like.

[0078] Referring to Fig. 10, a personalized search engine system 1000 typically includes one or more processing units (CPU's) 1002, one or more network or other communications interfaces 1010, memory 1012, and one or more communication buses 1014 for interconnecting these components. The system 1000 may optionally include a user interface 1004, for instance a display 1006 and a keyboard 1008. Memory 1012 may include high speed random access memory and may also include non-volatile memory, such as one or more magnetic disk storage devices. Memory 1012 may include mass storage that is remotely located from the central processing unit(s) 1002. The memory 1012 preferably stores:

- an operating system 1016 that includes procedures for handling various basic system services and for performing hardware dependent tasks;
- a network communication module 1018 that is used for connecting the system 1000 to other servers or computers via one or more communication networks (wired or wireless), such as the Internet, other wide area networks, local area networks, metropolitan area networks, and so on;
- a system initialization module 1020 that initializes other modules and data structures stored in memory 1012 required for the appropriate operation of system 1000;
- a search engine 1022 for processing a search query, identifying and ordering search results according to the search query and a user's profile;

- a user profile engine 1030 for gathering and processing user information, such as the user information identified in Fig. 2, and creating and updating a user's user profile that characterizes the user's search preferences and interests; and
- data structures 1040, 1060 and 1080 for storing a plurality of user profiles.

[0079] The search engine 1022 may further comprise:

- a generic rank module (or instructions) 1024 for processing a search query submitted by a user, identifying a list of documents matching the query and assigning each identified document a generic rank without reference to user specific information;
- a user profile rank module (or instructions) 1026 for correlating each of a plurality of documents identified by the generic rank module 1024 with the user's user profile and assigning the document a profile rank indicating the relevance of the document to the user's search preferences and interests; and
- a rank mixing module (or instructions) 1028 for combining the generic rank and the profile rank of an identified document into a personalized rank and re-ordering the list of documents according to their personalized ranks.

In some embodiments, these modules 1024, 1026, 1028 may be implemented within a single procedure or in a set of procedures that reside within a single software module.

[0080] The user profile engine 1030 may further comprise:

- a user information collection module 1032 for collecting and assorting various user information listed in Fig. 2;
- a document content extraction module 1034 for selecting and extracting content from the documents identified by the user, to identify content relevant to the user's interests, using techniques such as paragraph sampling (as discussed above); and
- a context analysis module 1036 for analyzing the content extracted by the document extraction module 1034 so as to identify terms that characterize a user's search preferences.

[0081] Each data structure hosting a user profile may further comprise:

- a data structure 1042, 1062 or 1082 for storing a term-based user profile;
- a data structure 1044, 1064 or 1084 for storing a category-based user profile; and
- a data structure 1046, 1066 or 1086 for storing a link-based user profile.

[0082] The foregoing description, for purpose of explanation, has been described with reference to specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the invention to the precise forms disclosed. Many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles of the invention and its practical applications, to thereby enable others skilled in the art to best utilize the invention and various embodiments with various modifications as are suited to the particular use contemplated.

What is claimed is:

1. A method of personalizing search results of a search engine, comprising:

accessing a user profile for a user based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user;

receiving a search query from the user;

identifying a set of search result documents that match the search query;

assigning a generic score to each of at least a plurality of the search result documents;

assigning a personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the user profile; and

ranking the set of search result documents according to their personalized scores.

2. The method of claim 1, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

3. The method of claim 1, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

4. The method of claim 1, including updating the user profile by:

updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

5. The method of claim 4, including updating the user profile by:

updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

6. The method of claim 5, wherein the link-based profile includes information about URLs or portions of URLs.

7. The method of claim 6, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

8. The method of claim 7, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

9. The method of claim 4, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

10. The method of claim 9, wherein the weight is a weight associated with occurrences of the term in the set of documents.

11. The method of claim 9, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

12. The method of claim 4, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

13. The method of claim 12, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

14. The method of claim 1, including updating the user profile by:

updating a category-based profile by classifying a document in the set of documents into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

15. The method of claim 1, wherein the user profile is stored on a server of the search engine.

16. The method of claim 1, wherein the user profile is stored on a client associated with the user.

17. The method of claim 1, wherein the user is a group of users.

18. A method of personalizing search results of a search engine, comprising:

creating a plurality of user profiles for a plurality of users, each user profile including at least a user's identification number and information derived from documents visited by the user;

receiving a search query from a user of the plurality of users, the search query including at least one query term and the user's identification number;

selecting a set of documents from the Internet, assigning to each document in the set a generic score that characterizes the relevance of the document to the at least one query term;

retrieving the user's user profile and assigning to each of the set of documents a profile score based on the user profile; and

ranking the set of documents according to their generic and profile scores.

19. The method of claim 18, wherein the step of creating a user's user profile further comprises:

creating a term-based profile by extracting a set of terms from documents visited by the user and associating a weight with each extracted term; and

creating a category-based profile by determining a plurality of categories associated with documents visited by the user and associating a weight with each determined category.

20. The method of claim 18, wherein the step of creating a user's user profile further comprises:

creating a link-based profile by analyzing links in documents visited by the user and associating weights with the link.

21. The method of claim 18, wherein the user profile for a particular user includes demographic and geographic information provided by the user.

22. The method of claim 18, wherein the documents visited by the user from which information is derived for use in a particular user's user profile are selected based on the user's activities when visiting the documents.

23. The method of claim 18, including storing the plurality of user profiles on a server of the search engine; and

the retrieving including identifying the user's user profile based on a user identification number associated with both the user and the user's user profile.

24. The method of claim 18, including storing the plurality of user profiles on client computers associated with the plurality of users.

25. The method of claim 18, wherein the ranked set of documents comprises two sets of documents, one set of documents ordered by their generic scores and the other set of documents ordered by personalized scored generated by combining the document's generic and profile scores.

26. The method of claim 25, including interleaving or intermixing the two sets of documents to form the ranked set of documents.

27. A search engine system, comprising:

one or more central processing units for executing programs;

an interface for receiving event messages; and

a search engine module executable by the one or more central processing units, the module comprising:

instructions for accessing a user profile for a user, the user profile based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine system,

documents accessed by the user, documents linked to the documents identified by search results from the search engine system, and documents linked to the documents accessed by the user;

instructions for receiving a search query from the user;

instructions for identifying a set of search result documents that match the search query;

instructions for assigning a generic score to each of at least a plurality of the search result documents;

instructions for assigning a personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the user profile; and

instructions for ranking at least the plurality of the search result documents according to their personalized scores.

28. The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

29. The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

30. The system of claim 27, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

31. The system of claim 30, further including:

instructions for updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

32. The system of claim 31, wherein the link-based profile includes information about URLs or portions of URLs.

33. The system of claim 32, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

34. The system of claim 33, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

35. The system of claim 30, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

36. The system of claim 35, wherein the weight is a weight associated with occurrences of the term in the set of documents.

37. The system of claim 35, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

38. The system of claim 30, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

39. The system of claim 38, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

40. The system of claim 27, wherein the user profile is stored on a server of the search engine.

41. The system of claim 27, wherein the user profile is stored on a client associated with the user.

42. The system of claim 27, wherein the user is a group of users.

43. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for accessing a user profile for a user based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from a search engine, documents accessed by the user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user;

instructions for receiving a search query from the user;

instructions for identifying a set of search result documents that match the search query;

instructions for assigning a generic score to each of at least a plurality of the search result documents;

instructions for assigning a personalized score to each document of the plurality of search result documents in accordance with the generic scores assigned to the document and the user profile; and

instructions for ranking at least the plurality of the search result documents according to their personalized scores.

44. The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

45. The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

46. The computer program product of claim 43, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

47. The computer program product of claim 46, further including:

instructions for updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

48. The computer program product of claim 46, wherein the link-based profile includes information about URLs or portions of URLs.

49. The computer program product of claim 48, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

50. The computer program product of claim 49, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

51. The computer program product of claim 46, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

52. The computer program product of claim 48, wherein the weight is a weight associated with occurrences of the term in the set of documents.

53. The computer program product of claim 48, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

54. The computer program product of claim 46, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

55. The computer program product of claim 54, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

56. The computer program product of claim 43, wherein the user profile is stored on a server of the search engine.

57. The computer program product of claim 43, wherein the user profile is stored on a client associated with the user.

58. The computer program product of claim 43, wherein the user is a group of users.

Personalization of Web Search

ABSTRACT OF THE INVENTION

A system and method for creating a user profile and for using the user profile to order search results returned by a search engine. The user profile is based on search queries submitted by a user, the user's specific interaction with the documents identified by the search engine and personal information provided by the user. Generic scores associated with the search results are modulated by the user profile to measure their relevance to a user's preference and interest. The search results are re-ordered accordingly so that the most relevant results appear on the top of the list. User profiles can be created and/or stored on the client side or server side of a client-server network environment.

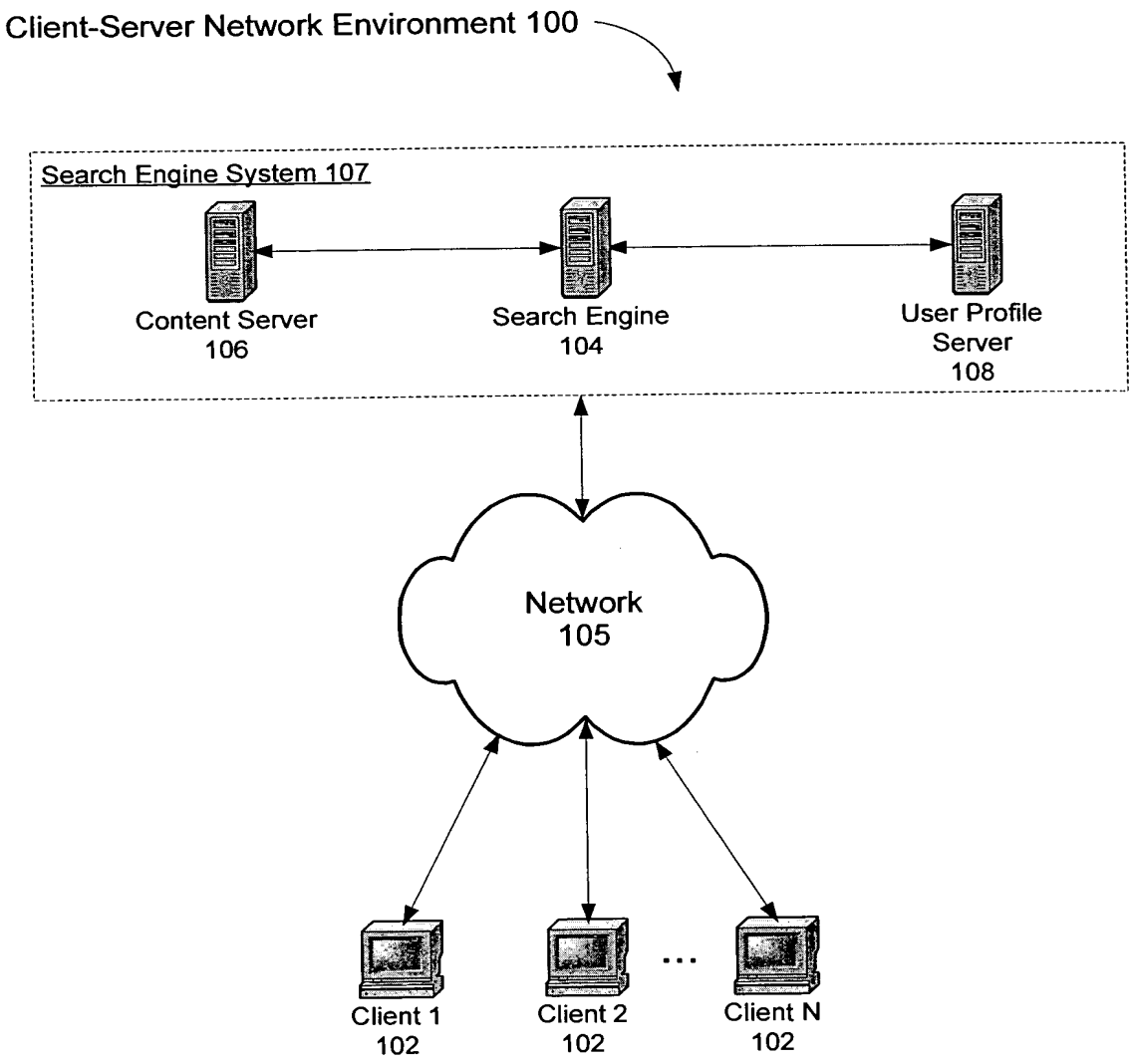

Client-Server Network Environment 100
Search Engine System 107
Content Server
106
Search Engine
104
User Profile
Server
108
Network
105
Client 1
102
Client 2
102
...
Client N
102

Fig. 1

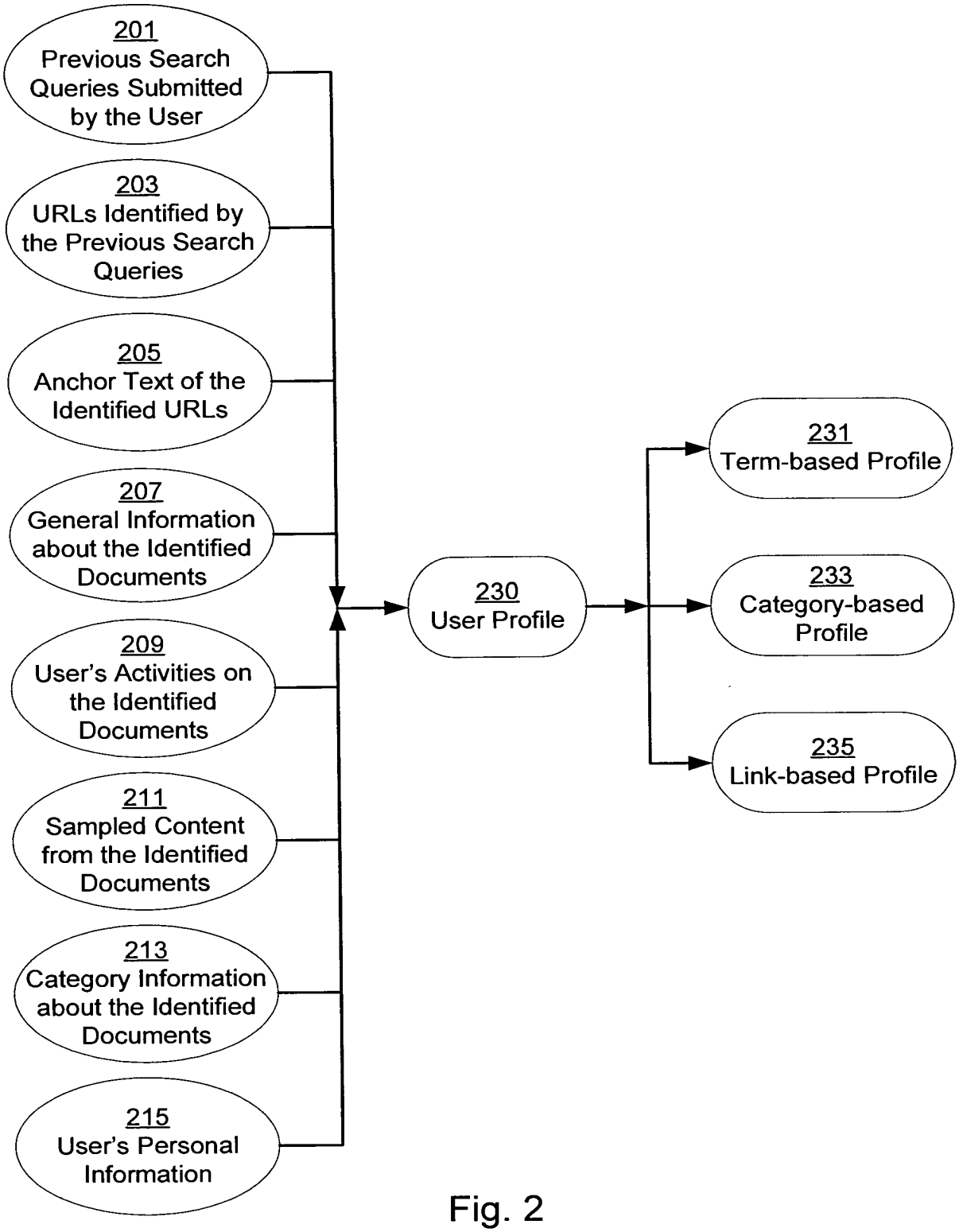

201
Previous Search Queries Submitted by the User
203
URLs Identified by the Previous Search Queries
205
Anchor Text of the Identified URLs
207
General Information about the Identified Documents
209
User's Activities on the Identified Documents
211
Sampled Content from the Identified Documents
213
Category Information about the Identified Documents
215
User's Personal Information
230
User Profile
231
Term-based Profile
233
Category-based Profile
235
Link-based Profile

Fig. 2

Term-based Profile Table 300

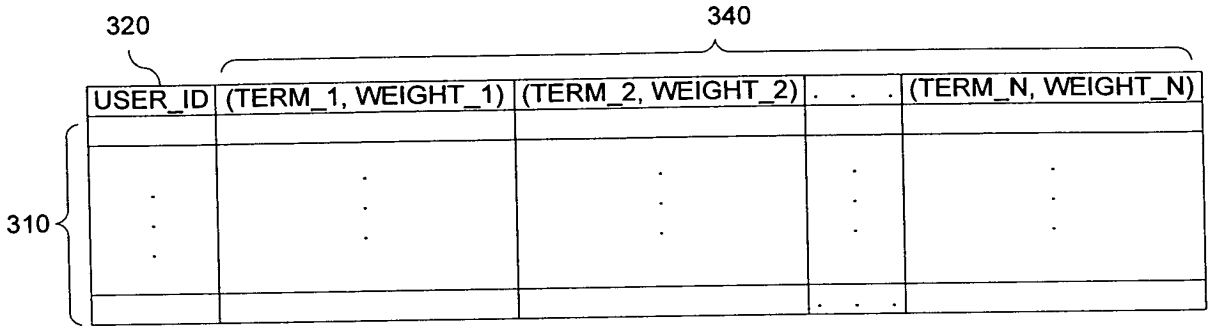


Fig. 3

Link-based Profile Table 500

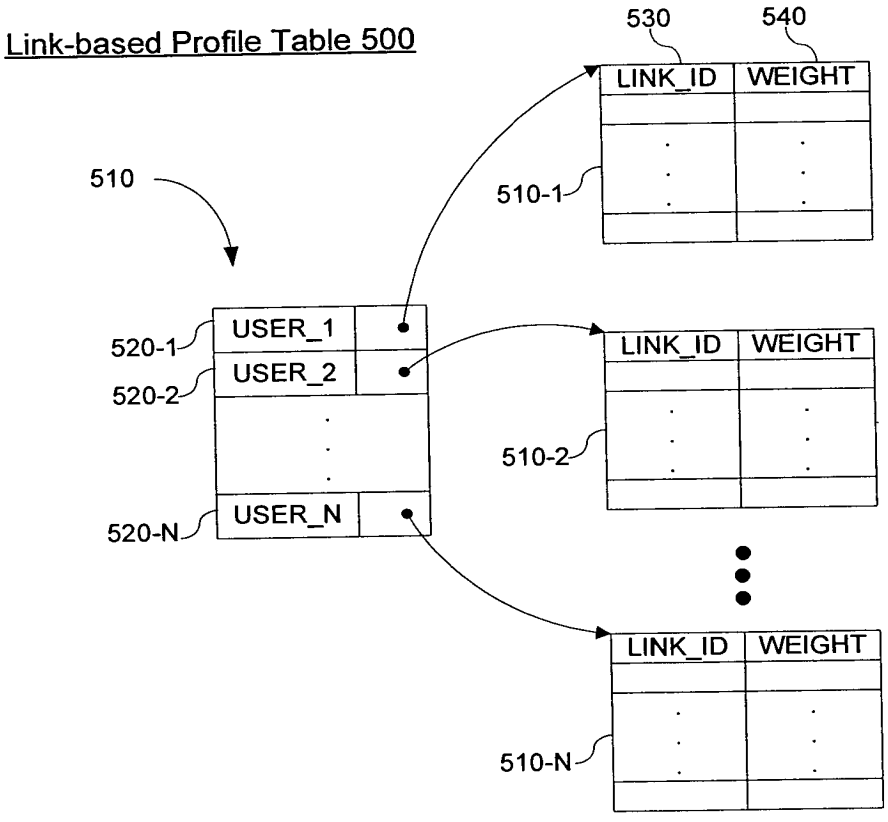


Fig. 5

Category Map 400

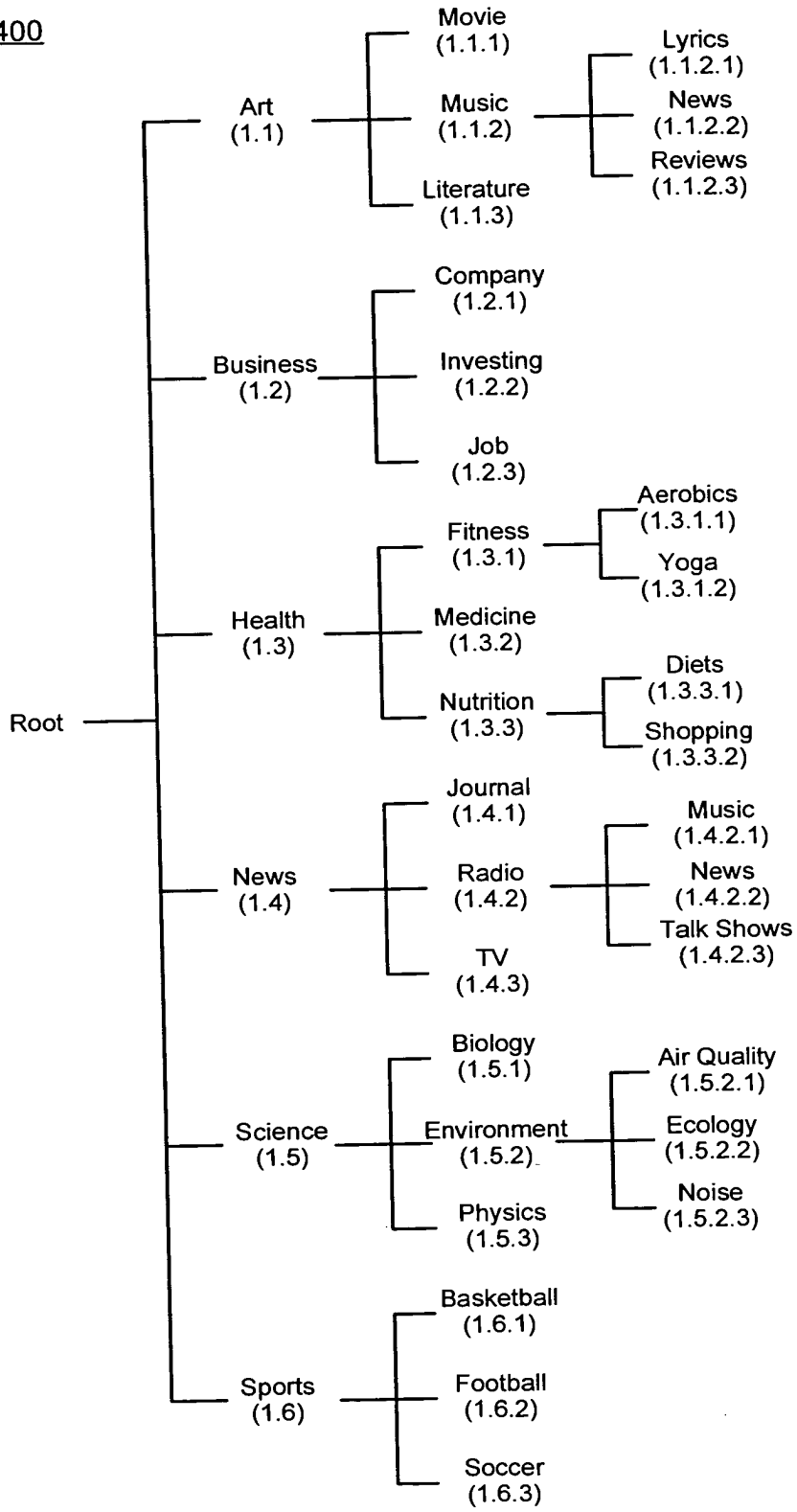


Fig. 4A

Category-based Profile Table 450

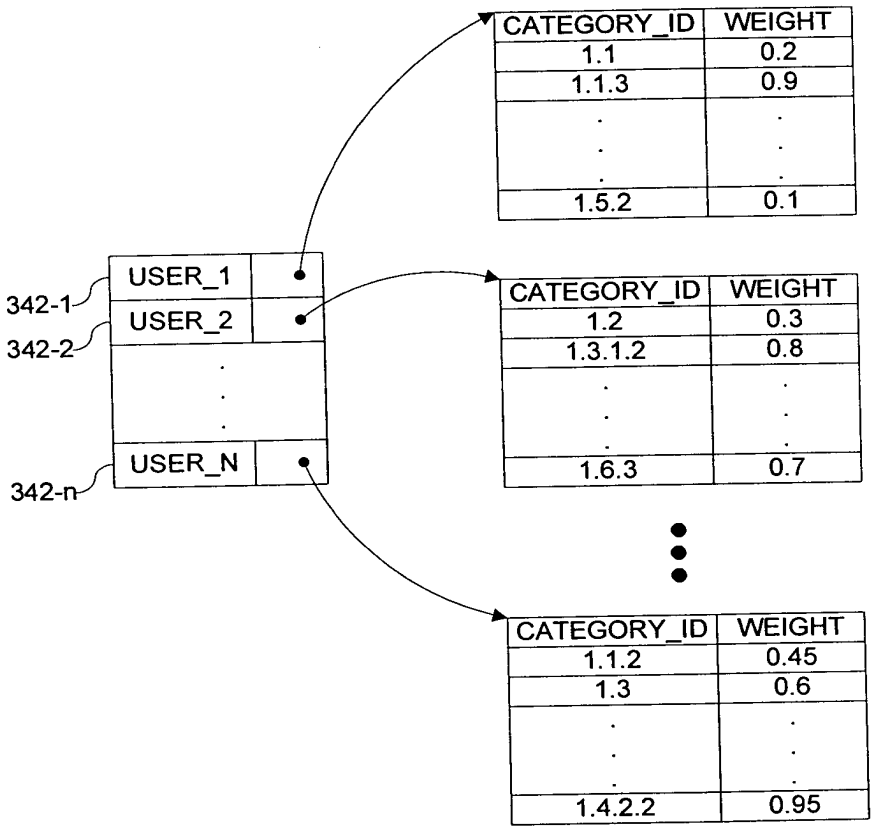


Fig. 4B

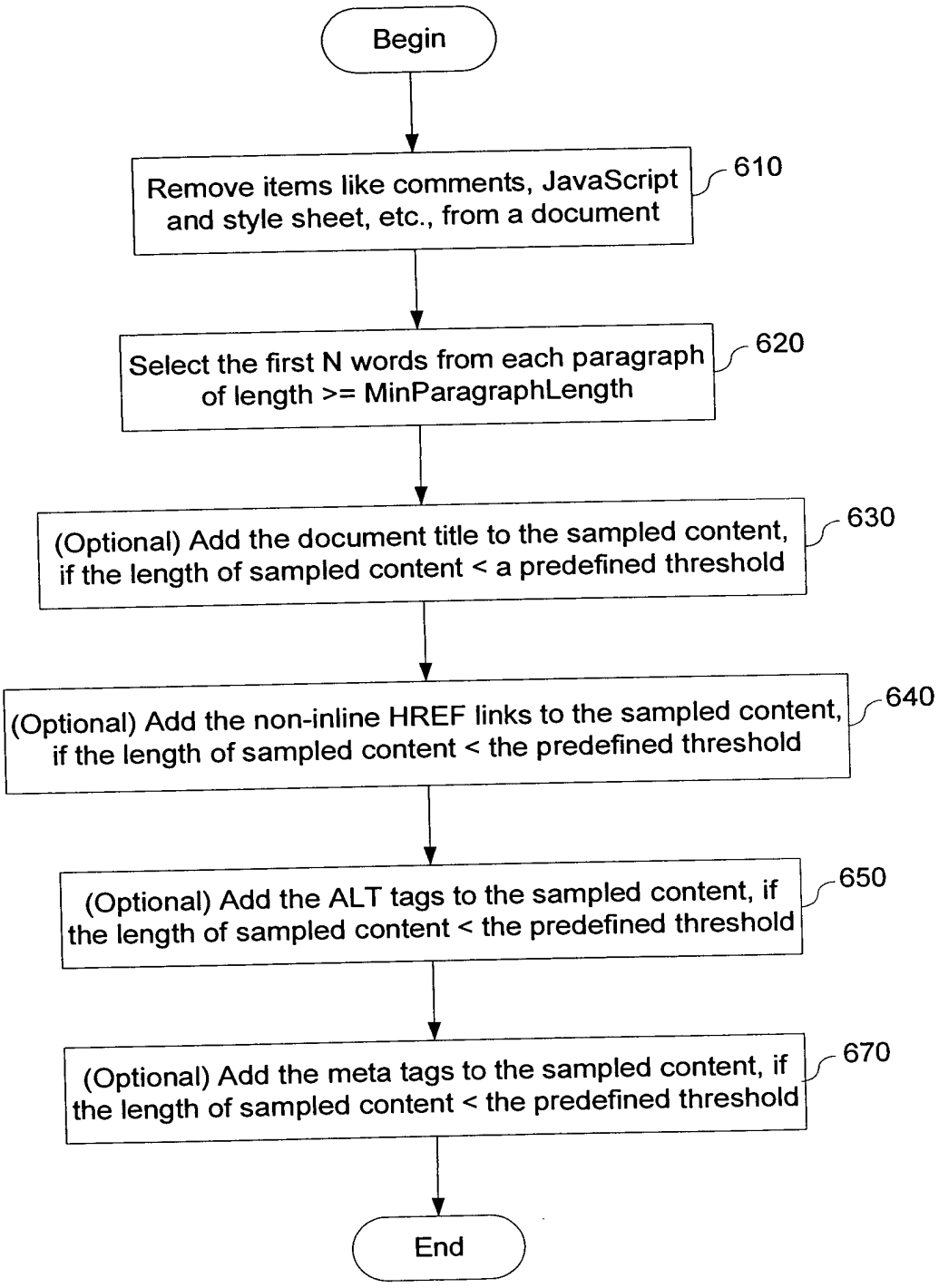

Begin
Remove items like comments, JavaScript and style sheet, etc., from a document
610
Select the first N words from each paragraph of length >= MinParagraphLength
620
(Optional) Add the document title to the sampled content, if the length of sampled content < a predefined threshold
630
(Optional) Add the non-inline HREF links to the sampled content, if the length of sampled content < the predefined threshold
640
(Optional) Add the ALT tags to the sampled content, if the length of sampled content < the predefined threshold
650
(Optional) Add the meta tags to the sampled content, if the length of sampled content < the predefined threshold
670
End

Fig. 6

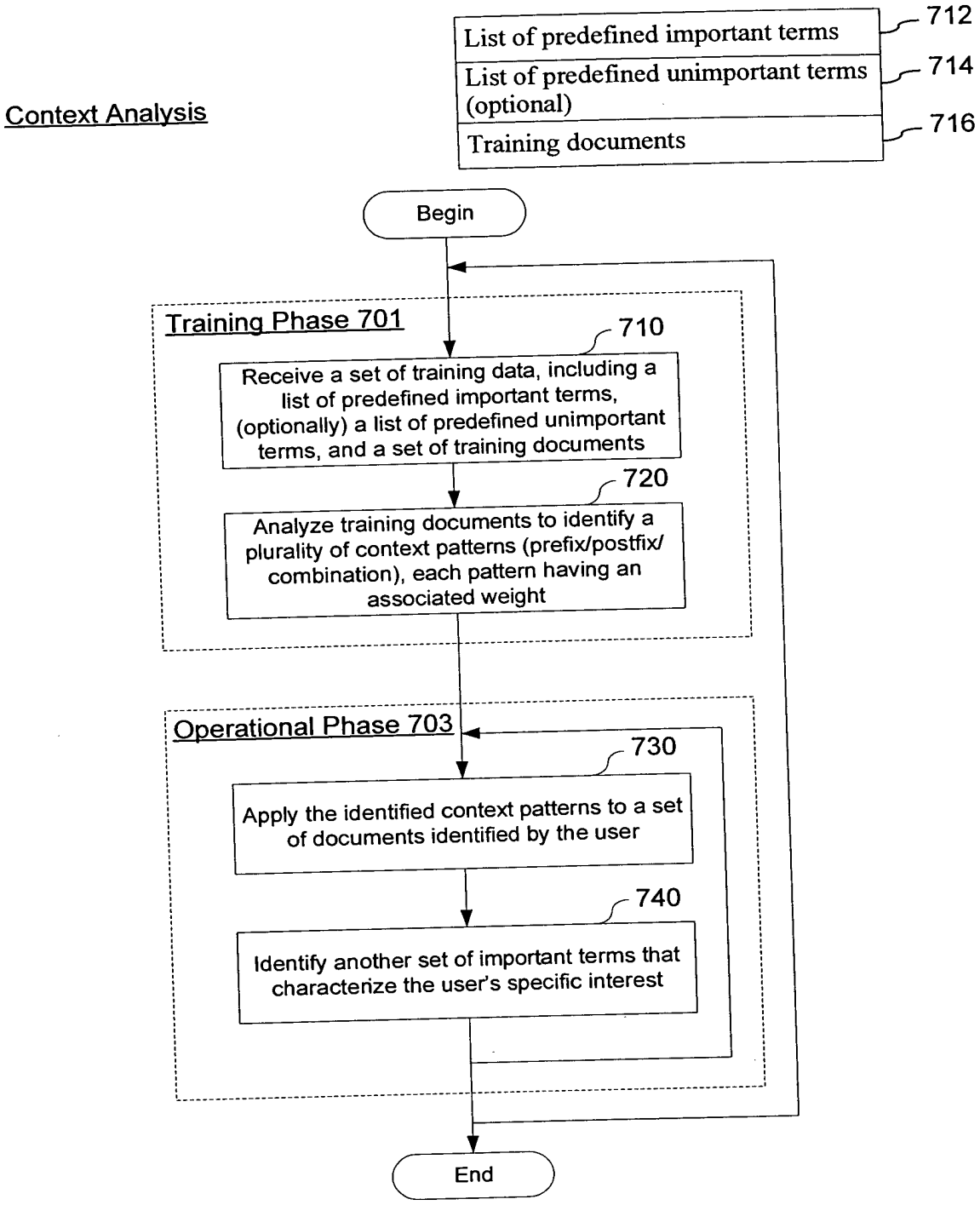

Context Analysis
List of predefined important terms
List of predefined unimportant terms (optional)
Training documents
712
714
716
Begin
Training Phase 701
710
Receive a set of training data, including a list of predefined important terms, (optionally) a list of predefined unimportant terms, and a set of training documents
720
Analyze training documents to identify a plurality of context patterns (prefix/postfix/ combination), each pattern having an associated weight
Operational Phase 703
730
Apply the identified context patterns to a set of documents identified by the user
740
Identify another set of important terms that characterize the user's specific interest
End

Fig. 7A

Context Analysis

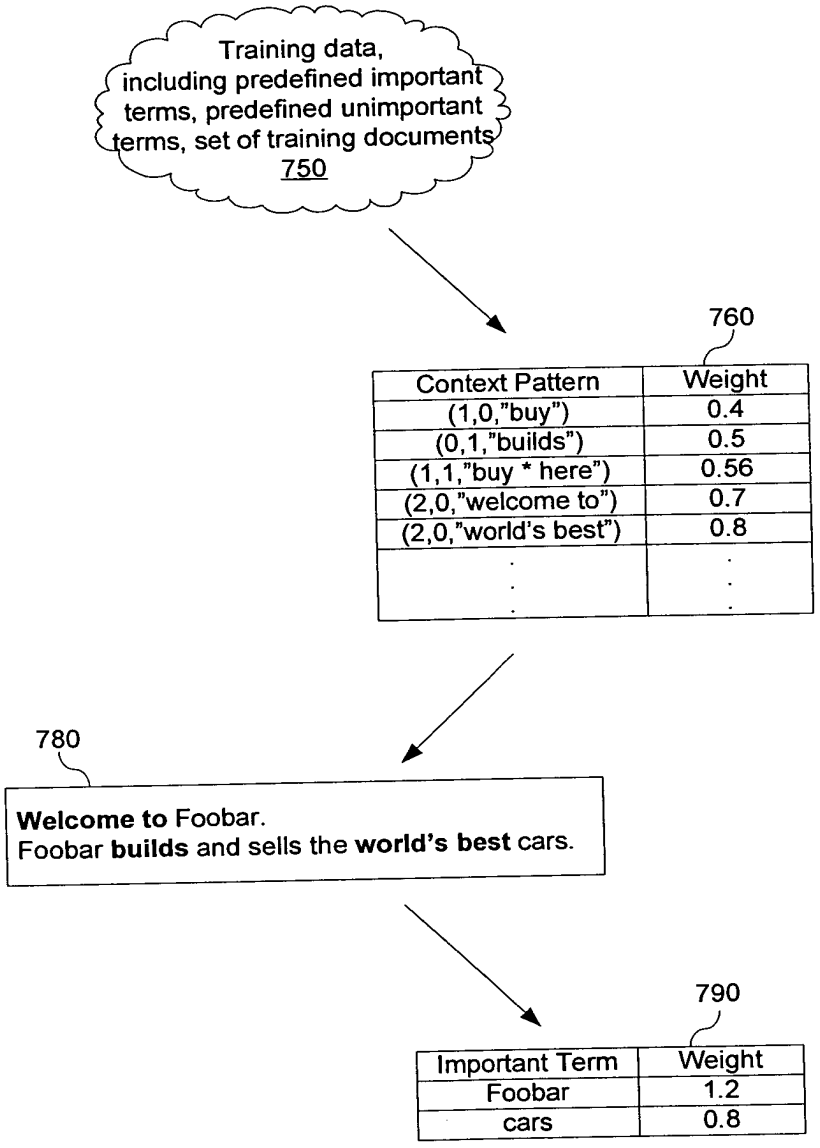


Fig. 7B

Term-based Document Information Table 810

DOC_ID	(TERM_1, WEIGHT_1)	(TERM_2, WEIGHT_2)	. . .	(TERM_X, WEIGHT_X)	Term-based Ranking Score
. . .	. . .	. . .	. . .	. . .	. . .
			. . .		

Category-based Document Information Table 830

DOC_ID	(CATEGORY_1, WEIGHT_1)	(CATEGORY_2, WEIGHT_2)	. . .	(CATEGORY_Y, WEIGHT_Y)	Category-based Ranking Score
. . .	. . .	. . .	. . .	. . .	. . .
			. . .		

Link-based Document Information Table 850

DOC_ID	(LINK_1, WEIGHT_1)	(LINK_2, WEIGHT_2)	. . .	(LINK_Z, WEIGHT_Z)	Link-based Ranking Score
. . .	. . .	. . .	. . .	. . .	. . .
			. . .		

Fig. 8

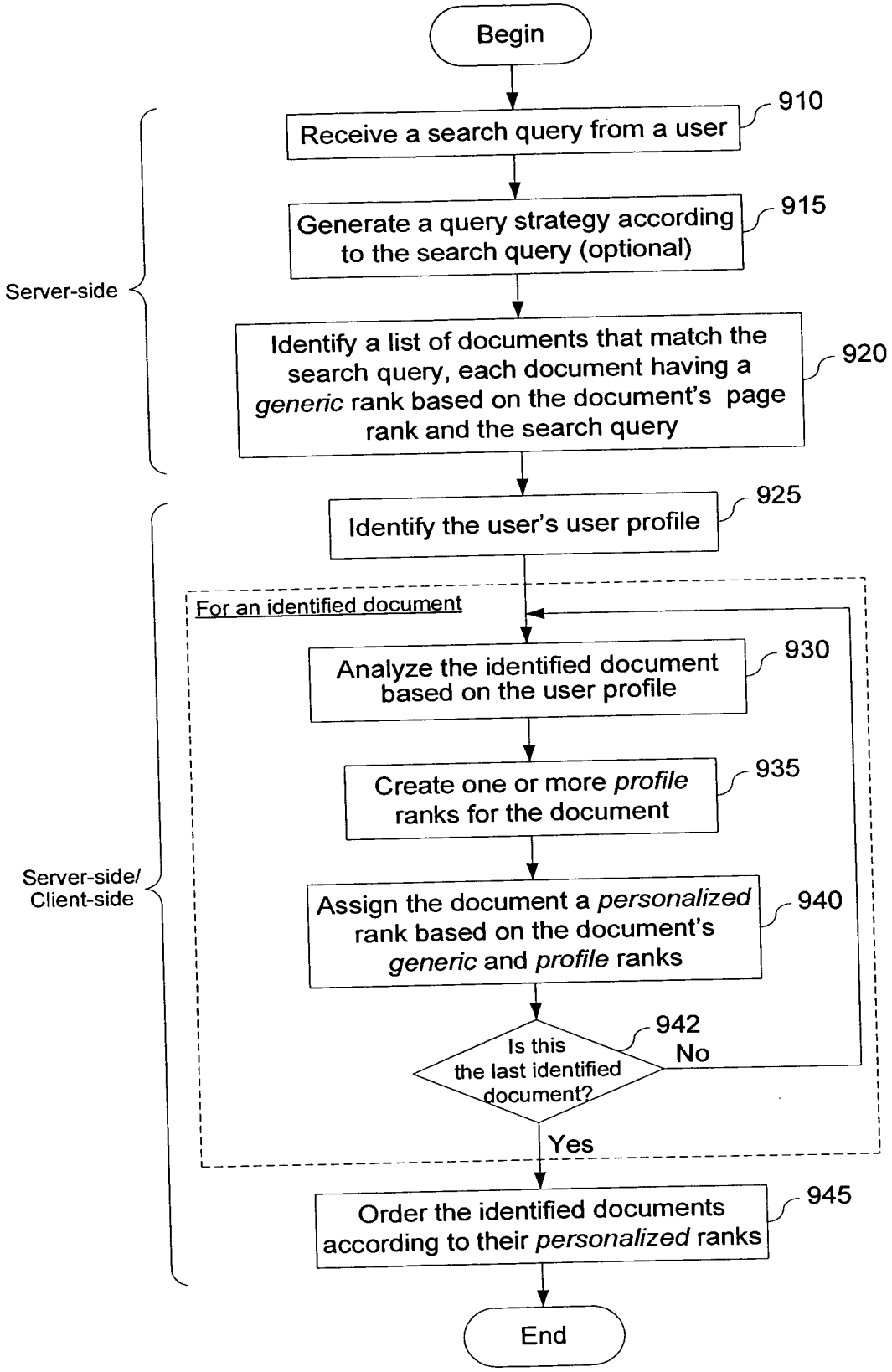

Begin
Receive a search query from a user
910
Generate a query strategy according to the search query (optional)
915
Server-side
Identify a list of documents that match the search query, each document having a *generic* rank based on the document's page rank and the search query
920
Identify the user's user profile
925
For an identified document
Analyze the identified document based on the user profile
930
Create one or more *profile* ranks for the document
935
Server-side/ Client-side
Assign the document a *personalized* rank based on the document's *generic* and *profile* ranks
940
Is this the last identified document?
942
No
Yes
Order the identified documents according to their *personalized* ranks
945
End

Fig. 9A

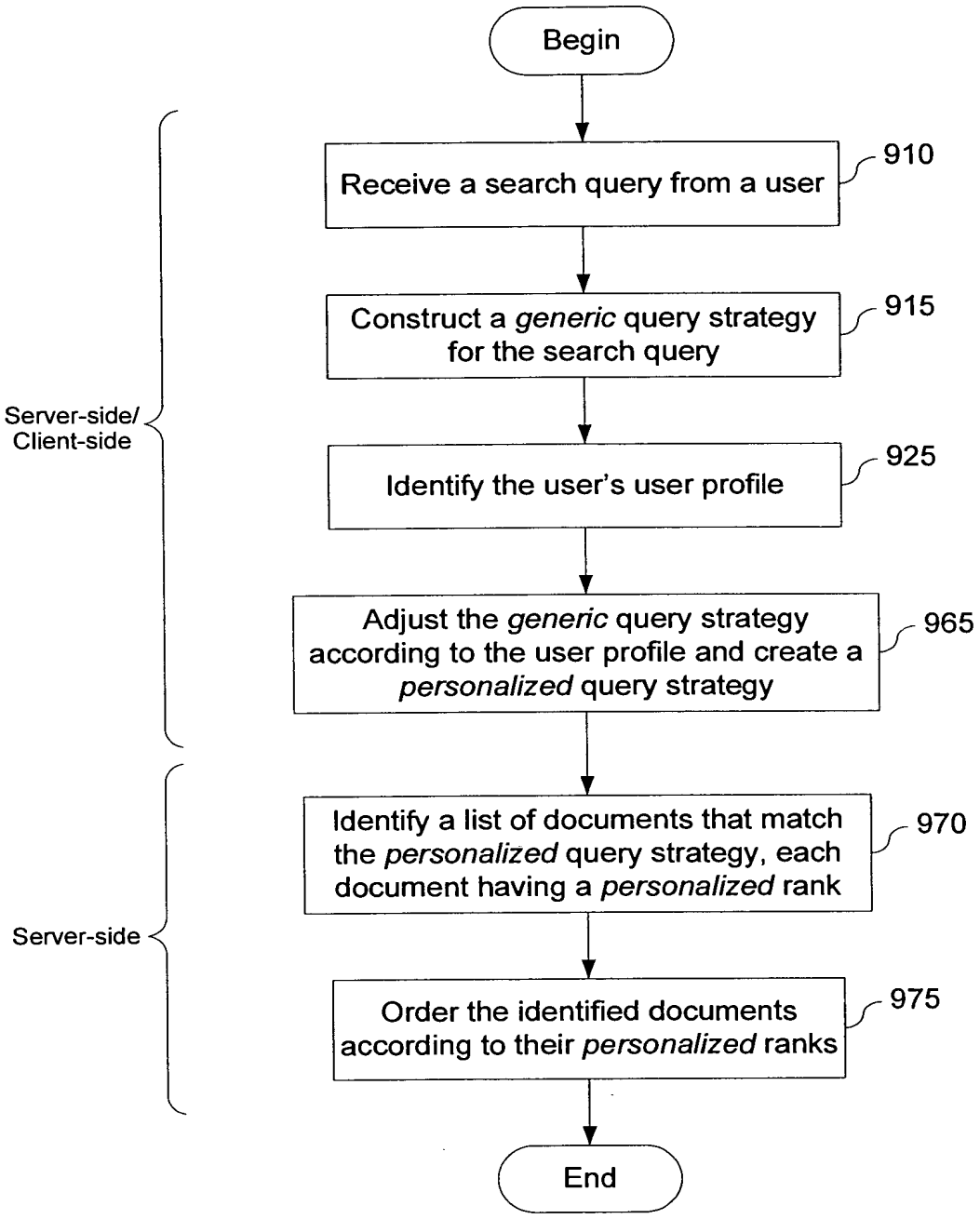

Begin
Receive a search query from a user
910
Construct a *generic* query strategy for the search query
915
Identify the user's user profile
925
Adjust the *generic* query strategy according to the user profile and create a *personalized* query strategy
965
Identify a list of documents that match the *personalized* query strategy, each document having a *personalized* rank
970
Order the identified documents according to their *personalized* ranks
975
End
Server-side/ Client-side
Server-side

Fig. 9B

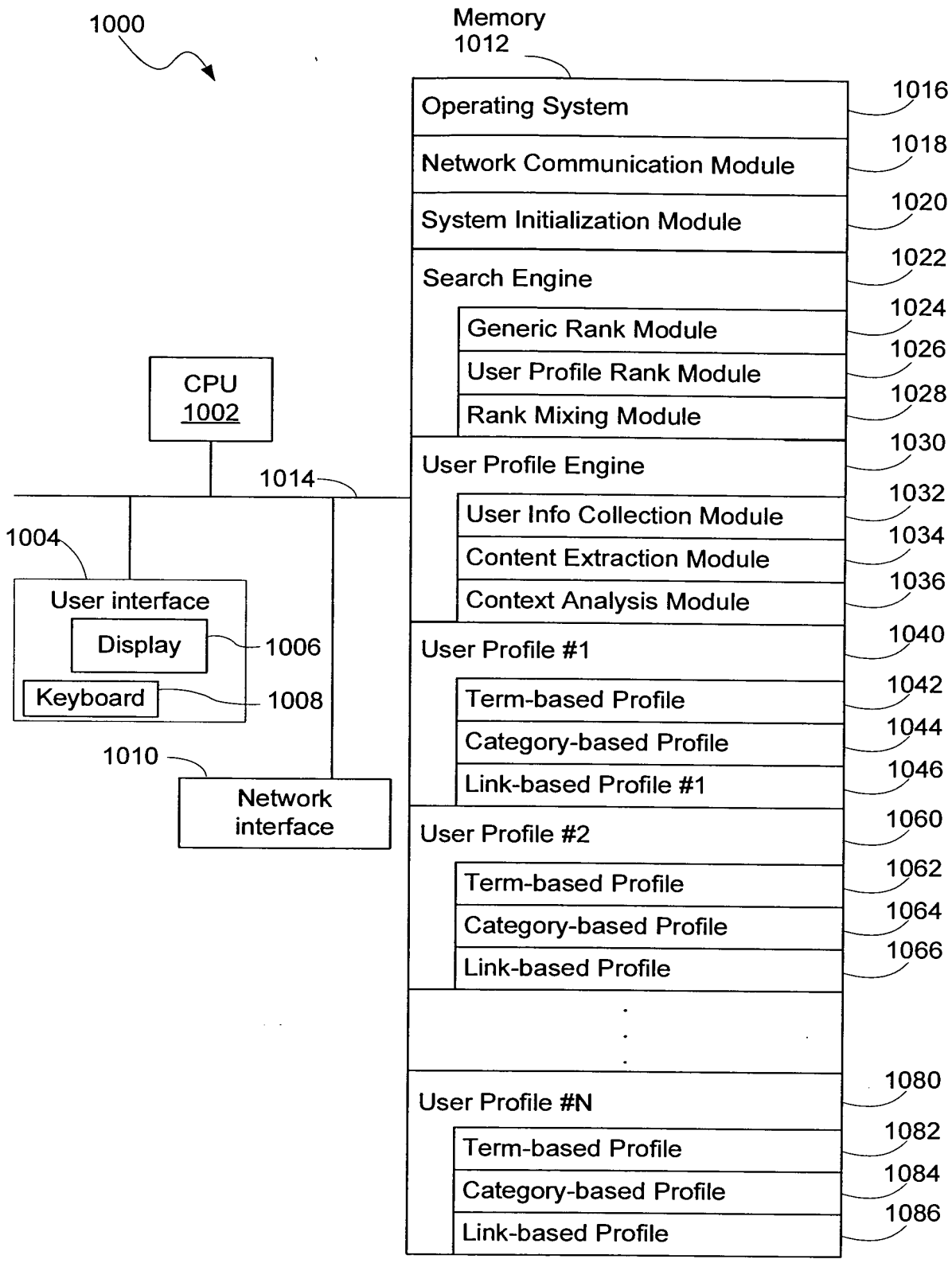

1000
Memory
1012
Operating System
1016
Network Communication Module
1018
System Initialization Module
1020
Search Engine
1022
Generic Rank Module
1024
User Profile Rank Module
1026
Rank Mixing Module
1028
User Profile Engine
1030
User Info Collection Module
1032
Content Extraction Module
1034
Context Analysis Module
1036
User Profile #1
1040
Term-based Profile
1042
Category-based Profile
1044
Link-based Profile #1
1046
User Profile #2
1060
Term-based Profile
1062
Category-based Profile
1064
Link-based Profile
1066
User Profile #N
1080
Term-based Profile
1082
Category-based Profile
1084
Link-based Profile
1086
CPU
1002
1014
1004
User interface
Display
1006
Keyboard
1008
1010
Network interface

Fig. 10

DECLARATION FOR NON-PROVISIONAL PATENT APPLICATION*

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below at 201 et seq. beneath my name.

I believe I am the original, first and sole inventor if only one name is listed at 201 below, or an original, first and joint inventor if plural names are listed at 201 et seq. below, of the subject matter which is claimed and for which a patent is sought on the invention entitled

PERSONALIZATION OF WEB SEARCH

and for which a patent application:

[x] is attached hereto and includes amendment(s) filed on (if applicable)

[ ] was filed in the United States on as Application No. (for declaration not accompanying application) with amendment(s) filed on (if applicable)

[ ] was filed as PCT international Application No. on and was amended under PCT Article 19 on (if applicable)

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

I acknowledge the duty to disclose information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations,§1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119(a)-(d) 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:

EARLIEST FOREIGN APPLICATION(S), IF ANY, FILED PRIOR TO THE FILING DATE OF THE APPLICATION				
APPLICATION NUMBER	COUNTRY	DATE OF FILING (day, month, year)	PRIORITY CLAIMED	
			YES [ ]	NO [ ]
			YES [ ]	NO [ ]

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

PROVISIONAL APPLICATION NUMBER	FILING DATE

I hereby claim the benefit under Title 35, United States Code, §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 §112, I acknowledge the duty to disclose information known to me which is material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

NON-PROVISIONAL APPLICATION SERIAL NO.	FILING DATE	STATUS		
		PATENTED	PENDING	ABANDONED

* for use only when the application is assigned to a company, partnership or other organization.

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 issuing thereon.

201	FULL NAME OF INVENTOR	LAST NAME Lawrence	FIRST NAME Stephen	MIDDLE NAME R.	
	RESIDENCE & CITIZENSHIP	CITY Mountain View	STATE OR FOREIGN COUNTRY California	COUNTRY OF CITIZENSHIP Australia	
	POST OFFICE ADDRESS	STREET 2400 West El Camino Real, #204	CITY Mountain View	STATE OR COUNTRY CA	ZIP CODE 94040
		SIGNATURE OF INVENTOR 201		DATE 9/30/03	

PATENT APPLICATION FEE DETERMINATION RECORD
Effective January 1, 2003

Application or Docket Number

CLAIMS AS FILED - PART I

	(Column 1)	(Column 2)
TOTAL CLAIMS	58	
FOR	NUMBER FILED	NUMBER EXTRA
TOTAL CHARGEABLE CLAIMS	58 minus 20=	* 38
INDEPENDENT CLAIMS	4 minus 3 =	* 1
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	375.00	OR	BASIC FEE	750.00
X$ 9=		OR	X$18=	684
X42=		OR	X84=	84
+140=		OR	+280=	
TOTAL		OR	TOTAL	1518

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=	
X42=		OR	X84=	
+140=		OR	+280=	
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=	
X42=		OR	X84=	
+140=		OR	+280=	
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=	
X42=		OR	X84=	
+140=		OR	+280=	
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.

PATENT APPLICATION SERIAL NO. _____

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

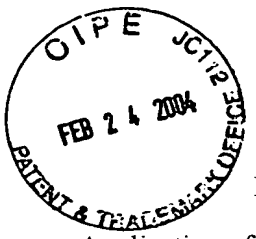
10/08/2003 JBALINAN 00000027 161150 10676711

01 FC:1001 750.00 DA
02 FC:1201 84.00 DA
03 FC:1202 684.00 DA

PTO-1556
(5/87)

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

02/26/04



Express Mail No. EV 371 773 284 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Lawrence	Confirmation No.:	8147
Serial No.:	10/676,711	Art Unit:	2171
Filed:	September 30, 2003	Examiner:	To be determined
For:	PERSONALIZATION OF WEB SEARCH	Attorney Docket No.:	60963-0014 (formerly 11378-0014-999)

INFORMATION DISCLOSURE STATEMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the duty of disclosure provisions of 37 C.F.R. §1.56, there is hereby provided certain information which the Examiner may consider material to the examination of the subject U.S. patent application. It is requested that the Examiner make this information of record if it is deemed material to the examination of the application.

1. Enclosures accompanying this Information Disclosure Statement are:

1a. ☒ A list of all patents, publications, applications, or other information submitted for consideration by the office.

1b. A legible copy of :

☐ Each U.S. patent application publication and U.S. and foreign patent;

☒ Each publication or that portion which caused it to be listed on the PTO-1449;

☐ For each cited pending U.S. application, the application specification including the claims, and any drawing of the application, or portion of the application which caused it to be listed on the PTO-1449 including any claims directed to that portion;

☒ all other information or portion which caused it to be listed on the PTO-1449.

1c. ☐ An English language copy of search report(s) from a counterpart foreign application or PCT International Search Report.

1d. ☐ Explanations of relevancy (ATTACHMENT 1(d), hereto) or English language abstracts of the non-English language publications.

2. ☒ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(b):

☐ Within three months of the filing date of a national application other than a continued prosecution application under §1.53(d);

☐ Within three months of the date of entry of the national stage as set forth in §1.491 in an international application;

☒ Before the mailing of the first Office action on the merits;

☐ Before the mailing of a first Office action after the filing of a request for continued examination under §1.114.

3. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(c) after the period specified in 37 C.F.R §1.97(b), but before the mailing date of any of a final action under 37 C.F.R. §1.113, a notice of allowance under 37 C.F.R. §1.311 or an action that otherwise closes prosecution in the application.

(Check either Item 3a or 3b)

3a. ☐ The Certification Statement in Item 5 below is applicable. Accordingly, no fee is required.

3b. ☐ The $180.00 fee set forth in 37 C.F.R. §1.17(p) in accordance with 37 C.F.R. §1.97(c) is:
☐ enclosed
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

(Item 3b to be checked if any reference known for more than 3 months)

4. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(d) after the period specified in 37 C.F.R. §1.97(c), but on or before the date of payment of the issue fee.

The Certification Statement in Item 5 below is applicable.

The $180.00 fee set forth in 37 C.F.R. §1.17(p) is:
☐ enclosed.
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

5. ☐ Certification Statement (applicable if Item 3a or Item 4 is checked)

(Check either Item 5a or 5b)

5a. ☐ In accordance with 37 C.F.R. §1.97(e)(1), it is certified that each item of information contained in this Information Disclosure Statement was first cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of this Information Disclosure Statement.

5b. ☐ Each item of information contained in this information disclosure statement was cited in a communication from a foreign patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

5c. ☐ Pursuant to 37 C.F.R. §1.704(d), each item of information contained in this information disclosure statement was cited in a communication from a foreign patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

6. ☐ This application is a continuation application under 37 C.F.R. §1.53(b) or (d).

(Check appropriate Items 6a, 6b and/or 6c)

6a. ☐ A Petition to Withdraw from issue under 37 C.F.R. §1.313(b)(5) is concurrently filed herewith.

6b. ☐ Copies of publications listed on Form PTO-1449 from prior application Serial No. , filed on , of which this application claims priority under 35 U.S.C. §120, are not being submitted pursuant to 37 C.F.R. §1.98(d).

6c. ☐ Copies of the publications listed on Form PTO-1449 were not previously cited in prior application Serial No. , filed on , and are provided herewith.

7. ☐ This is a Supplemental Information Disclosure Statement. (Check Item 7a)

7a. ☐ This Supplemental Information Disclosure Statement under 37 C.F.R. §1.97(f) supplements the Information Disclosure Statement filed on . A bona fide attempt was made to comply with 37 C.F.R. §1.98, but inadvertent omissions were made. These omissions have been corrected herein. Accordingly, additional time is requested so that this Supplemental Information Disclosure Statement can be considered as if properly filed on .

8. ☐ In accordance with 37 C.F.R. §1.98, a concise explanation of what is presently understood to be the relevance of each non-English language publication is:

(Check Item 8a, 8b, or 8c)

8a. ☐ satisfied because all non-English language publications were cited on the enclosed English language copy of the PCT International Search Report or the search report from a counterpart foreign application indicating the degree of relevance found by the foreign office.

8b. ☐ set forth in the application.

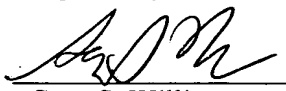
8c. ☐ enclosed as an attachment hereto.

9. ☒ The Commissioner is authorized to charge any additional fee required or credit any overpayment for this Information Disclosure Statement and/or Petition to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

10. ☒ No admission is made that the information cited in this Statement is, or is considered to be, material to patentability nor a representation that a search has been made (other than a search report of a foreign counterpart application or PCT International Search Report if submitted herewith). 37 C.F.R. §§1.97(g) and (h).

Respectfully submitted,

Date: February 24, 2004



31,066

Gary S. Williams (Reg. No.)
MORGAN, LEWIS & BOCKIUS LLP
3300 Hillview Avenue
Palo Alto, CA 94304
(650) 493-4935

INFORMATION DISCLOSURE CITATION PTO-1449				*Complete If Known*	
				Application Number	10/676,711
				Filing Date	September 30, 2003
				First Named Inventor	Lawrence
				Art Unit	2171
				Examiner Name	To be determined
Sheet	1	of	1	Attorney Docket Number	60963-0014 (formerly 11378-0014-999)

U.S. PATENT DOCUMENTS						
Examiner Initials	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate

FOREIGN PATENT DOCUMENTS							
Examiner Initials	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation	
						Yes	No
						X	

OTHER PRIOR ART - NON PATENT LITERATURE DOCUMENTS	
Examiner Initials	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published
	Haveliwala, Taher H., "Topic-Sensitive PageRank", *Proceedings of the Eleventh International World Wide Web Conference*, Honolulu, Hawaii, May 2002.
	Jeh, Glen, et al., "Scaling Personalized Web Search", Stanford University Technical Report, 2002.
	Pretschner, Alexander, et al., "Ontology Based Personalized Search", *Proc. 11th IEEE Int. Conf. on Tools with Artificial Intelligence*, Chicago, Nov. 1999, pgs. 391-398.

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

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/RECEIPT DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO.
10/676,711	09/30/2003	Stephen R. Lawrence	11378-0014-999

CONFIRMATION NO. 8147

24341
MORGAN, LEWIS & BOCKIUS, LLP.
2 PALO ALTO SQUARE
3000 EL CAMINO REAL
PALO ALTO, CA 94306

Date Mailed: 09/14/2004

Communication Regarding Rescission Of Nonpublication Request and/or Notice of Foreign Filing

Applicant's rescission of the previously-filed nonpublication request and/or notice of foreign filing is acknowledged. The paper has been reflected in the Patent and Trademark Office's (USPTO's) computer records so that the earliest possible projected publication date can be assigned.

The projected publication date is 03/31/2005.

If applicant rescinded the nonpublication request before or on the date of "foreign filing,"[1] then no notice of foreign filing is required.

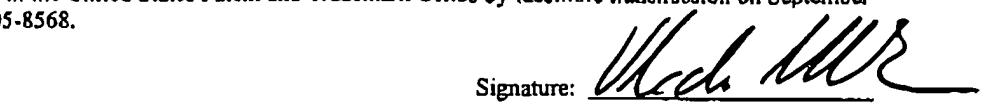
If applicant foreign filed the application after filing the above application and before filing the rescission, and the rescission did not also include a notice of foreign filing, then a notice of foreign filing (not merely a rescission) is required to be filed within 45 days of the date of foreign filing. See 35 U.S.C. § 122(b)(2)(B)(iii), and Clarification of the United States Patent and Trademark Office's Interpretation of the Provisions of 35 U.S.C. § 122(b)(2)(B)(ii)-(iv), 1272 Off. Gaz. Pat. Office 22 (July 1, 2003).

If a notice of foreign filing is required and is not filed within 45 days of the date of foreign filing, then the application becomes abandoned pursuant to 35 U.S.C. § 122(b)(2)(B)(iii). In this situation, applicant should either file a petition to revive or notify the Office that the application is abandoned. See 37 CFR 1.137(f). Any such petition to revive will be forwarded to the Office of Petitions for a decision. Note that the filing of the petition will not operate to stay any period of reply that may be running against the application.

Questions regarding petitions to revive should be directed to the Office of Petitions at (703) 305-9282. Questions regarding publications of patent applications should be directed to the patent application publication hotline at (703) 605-4283 or by e-mail pgpub@uspto.gov.

[1] Note, for purpose of this notice, that "foreign filing" means "filing an application directed to the same invention in another country, or under a multilateral international agreement, that requires publication of applications 18 months after filing".

I certify that this paper is being filed with the United States Patent and Trademark Office by facsimile transmission on September 14, 2004, to facsimile number (703) 305-8568.

Signature: 

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Lawrence	Confirmation No.:	8147
Serial No.:	10/676,711	Art Unit:	2171
Filed:	September 30, 2003	Examiner:	To be determined
For:	PERSONALIZATION OF WEB SEARCH	Attorney Docket No.:	60963-0014

REQUEST TO RESCIND PREVIOUS NON-PUBLICATION REQUEST
(35 U.S.C. 122((b)(2)(B)(ii)) AND NOTICE OF FOREIGN FILING (35 U.S.C. 122(b)(2)(B)(iii))

Mail Stop PG Pub
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

A request that the above-identified application not be published under 35 U.S.C. 122(b) (nonpublication request) was include with the above-identified application on filing pursuant to 35 U.S.C. 122(b)(2)(B)(i). I hereby **rescind** the previous nonpublication request.

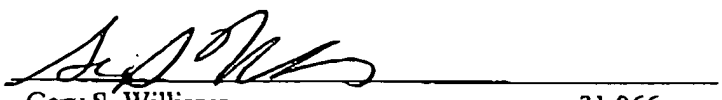
If a notice of foreign or international filing is or will be required by 35 U.S.C. 122(b)(2)(B)(iii) and 37 CFR 1.213(c), I hereby provide such notice. This notice is being provided no later than forty-five (45) days after the date of such foreign or international filing.

If a notice of subsequent foreign or international filing required by 35 U.S.C. 122(b)(2)(B)(iii) and 35 CFR 1.213(c) was not filed within forty-five (45) days after the date of filing of the foreign or international application, the application is ABANDONED, and a petition to revive under 35 CFR 1.137(b) is required. See 35 CFR 1.137(f).

No fee is believed to be due with this request; however, the Commissioner is authorized to charge any requisite fee to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no. 60963-0014). A copy of this sheet is attached for such purpose.

Respectfully submitted,

Date: 9/13/04



Gary S. Williams
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, CA 94306
(650) 843-4000

31,066
(Reg. No.)

Morgan, Lewis & Bockius LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, CA 94306
TEL: 650.843.4000
FAX: 650.843.4001
eFax: 877.432.9652
www.morganlewis.com

Morgan Lewis
COUNSELORS AT LAW

SEND TO

Name:		Firm:	USPTO
FAX Number:	703 305 8568	Telephone Number:	

FROM

Name:	Vladimir Skliba	Date Sent:	September 14, 2004
Operator Sending:		Telephone Number:	650 843 7573
FAX Number:	650 843 4001	Floor: 8	Number of Pages: 2 *(including cover page)*

FAX MESSAGE

COMMENTS

Application of: Lawrence
Application No.: 10/676,711
Filing Date: September 30, 2003
Applicant's File Ref: 60963-0014US
Title: PERSONALIZATION OF WEB SEARCH

Enclosed is:

(1) Request to Rescind Previous Non-publication Request and Notice of Foreign Filing

DO NOT SEND TO CLIENTS | TEAR ALONG DOTTED LINE | THIS PORTION FOR BILLING PURPOSES ONLY

Date Sent:		Attorney:	Vladimir Skliba	ID: 64049	
Time:		Client:	Google, Inc.	Client/Matter Number:	060963-0014
(1) Fax No.:	703 305 8568				
Receipt Confirmed:		Number of Pages: *(including cover page)*	2	☐ Firm Charge *(check box if applicable)*	[PA-25]

3-7-05

IFW

Express Mail No. EV 534 876 651 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Stephen R. Lawrence	Confirmation No.:	8147
Serial No.:	10/676,711	Art Unit:	2171
Filed:	September 30, 2003	Examiner:	To be determined
For:	*Personalization of Web Search*	Attorney Docket No.:	60963-0014-US

INFORMATION DISCLOSURE STATEMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the duty of disclosure provisions of 37 C.F.R. §1.56, there is hereby provided certain information which the Examiner may consider material to the examination of the subject U.S. patent application. It is requested that the Examiner make this information of record if it is deemed material to the examination of the application.

1. Enclosures accompanying this Information Disclosure Statement are:

1a. ☒ A list of all patents, publications, applications, or other information submitted for consideration by the office.

1b. A legible copy of :

☐ Each U.S. patent application publication and U.S. and foreign patent;

☒ Each publication or that portion which caused it to be listed on the PTO-1449;

☐ For each cited pending U.S. application, the application specification including the claims, and any drawing of the application, or portion of the application which caused it to be listed on the PTO-1449 including any claims directed to that portion;

☒ all other information or portion which caused it to be listed on the PTO-1449.

1c. ☒ An English language copy of search report(s) from a counterpart foreign application or PCT International Search Report.

1d. ☐ Explanations of relevancy (ATTACHMENT 1(d), hereto) or English language abstracts of the non-English language publications.

2. ☒ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(b):

☐ Within three months of the filing date of a national application other than a continued prosecution application under §1.53(d);

☐ Within three months of the date of entry of the national stage as set forth in §1.491 in an international application;

☒ Before the mailing of the first Office action on the merits;

☐ Before the mailing of a first Office action after the filing of a request for continued examination under §1.114.

3. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(c) after the period specified in 37 C.F.R §1.97(b), but before the mailing date of any of a final action under 37 C.F.R. §1.113, a notice of allowance under 37 C.F.R. §1.311 or an action that otherwise closes prosecution in the application.

(Check either Item 3a or 3b)

3a. ☐ The Certification Statement in Item 5 below is applicable. Accordingly, no fee is required.

3b. ☐ The $180.00 fee set forth in 37 C.F.R. §1.17(p) in accordance with 37 C.F.R. §1.97(c) is:
☐ enclosed
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

(Item 3b to be checked if any reference known for more than 3 months)

4. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(d) after the period specified in 37 C.F.R. §1.97(c), but on or before the date of payment of the issue fee.

(Check either Item 4a or 4b)

4a. ☐ The Certification Statement in Item 5 below is applicable.

4.b ☐ The $180.00 fee set forth in 37 C.F.R. §1.17(p) is:
☐ enclosed.
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

5. ☐ Certification Statement (applicable if Item 3a or Item 4a is checked)

(Check either Item 5a, 5b or 5c)

5a. ☐ In accordance with 37 C.F.R. §1.97(e)(1), it is certified that each item of information contained in this Information Disclosure Statement was first cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of this Information Disclosure Statement.

5b. ☐ Each item of information contained in this information disclosure statement was cited in a communication from a foreign patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

5c. ☐ Pursuant to 37 C.F.R. §1.704(d), each item of information contained in this information disclosure statement was cited in a communication from a foreign

patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

6. ☒ Copies of each cited U.S. patent and each U.S. patent application publication are not enclosed pursuant to the USPTO OG Notice dated 05 August 2003 waiving the requirement under 37 C.F.R. 1.98(a)(2)(i) for U.S. patent applications filed after June 30, 2003.

7. ☐ This application is a continuation application under 37 C.F.R. §1.53(b) or (d).

(Check appropriate Items 7a, 7b and/or 7c)

7a. ☐ A Petition to Withdraw from issue under 37 C.F.R. §1.313(b)(5) is concurrently filed herewith.

7b. ☐ Copies of publications listed on Form PTO-1449 from prior application Serial No. , filed on , of which this application claims priority under 35 U.S.C. §120, are not being submitted pursuant to 37 C.F.R. §1.98(d).

7c. ☐ Copies of the publications listed on Form PTO-1449 were not previously cited in prior application Serial No. , filed on , and are provided herewith.

8. ☐ This is a Supplemental Information Disclosure Statement. (Check Item 8a)

8a. ☐ This Supplemental Information Disclosure Statement under 37 C.F.R. §1.97(f) supplements the Information Disclosure Statement filed on . A bona fide attempt was made to comply with 37 C.F.R. §1.98, but inadvertent omissions were made. These omissions have been corrected herein. Accordingly, additional time is requested so that this Supplemental Information Disclosure Statement can be considered as if properly filed on .

9. ☐ In accordance with 37 C.F.R. §1.98, a concise explanation of what is presently understood to be the relevance of each non-English language publication is:

(Check Item 9a, 9b, or 9c)

9a. ☐ satisfied because all non-English language publications were cited on the enclosed English language copy of the PCT International Search Report or the search report from a counterpart foreign application indicating the degree of relevance found by the foreign office.

9b. ☐ set forth in the application.

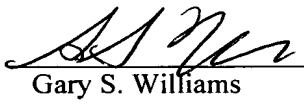
9c. ☐ enclosed as an attachment hereto.

10. ☒ The Commissioner is authorized to charge any additional fee required or credit any overpayment for this Information Disclosure Statement and/or Petition to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no. 60963-0014-US).

11. ☒ No admission is made that the information cited in this Statement is, or is considered to be, material to patentability nor a representation that a search has been made (other than a search report of a foreign counterpart application or PCT International Search Report if submitted herewith). 37 C.F.R. §§1.97(g) and (h).

Respectfully submitted,

Date: March 3, 2005



31,066

Gary S. Williams (Reg. No.)
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, CA 94306
(650) 843-4000

INFORMATION DISCLOSURE CITATION PTO-1449				*Complete If Known*	
				Application Number	10/676,711
				Filing Date	September 30, 2003
				First Named Inventor	Stephen R. Lawrence
				Art Unit	2171
				Examiner Name	To be determined
Sheet	1	of	1	Attorney Docket Number	60963-0014-US

U.S. PATENT DOCUMENTS

Examiner Initials	Cite No.	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate
		6,285,999	09-04-2001	Page	707	5	
		6,327,590	12-04-2001	Chidlovskii et al	707	5	
		2002/0024532 A1	02-28-2002	Fables et al.	345	700	
		2002/0198882 A2	12-26-2002	Linden et al.	707	10	
		2003/0233345 A1	12-18-2003	Perisic et al.	707	3	

FOREIGN PATENT DOCUMENTS

Examiner Initials	Cite No.	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation Yes	Translation No
		EP 1050830A2	11-08-2000	Xerox Corp.				
		WO 03/107127A2	12-24-2003	Entopia, Inc.				

OTHER NON PATENT LITERATURE DOCUMENTS

Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published
		Sergey Brin et al., "The Anatomy of a Large-scale Hypertextual Web Search Engine", *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, April 1998, pgs. 107-117.
		Junghoo Cho et al., "Efficient crawling through URL ordering", *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, April 1998, pgs. 161-171.

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

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

(11) EP 1 050 830 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
08.11.2000 Bulletin 2000/45

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

(21) Application number: **00303613.4**

(22) Date of filing: **28.04.2000**

(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: **05.05.1999 US 305435**

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

(72) Inventors:
- **Chidlovski, Boris**
  38240 Meylan (FR)
- **Glance, Natalie S.**
  38240 Meylan (FR)
- **Grasso, Antonietta**
  38000 Grenoble (FR)

(74) Representative:
Skone James, Robert Edmund
GILL JENNINGS & EVERY
Broadgate House
7 Eldon Street
London EC2M 7LH (GB)

(54) System and method for collaborative ranking of search results employing user and group profiles

(57) A system for ranking search results obtained from an information retrieval system includes a search pre-processor (30), a search engine (20) and a search post-processor (40). The search pre-processor (30) determines the context of the search query by comparing the terms in the search query with a predetermined user context profile. Preferably, the context profile is a user profile or a community profile, which includes a set of terms which have been rated by the user, community, or a recommender system. The search engine generates a search result comprising at least one item obtained from the information retrieval system. The search post-processor (40) ranks each item returned in the search result in accordance with the context of the search query.

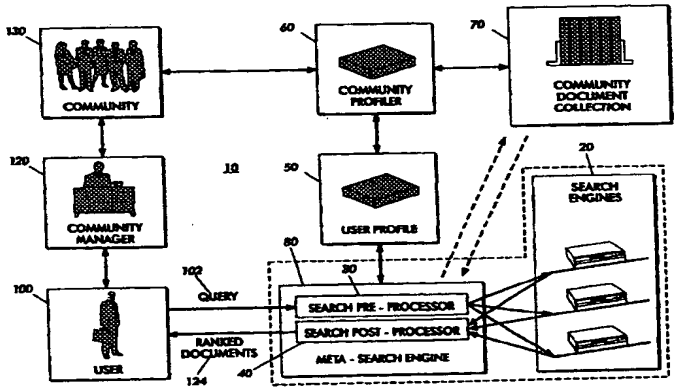


FIG. 2

EP 1 050 830 A2

Description

[0001] This invention relates generally to information retrieval systems and more particularly, to a system and method of collaboratively ranking results returned from a search engine using user and group profiles.

[0002] The World Wide Web (the "web" or "WWW") is an architectural framework for accessing documents (or web pages) stored on a worldwide network of distributed servers called the Internet. Documents stored on the Internet are defined as web pages. The architectural framework of the web integrates web pages stored on the Internet using links. Web pages consist of elements that may include text, graphics, images, video and audio. A web page, which points to the location of another web page, is said to be linked to that other web page. Links that are set forth in a web page usually take the form of a text fragment or an image. A user follows a link by selecting it.

[0003] With the advent of networking technology and the World Wide Web, the ability to access information from external sources has greatly increased. Various search engines enable a user to submit a query, which returns a collection of items or documents. A well-crafted query may return a manageable set of documents, typically from 30 to 50 documents. A less narrow query may return over 1000 documents. An overly narrow query may return no documents (in which case no ranking is required). Various techniques are available for assisting the user in refining or narrowing his/her search query. However, once the search result has been properly narrowed, a significant problem in information retrieval is how to rank the results returned by the search engine or the combination of search engines.

[0004] For individual search engines, there are many different techniques for ranking results, ranging from counting the frequency of the appearance of the various search terms in the search query to calculating vector similarities between a search term vector and each returned document vector. In a networked environment such as the World Wide Web, meta-searchers access different and often heterogeneous search engines and face the additional difficulty of combining the ranking information returned by the individual engines. Meta-searcher is a Web information retrieval system aimed at searching answers to a user's query in the heterogeneous information providers distribute over the Web. When a meta-searcher receives responses (usually in the form of HTML files) from the information providers, a special component of a meta-searcher called a wrapper, process the responses to answer the original query. Since many search engines, including meta-searchers, hide the mechanism used for document ranking, the problem of merging search results is compounded. A problem common to both individual search engines and meta-search engines is that these approaches ignore, or knowing nothing about, the user conducting the search, or the user's context for conducting the search.

[0005] Relevance feedback is one approach that elicits information about the user and his/her search context. Relevance feedback techniques re-rank the search results by using user feedback to recalculate the relative importance of key words in the query. While powerful from a technical point of view, relevance feedback approaches suffer from user interface issues. The relevance information required is often difficult to elicit successfully from users during the search process. U.S. Patent No. 4,996,642 to Hey, System and Method for Recommending Items, describes a system for providing recommendations to users based on others items previously sampled by the user and the availability of the items.

[0006] Knowledge Pump, a Xerox system, provides community-based recommendations by initially allowing users to identify their interests and "experts" in the areas of those interests. Knowledge Pump is then able to "push" relevant information to the users based on those preferences. This is accomplished by monitoring network traffic to create profiles of the users, including their interests and "communities of practice," thus refining the community specifications. However, monitored or automatically created profiles for establishing context may not accurately reflect the user's context at all times.

[0007] There is a need for a system and method of ranking search results which does not require user solicited relevance information. There is also a need for a system of ranking search results which takes into account a predetermined user context profile. There is also a need for a system and method of ranking search results which ranks results based on a user selected context. There is also a need for a system and method of ranking search results which takes into account a group or community to which the user belongs. There is a further need for a system and method of creating a user and community profile for ranking search results.

[0008] A system for ranking search results obtained from an information retrieval system, according to the invention, includes a search pre-processor, a search engine and a search post-processor. The search pre-processor, responsive to a search query, determines the context of the search query by comparing the terms in the search query with a predetermined user context profile. The user's context profile may include, for example, the user's identity, the community or set of communities applicable to the search, and the point of view the user wishes to adopt (e.g., that of a domain expert) for the search. Preferably, the context profile is a user profile or a community profile, which includes a hierarchical set of terms that have been rated by the user or community. Also, a recommender system may be used to generate the user or community context profile.

[0009] The search engine, responsive to the search query, generates a search result comprising at least one item obtained from the information retrieval system. (If

no items are returned, such as when the search is overly narrow, no ranking is required.) Generally, a great number of items will be generated, which the search engine will provide in its own predetermined form of hierarchical valuation. The search post-processor, responsive to a non-empty search result, ranks each item returned in the search result in accordance with the context of the search query. The ranked results may then be provided or displayed in any normal fashion, such as on a computer display or printed out. If more than one search engine is used, each search engine returns its own list of search results. The post-search processor then ranks all items returned, regardless of search engine, in accordance with the context of the search query.

[0010] A method of ranking search results obtained from an information retrieval system, according to the invention, includes providing a predetermined user context profile, generating a search query, and applying the context profile to the search query to generate a context of the search query. A search is then performed based on the search query, which includes at least one item obtained from the information retrieval system. The search results are then ranked in accordance with the context of the search query.

[0011] The system and method according to the invention couples a predetermined user context profile (e.g., user profiling, community profiling or recommender profiling) with the search process. By coupling context profiling with the search process, search results are no longer an isolated event, but are ranked within the context of a particular user or community or recommender system point of view. Depending on the user's context for the search, a different predetermined context profile may be selected, thus customizing the ranking of each particular search.

[0012] The user and community profiles are built by analyzing document collections put together by the users and the communities to which the users belong, If any of the retrieved search results are considered relevant to the user or the community, they can be used to tune or modify the particular user or community profile by re-weighting the profile terms.

[0013] User and community profiling is particularly useful in the invention. First, the post-processor uses a particular context profile (either the user's profile, the community's profile or another user's profile - such as a known expert in a domain outside the user's expertise) to rank the results of a search query. Preferably, a user profile is build from a user selecting a particular document collection and ranking or rating the various terms within the document collection. The user profile becomes the document collection with rating information attached to each document. A user can have more than one user context profile, or use another user's context profile in order to rank the search results most expeditiously according to a particular point of view. The ranked search results can be used to update the user profile based on new submissions or documents produced in the search and ranked using the user's context. This approach is similar to relevance feedback.

[0014] Similarly, a community (i.e., a group of users having similar interests) profile can be built by categorizing the documents in a document collection into the communities (when such a construct exists in the document collection) and then ranking the various documents according to the particular users belonging to the community. A user's ability to rank documents within the community will vary according to his/her levels of expertise. Various methods to approximate a user's level of expertise within a community can be used (e.g., by agreement, by statistics, etc.).

[0015] The system of the invention provides an architecture that allows these methods to work together in support of community-based relevance feedback. The system and method of the invention provide the ability to rank results returned across multiple search engines and the ability to take into account the user's context through use of user, community or expert user profiles.

Figure 1 illustrates an example of a distributed operating environment for performing the present invention;

Figure 2 is a block diagram of a system for ranking search results obtained from an information retrieval system in accordance with a predetermined context profile.

[0016] Referring now to the figures, Figure 1 illustrates an example of a distributed operating environment for performing the present invention. In the distributed operating environment illustrated in Figure 1, client computers 102, request searches, communicate with other client computers and retrieve documents (i.e., web pages) stored on servers 104 for either viewing, storing, or printing. The client computers 102 are coupled to the servers 104 through Internet 106. Some client computers 102, which are located on an Intranet 110, communicate indirectly with servers 104 located on the Internet 106 through a proxy server 112. The client computers 102 may consist of either workstations 114 or laptops 116. Alternatively, the client computers 102 may request searches, communicate with other client computers and retrieve documents (i.e., web pages) stored on Intranet servers such as proxy server 112 for either viewing, storing, or printing.

[0017] Referring to Figure 2, a system for ranking search results obtained from an information retrieval system according to a predetermined user context profile is generally shown therein and referred to by numeral 10. System 10 includes search pre-processor 30, which takes a query 102 from a user 100 and applies a predetermined user context profile to determine the context of the search query. The user context

profile may be a user profile generated by user profiler 50 or a community profile generated by a community profiler 60. Results from the search query, which generally include a plurality of hierarchically-ranked search results based on the query, are returned by the various search engines 20, or meta-search engine 80 by searching an information retrieval system (such as the Internet). These search results are then ranked by the search post-processor 40 and provided to the user in the form of ranked documents 124.

[0018] Community profiler 60 ranks community document collection 70 in accordance with evaluations or rankings determined by the members of community 130. In some cases, community manager 120 may determine from time to time whether a particular user may join or continue to be a part of the community. User profiler 50 ranks a selected document collection (which may also be the community document collection 70) in accordance with evaluations or ratings by user 100.

[0019] The system may be extended to support community-based relevance feedback. In addition to the search pre-processor, one or more search engines or meta-search engines and search post-processor, the extended system may include one or more document collections with associated user, community/group, and rating attributes, a user profiler, a community profiler and a community manager. Additionally, the extended system may include wrappers that allow the profilers to extract document content (or document reference, such as its URL), user, community and rating information from the document collections and wrappers that allow the search pre-processor to submit queries to the search engine and the search post-processor to extract the results.

[0020] The document collection may be one (or a combination) of several different types: documents residing in a document management system or a file system or documents referenced by a recommender system. In each case, the document collection provides a specific methodology for associating content with users and potentially with communities of users. In each case, the document collection provides the basis for establishing the user context profile, in that the document collection and user ratings establish the environment or the interrelated conditions under which the user desires to rank search results.

[0021] Preferred document collections include those provided by community recommender systems which attach user identification, community categorizations and user ratings to the documents. Using document collections generated by community recommender systems allows use of the most sophisticated of the user and community profiling techniques described below. Preferably, a community-based relevance feedback system includes a recommender system as one of the document collections, or preferably as the principal one.

[0022] An important aspect of the system for ranking is the document collection used to generate the context profile. The document collection may include an application program interface (API) for allowing the profilers to query for all documents submitted and/or reviewed by a user (who may be associated with a particular community). If such an API is not provided, then a wrapper suitable for extracting the information may be used. A wrapper is a tool used by a meta-searcher that scans the HTML files returned by the search engine, drops the markup instructions and extracts the information related to the query. Then the wrapper takes the answers from the different providers, puts them in a new format and generates an HTML file that can be viewed by the user. The API or the wrapper generates "metadata" which is used by the profilers to construct and to incrementally update the user and community profiles from the set of documents relevant to the user and in the context of the community. In the case of standard document collections, (such as file systems or document management systems), it is generally assumed that any document filed or stored by the user is relevant to the user. In the case of a recommender system, it is generally assumed that any document submitted or reviewed by a user with an average or higher rating is relevant.

[0023] The search engine may include an API for submitting a search and retrieving results. If not, a suitable wrapper may be used. The problem of query translation across multiple, heterogeneous search engines and databases and the extraction of the search results is well known. Thus, any commercially available translation and extraction product may be used. However, it should be noted that search engines do not necessarily cover the documents in the system's document collections, although overlap is always possible.

[0024] The search pre-processor determines the context of the user's search. The search pre-processor applies a predetermined context profile to search query. For example, the context profile may include the user's identity, the community or set of communities appropriate to the search, and the point of view the user wishes to adopt for the search, if any (such as that of another user or a domain expert). The context profile can be retrieved explicitly by asking the user to identify him/herself and by asking the user to select the appropriate community or communities and/or point of view. Context can also be determined (deduced) automatically by matching the query with a query memory associated with a community (if selected) or the collection of users using the system.

[0025] A preferred context profile is that of the user. The user profile is created or generated by the user profiler, which constructs a term-weight vector for each user which is extracted from the set of documents submitted and/or reviewed into each of the document collections to which the user participates. Matching a user across several different document collections is not always simple. One method of accomplishing this is to ask the user to provide his/her identifier (and password,

if needed) for each document collection. If a user withholds some of this information, then his/her profile will be less complete than for other users who do not withhold this information. However, this is not always the case as a user may choose only to provide access to particular document collections deemed appropriate by the user. This problem only occurs is more than one document collection is used in the system, which is preferably the one provided by the community recommender system.

[0026] The term-weight vector or user profile P^u is calculated in a standard way, although various linguistic-based enhancements are possible as noted below. For a user, u, the vector includes the set of terms $\{t_i\}$ with their weights w^u_i, $P^u = \{t_i, w^u_i\}$. If the term-weight vector is at least calculated in part from documents that have been evaluated (implicitly or explicitly rated) in some way by the user, then the ratings given to the documents can be used to bias the term-weight vector.

[0027] The user profiler may also calculate the profile of the user in the context of a community or a specific domain or domains. In this case, the user profiler would take into account only those documents submitted or reviewed by the user and classified (either by the user or automatically) into the domain. An added difficulty in this case is matching communities/domains across document collections. Again, if there is only a single document collection, the recommender system, this difficulty disappears. The user profiler provides an API that returns a term-weight vector in response to a user identification and possibly a community/domain identifier.

[0028] Similarly, the community profile P^c is created or generated by a community profiler. The community profiler constructs a term-weight vector for each community, which is extracted from the set of documents classified into a community within each of the document collections. The term-weight vector for the community is determined in a way analogous to that employed for users. The community vector contains the set of terms $\{t_i\}$ with their weights w^c_i, $P^c = \{t_i, w^c_i\}$. The weight of each term is calculated from the weights w^u_i of the individual community members (users). Since the contributions of the members are frequently much different from one another, the community profile can be biased to weigh more heavily the contribution of "experts" in the community (special users). Experts are those community members whose recommendations are most frequently followed by the whole community. Formally, each member u in the community is assigned a weight α_u. Experts have the highest α_u and for the whole community:

$$\sum_u \alpha_u = 1.$$

The individual α_u must be re-normalized whenever a user enters or leaves the community. Then, the weight of term t_i in the community profile is evaluated as:

$$w^c_i = \sum_u \alpha_u w^u_i,$$

where w^u_i is the weight of t_i in the profile of user u. Beyond the community and personal (user) profile, the user can request the profile of the community expert(s), which contains weight w^{exp}_i for each profile term t_i. The community profiler provides an API that returns a term-weight vector in response to a community identifier.

[0029] When registering a new user u to community, the initial user profile $P^u = (t_i, w_i)$ can be one of the following options: the community profile (t_i, w_i); a list of user defined keywords t_i (the weights w_i are equal or induced from community profile); or empty. Any document reviewed or submitted by the user changes her/his profile as follows.

[0030] If (a new) document D is submitted, all terms and their associated weights are extracted from the document. k top-weighted terms are then selected such that a document profile is created: $D = (t'_i, w'_i)$, for i= 1,...,k, where k is a system/application-dependent constant. Each document D reviewed and in the document collection has its own profile $D = (t'_i, w'_i)$. Otherwise, for a reviewed document, its profile $D = (t'_i, w'_i)$ is retrieved from the repository where it is stored along with the document.

[0031] The current user profile vector $P^u=(t_i, w_i)$ and new document profile $D=(t'_i, w'_i)$, are used as follows to update the user profile. For each term in set $\{t_i \cup t'_i\}$, we evaluate

$$w^{new}_i = \gamma \times w_i + (1-\gamma) \times w'_i,$$

where γ is a "profile conservativeness" constant, $0 < \gamma < 1$. The closer γ to 1, the slower the profile changes with new submissions. Practical values of γ are in range [0.5,0.95] and can depend on the number of user submissions (over last m days).

[0032] Only the k top-weighted terms t^{new}_i are chosen for the new user profile P^u and their weights get normalized:

$$\Sigma w^{new}_i = 1.$$

[0033] When creating a community, the community administrator (which can be a human or a software program) can use for the initial community profile $P^c=(t_i, w_i)$ one of the following options: process sample document(s) relevant to the community and extract terms and weights as with a user submission described above; use a list of community keywords t_i given by

administrator, or leave it empty. Any document reviewed or submitted by a community member changes the member profile. Beyond the member profiles, the community profiler maintains values of contribution α_u for each community member u.

[0034] For user u in the community, its contribution α_u can be evaluated as

$$v \cdot r_u / f_u,$$

where r_u is a number of documents submitted by the user (over the last m days) and f_u is the number of community users that followed those recommendations. v is a customization coefficient; it may favor a user with numerous, but moderate, recommendations rather than a user with one but a popular recommendation. The user with the highest value α_u is called the community expert and his/her profile can be used as an expert profile P^{exp} by other users for re-ranking the search results. Values α_u are kept normalized such that $\Sigma\alpha_u = 1$. Optionally, experts can be chosen or assigned by community members without statistical evaluation.

[0035] For all members of the community, their profiles $P^u = (t_i{}^u, w_i{}^u)$ and their contributions α_u are used for updating the community profile. For each term t_i in

$$\left\{\bigcup_u t_i^u\right\},$$

its weight is

$$w_i^c = \sum_u \alpha_u \times w_i^u.$$

The community profile can keep all terms from its members' profiles or only the k top-weighted terms; in either case, their weights are kept normalized: $\Sigma w_i^c = 1$.

[0036] The update of the community profile P^c is performed preferably when a minimally required number of user profile changes have occurred. The community profile update is processor time-consuming; thus it is preferably to update the profile off-line.

[0037] In a preferred embodiment, the system according to the invention incorporates a recommender system. In addition to the storing of document profiles, user profiles community profiles and expert profiles, a recommender system would also include or provide tools for profile retrieval and profile updates.

[0038] Matching community definitions across document collections and maintaining a coherent list of communities and users participating in those communities is frequently difficult. If a community recommender system exists within the system, then its list of communities is a likely candidate for adoption for the community relevance feedback system. Alternatively, an administrator of the system could be responsible for matching groups or collections in other document collections with the community list The task of constructing such a list from scratch would fall to the administrator in the absence of such a list. It is also possible to create automatic methods of performing the matching, although this may possibly reduce the accuracy of the community profiles. It is possible that the way in which the community list is constructed and matched across collections that the end result will be that the community profile is entirely determined from the data in the community recommender system, if such exists, or other document collection with a notion of the community.

[0039] It should be noted that a community profile is not required in order to practice the system and method of the invention. In some instances, it may be appropriate to take into account only the user's context, in the absence of the community, although such a system will be more difficult to accept and add new users.

[0040] A method of ranking search results obtained from an information retrieval system using a predetermined user context profile would include the following steps. Before a user starts formulating queries, because of the greater benefits available from a recommender system, the user is assumed to have registered with a recommender system. This permits the system to upload the user profile, community profile and/or expert profiles chosen by the users for search result ranking and profile re-weighting.

[0041] Once a user has formulated a query, the search pre-processor takes the query and processes the keywords in the query to a query profile $P^q = (t_i{}^q, w_i{}^q)$. This profile is used by the search post-processor later.

[0042] When the query is submitted to a search engine and the search result returns, the user sees the documents determined from the search query, listed or ranked in accordance with the algorithm provided by the particular search engine(s), if any. Since this ranking will likely not rank the results within the context desired by the user, the user can request re-ranking. Alternatively, the re-ranking by the search post-processor can be automatic.

[0043] The search post-processor takes as input the list of search results returned by the search engines. It has two preferred ways to evaluate the relative rank of the documents. One is by matching the document or its pointer (e.g., its URL) to one already existing in one of the document collections. For example, if the document has been rated by a community recommender system connected to the system within the appropriate community context, then this rating is given a high weight in determining the relative rank of the result.

[0044] The second preferred means of evaluating the relative ranking of documents is by using the profile term-weight vectors as a source of relevance feedback. Depending on the context (i.e., user, community or expert), the appropriate profile is requested.

[0045] Each document in the search result is downloaded, i.e., the full contents of each document is downloaded for term extraction. The extraction is similar to that for submitted documents when a user context profile is being created from a document collection as described above. The search post-processor returns a document profile $P^d = (t^d{}_i, w^d{}_i)$ for each document d that contains k' top-weighted terms. Generally, a document in a search response is not considered as important as a submitted document in the document collection, thus the number k' of terms chosen for the search returned document profile may be less than that for a submitted document. If a document in the search result is already in the document collection, its profile is not extracted from the document but is up-loaded from the document storage.

[0046] Once a profile $P^d = (t^d{}_i, w^d{}_i)$ is obtained for each document d in the search response, the relevance (or ranking) of document d with respect to the chosen user/community/expert profile is obtained by the formula:

$$relevance(d) = \frac{\sum_i w_i^d \cdot w_i^{prof} \cdot w_i^q}{\sqrt{\sum_i (w_i^{prof} \cdot w_i^d)^2}}$$

where w_i^{prof} are term weights in the selected user, community, or expert profile. Then, documents are sorted based on their calculated rank values and presented to user.

[0047] Since the profile-based document re-ranking takes some time (needed for the documents down-loading, term extraction and rank calculation), the user may request for re-ranking, switch to other activity (or continue search) and return back to re-ranked results later. Alternatively, the user may request persistent queries, when user queries are executed off-line.

[0048] Note that the term weights $w^d{}_i$ are evaluated from the document d content only. Since a response can actually be a brief description of an original document, the term weights $w^d{}_i$ in this case may be quite different from the term weights $w^d{}_i$ in the case when the full document is available. Although the relevance ranking is then biased by the profile (through the term w_i^{prof}), there is a standard tradeoff between the length of the response documents and the quality of the ranking: the longer the document, the more precise the ranking (and the longer it takes to perform the ranking). After each document in the search result is ranked, the results are displayed in ranked order to the user.

[0049] Either kind of user search (with or without re-ranking) can lead to finding documents relevant to the community. Such documents submitted by the user after the search may be used to change or modify the user profile and consequently the community profile, as discussed above.

[0050] Not all search engines are necessarily external to the search. An internal document collection can also be searched with a query. The post-processing in this case is simpler. Indeed, the term frequency vector for a document can be extracted a prior, stored along with the document in the collection and reused in the relevance ranking each time the document fits a query, thus reducing the ranking time.

[0051] If a user provides a positive feedback to the search result or documents retrieved over the process, the search results can be included in the document collection like any other recommendation. Additionally, the search results can be used to modify the user profile by re-weighting term weights. In such a case, the query terms and/or most frequent terms in the response form a set {*Rel*} of relevant terms. Using this set, the standard Rocchio formula for relevance feedback can be used. The main difference between the approach described herein and standard relevance feedback is that the approach of the invention does not take into account non-relevant terms since this approach does not have a reliable way of extracting this kind of information from both document collections and search results. As a result of re-weighting, the relevant terms from {*Rel*} have their weights increased in the user profile.

[0052] The search post-processor generally requires textual content in order to evaluate the comparative relevance of the returned items in the context of a given user or community profile. This means downloading either an abstract (if available) or the entire document. Th entire process could become quite lengthy especially if the number of documents returned by the query is large. A first step of prefiltering may be necessary in order to prune the list to a manageable number. However, while the time cost is high, it should be remembered that the time cost to the user of evaluating the returned documents him/herself is even higher. In many cases, users may be willing to turn collaborative ranking on, return to other work at hand, and wait until an alert indicating the collaborative ranking process has terminated. As a further incentive to use the collaborative ranking feature, the items downloaded in the process can be cached locally, so that subsequent browsing by the user will be much less time-consuming.

[0053] Document content comes in many formats. In order to operate across as many formats as possible, the search post-processor will need to be able to connect to other modules that transform content format into the search format. Preferably, all formats will be transformed into ASCII format. Some documents may fall outside the system's ability to rank them. The system will need to distinguish these from those documents that

are ranked in a way meaningful to the user.

[0054] It will be appreciated that the present invention may be readily implemented in software using software development environments that provide portable source code that can be used on a variety of hardware platforms. Alternatively, the disclosed system may be implemented partially or fully in hardware using standard logic circuits. Whether software or hardware is used to implement the system varies depending on the speed and efficiency requirements of the system and also the particular function and the particular software or hardware systems and the particular microprocessor or microcomputer systems being utilized.

[0055] The invention has been described with reference to a particular embodiment. 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 ranking search results obtained from an information retrieval system, comprising

   a search pre-processor, responsive to a search query, for determining a context of the search query in accordance with a predetermined user context profile;
   a search engine, responsive to the search query, for generating a search result comprising at least one item obtained from the information retrieval system; and
   a search post-processor, responsive to the search result, for ranking the item in accordance with the context of the search query.

2. The system of claim 1, wherein the user context profile comprises one or more of a user profile comprising a set of terms rated by the user, a community profile comprising a set of terms rated by members of the community, and a relevance profile comprising a set of terms generated by a recommender system.

3. The system of claim 1 or claim 2, wherein the search engine comprises a plurality of individual search engines, each search engine, responsive to the search query, generating a search result comprising at least one item obtained from the information retrieval system; and
   wherein the post-processor, responsive to the search results, ranks the items in accordance with the context of the search query.

4. The system of any of the preceding claims, wherein the search pre-processor determines a profile of the query $P^q = (t_i{}^q, w_i{}^q)$ in accordance with the predetermined user profile wherein t_i^q comprise the query terms having term weight, w_i^q.

5. The system of any of the preceding claims, further comprising a context profiler for generating a context profile and a document collection comprising a set of documents, the context profiler comprising a user profiler for constructing a term-weight vector for the user, the term-weight vector being extracted from each document in the document collection.

6. The system of any of claims 1 to 4, further comprising a context profiler for generating a context profile and a document collection comprising a set of documents, the context profiler comprising a community profiler for constructing a term-weight vector for the community, the term-weight vector being extracted from each document in the document collection and the community comprising a plurality of users u.

7. The system of claim 6, wherein the community profile $P^c = (t_i, w^c{}_i)$ comprises the set of terms $\{t_i\}$ with their weights w^c_i for each of the individual users u in the community.

8. The system of claim 7, wherein each member u in the community is assigned a weight α_u and for the whole community:

$$\sum_u \alpha_u = 1$$

   and wherein the weight of term t_i in the community profile is evaluated as:

$$w^c{}_i = \sum_u \alpha_u w_i{}^u.$$

9. The system of claim 4, wherein the search post-processor evaluates each item d in the search result and generates a document profile $P^d = (t^d{}_i, w^d{}_i)$ for each item d, where t^d_i is the profile term and w^d_i the weight of each term.

10. The system of claim 9, wherein the predetermined user context profile comprises a community profile, $P^c = (t_i, w^c{}_i)$ where

$$w^c{}_i = \sum_u \alpha_u w_i{}^u,$$

and $w^u{}_i$ is the weight of t_i in a profile of user u in the community.

11. The system of claim 10, wherein the search post-processor determines the relevance of each item d in the search result in accordance with:

$$relevance_{q,d} = \frac{\sum_i w_i^d \times w_i^{prof} \times w_i^q}{w_d}$$

where $w^q{}_i$ is the weight of the term t in a query q, and w_d is the vector length "projected" on the context profile and evaluated as

$$w_d = \sqrt{\sum_t (w_i{}^{prof} w_i^d)^2}.$$

12. A method of ranking search results obtained from an information retrieval system, comprising:

providing a predetermined user context profile;
generating a search query;
applying the context profile to the search query to generate a user context of the search query;
generating a search result in response to the search query, comprising at least one item obtained from the information retrieval system;
ranking the item in accordance with the context of the search query.

13. The method of claim 12, wherein the user context profile comprises one or more of a user profile comprising a set of terms rated by the user from a user provided document collection, a community profile comprising a set of terms rated by members of the community, and a relevance profile comprising a set of terms generated by a recommender system.

14. The method of claim 12 or claim 13, further comprising:

generating a plurality of search results, each search result being obtained from the information retrieval system;
ranking the search results in accordance with the context of the search query, and updating the predetermined user context profile using highly ranked items returned from the search query.

15. A method of creating a user context profile for use in ranking search results obtained from an information retrieval system, comprising:

providing a document collection comprising a plurality of relevant documents;
assigning a rating to each of the documents in the document collection to generate a document profile $P^d = (t^d{}_i, w^d{}_i)$ for each document in the collection;
constructing a weight-term vector, wherein the weight-term vector includes a portion of the set of terms $\{t^d{}_i\}$ with their weights $w^d{}_i$ to form a user profile $P^u = (t_i, w_i)$. 16. The method of claim 15, further comprising:
for each document d returned in response to a search query q, generating a document profile, $P^q = \{t^q{}_i, w_{q,t}\}$;
evaluating the search document rank in accordance with:

$$relevance_{q,d} = \frac{\sum_i w_i^d \times w_i^{prof} \times w_i^q}{w_d}$$

where $w^d{}_i$ is the weight of term t in the response document d, w_i^{prof} is the weight of term t in the user context profile, $^{wq}{}_i$ is the weight of the term t in the query q, and w_d is the vector length "projected" on the context profile and evaluated as

$$w_d = \sqrt{\sum_t (w_i{}^{prof} w_i^d)^2}; \text{ and}$$

updating the user context profile using the highest ranked item returned from the search query.

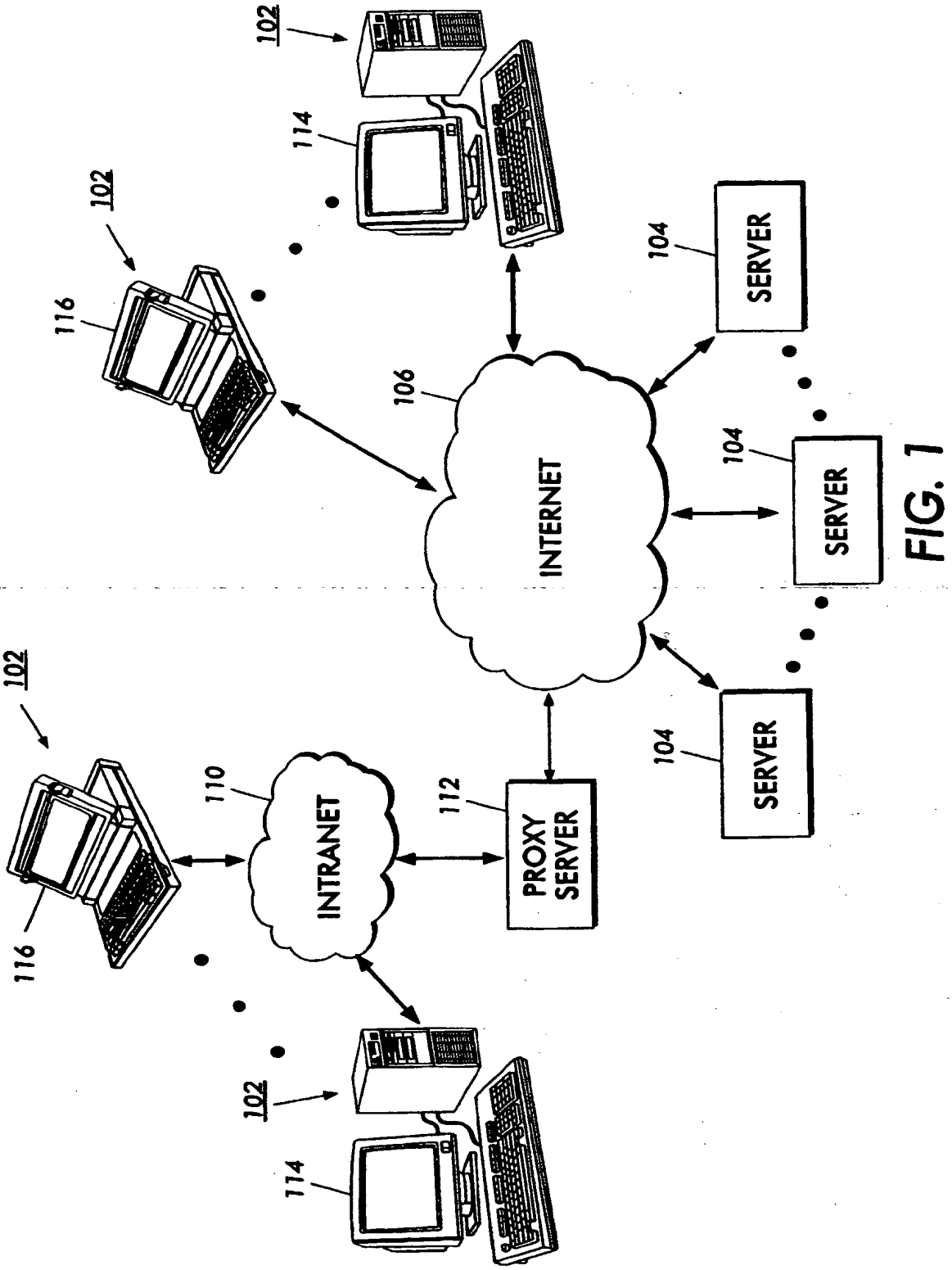


FIG. 1

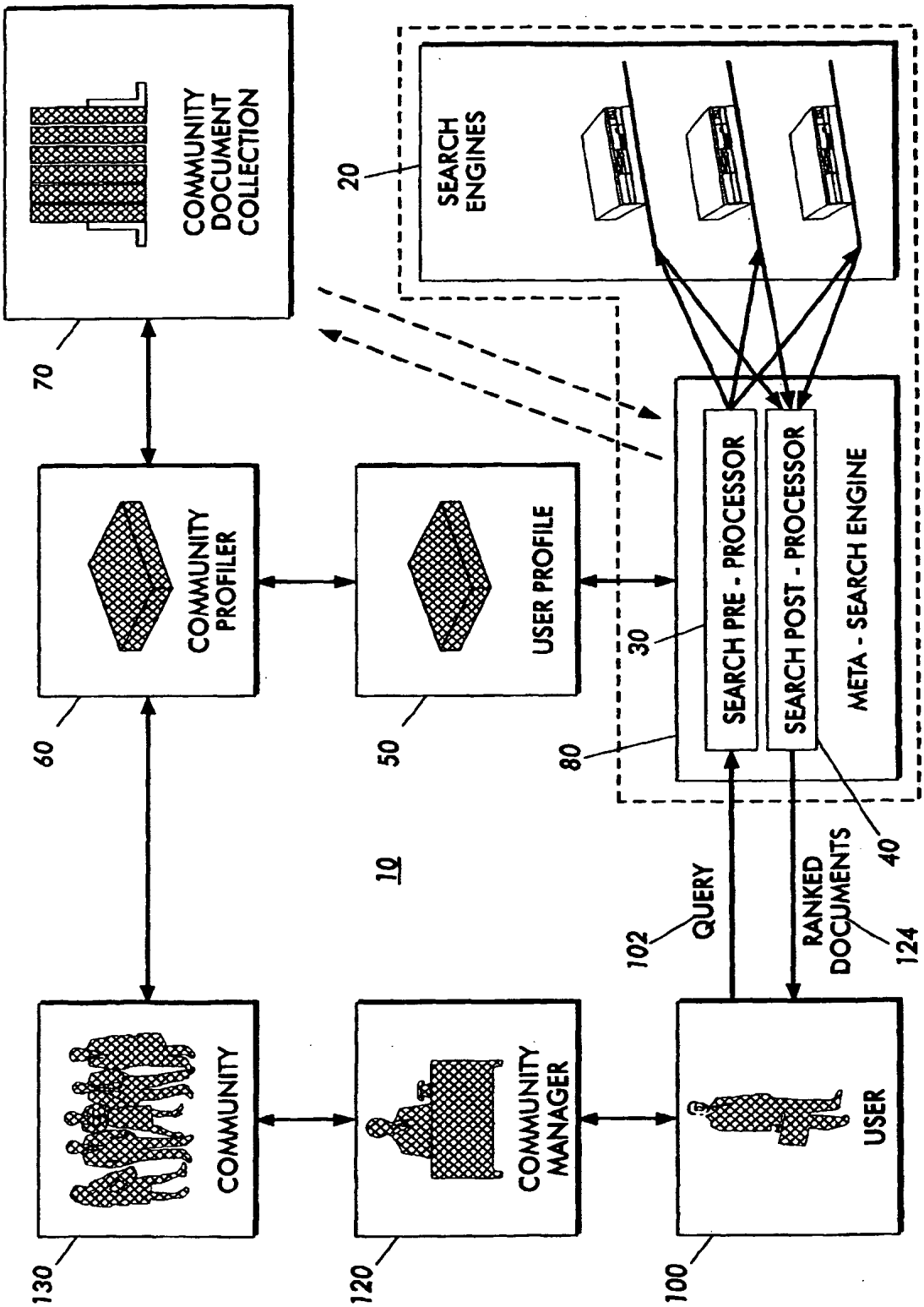


FIG. 2

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

(11) EP 1 050 830 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
17.04.2002 Bulletin 2002/16

(43) Date of publication A2:
08.11.2000 Bulletin 2000/45

(21) Application number: **00303613.4**

(22) Date of filing: **28.04.2000**

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

(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: **05.05.1999 US 305435**

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

(72) Inventors:
- **Chidlovski, Boris**
  38240 Meylan (FR)
- **Glance, Natalie S.**
  38240 Meylan (FR)
- **Grasso, Antonietta**
  38000 Grenoble (FR)

(74) Representative: **Skone James, Robert Edmund**
GILL JENNINGS & EVERY
Broadgate House
7 Eldon Street
London EC2M 7LH (GB)

(54) System and method for collaborative ranking of search results employing user and group profiles

(57) A system for ranking search results obtained from an information retrieval system includes a search pre-processor (30), a search engine (20) and a search post-processor (40). The search pre-processor (30) determines the context of the search query by comparing the terms in the search query with a predetermined user context profile. Preferably, the context profile is a user profile or a community profile, which includes a set of terms which have been rated by the user, community, or a recommender system. The search engine generates a search result comprising at least one item obtained from the information retrieval system. The search post-processor (40) ranks each item returned in the search result in accordance with the context of the search query.

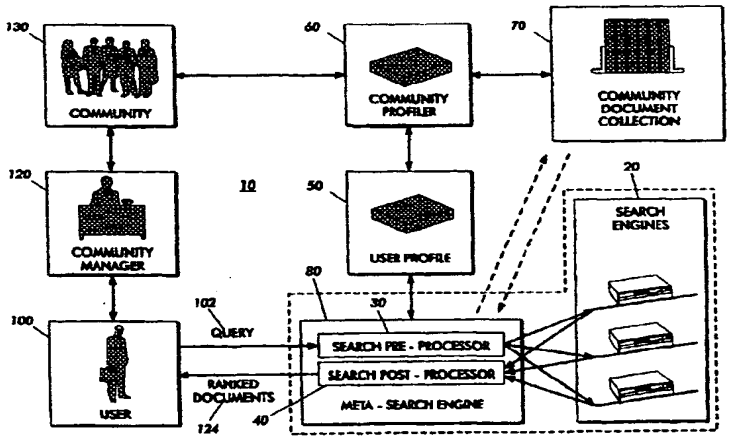


FIG. 2

EP 1 050 830 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 00 30 3613

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)
X	WO 98 33135 A (FIREFLY NETWORK INC) 30 July 1998 (1998-07-30) * column 2, line 48 - column 3, line 12 * * column 6, line 24 - column 9, line 48 *	1,12,15	G06F17/30
A	US 5 890 152 A (RAPAPORT JEFFREY ALAN ET AL) 30 March 1999 (1999-03-30) * the whole document *		
D,A	US 4 996 642 A (HEY JOHN B) 26 February 1991 (1991-02-26) * the whole document *		
			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	25 February 2002	DE CASTRO PALOM.., L

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 00 30 3613

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.

25-02-2002

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9833135 A	30-07-1998	US 6041311 A	21-03-2000
		AU 6041898 A	18-08-1998
		WO 9833135 A1	30-07-1998
US 5890152 A	30-03-1999	WO 9810597 A2	12-03-1998
US 4996642 A	26-02-1991	CA 1324675 A1	23-11-1993
		US 4870579 A	26-09-1989

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

This Page is Inserted by IFW Indexing and Scanning Operations and is not 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 include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☑ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2003 (24.12.2003)

PCT

(10) International Publication Number
WO 03/107127 A2

(51) International Patent Classification[7]: G06F

(21) International Application Number: PCT/US03/18685

(22) International Filing Date: 12 June 2003 (12.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/172,165 14 June 2002 (14.06.2002) US

(71) Applicant: **ENTOPIA, INC.** [US/US]; 1301 Shoreway Road, Suite 302, Belmont, CA 94002 (US).

(72) Inventors: **PERISIC, Igor**; 739 West Capistrano Way, San Mateo, CA 94402-2012 (US). **POSSE, Christian**; 2833 21st Avenue W., Seattle, WA 98199-2910 (US).

(74) Agents: **MORRIS, Francis, E.** et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

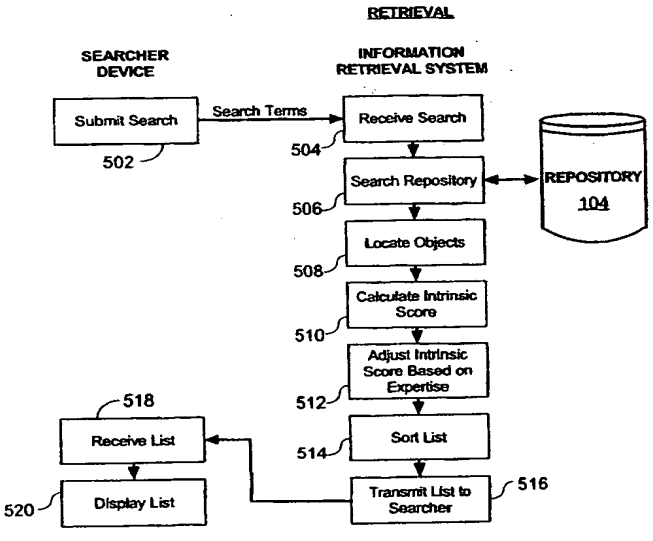
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM AND METHOD FOR PERSONALIZED INFORMATION RETRIEVAL BASED ON USER EXPERTISE



(57) Abstract: A search request is received at an information retrieval system from a searcher. The search request preferably contains at least one search term and a user identifier. A plurality of objects are then searched based on the at least one search term. At least one located object is found from the plurality of objects. The at least one located object is associated with the search term(s). An intrinsic score based on the search term(s) is subsequently calculated for each located object. The intrinsic score is then adjusted to an adjusted score based on the difference between a creator expertise of a creator of the at least one located object and/or a contributor expertise of a contributor to the at least one located object, and a searcher expertise of the searcher.

WO 03/107127 A2

SYSTEM AND METHOD FOR PERSONALIZED INFORMATION RETRIEVAL BASED ON USER EXPERTISE

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0001] The present invention relates generally to information retrieval, and more particularly to a system and method for adjusting search results based on the relative expertise between a searcher and the creator/s and/or contributor/s of a document.

DESCRIPTION OF RELATED ART

[0002] With the proliferation of corporate networks and the Internet, an ever increasing amount of information is being made available in electronic form. Such information includes documents, graphics, video, audio, or the like. While corporate information is typically well indexed and stored on corporate databases within a corporate network, information on the Internet is generally highly disorganized.

[0003] Searchers looking for information typically make use of an information retrieval system. In corporate networks, such an information retrieval system typically consists of document management software, such as Applicant's QUANTUM™ suite, or iManage Inc's INFORITE™ or WORKSITE™ products. Information retrieval from the internet, however, is typically undertaken using a search engine, such as YAHOO™ or GOOGLE™.

[0004] Generally speaking, these information retrieval systems extract keywords from each document in a network. Such keywords typically contain no semantic or syntactic information. For each document, each keyword is then indexed into a searchable data structure with a link back to the document itself. To search the network, a user supplies the information retrieval system with a query containing one or more search terms, which may be separated by Boolean operators, such as "AND" or "OR." These search terms can be further expanded

through the use of a Thesaurus. In response to the query, which might have been expanded, the information retrieval system attempts to locate information, such as documents, that match the searcher supplied (or expanded) keywords. In doing so, the information retrieval system searches through its databases to locate documents that contain at least one keyword that matches one of the search terms in the query (or its expanded version). The information retrieval system then presents the searcher with a list of document records for the documents located. The list is typically sorted based on document ranking, where each document is ranked according to the number of keyword to search term matches in that document relative to those for the other located documents. An example of a search engine that uses such a technique, where document relevancy is based solely on the content of the document, is INTELISEEK™. However, most documents retrieved in response to such a query have been found to be irrelevant.

[0005] In an attempt to improve precision, a number of advanced information retrieval techniques have been developed. These techniques include syntactic processing, natural language processing, semantic processing, or the like. Details of such techniques can be found in U.S. Patent Nos. 5,933,822; 6,182,068;6,311,194; and 6,199,067, all of which are incorporated herein by reference.

[0006] However, even these advanced information retrieval techniques have not been able to reach the level of precision required by today's corporations. In fact, a recent survey found that forty four percent of users say that they are frustrated with search engine results. See *Internet Usage High, Satisfaction low: Web Navigation Frustrate Many Consumers*, Berrier Associates - sponsored by Realnames Corporation (April 2000).

[0007] In addition, other advanced techniques have also proven to lack adequate precision. For example, GOOGLE™ and WISENUT™ rank document relevancy as a function of a network of links pointing to the document, while methods based on Salton's work (such as ORACLE™ text) rank document relevancy as a function of the number of relevant documents within the repository.

[0008] This lack of precision is at least partially caused by current information retrieval systems not taking the personal profiles of the document creator, searcher,

and any contributors into account. In other words, when trying to assess the relevancy of documents within a network, most information retrieval systems ignore the searcher that performs the query, *i.e.*, most information retrieval systems adopt a one-fit-all approach. For example, when a neurologist and a high school student both perform a search for "brain AND scan," an identical list of located documents is presented to both the neurologist and the high school student. However, the neurologist is interested in high level documents containing detailed descriptions of brain scanning techniques, while the student is only interested in basic information on brain scans for a school project. As can be seen, a document query that does not take the searcher into account can retrieve irrelevant and imprecise results.

[0009] Moreover, not only should the profession of a searcher affect a search result, but also the expertise of the searcher within the search domain. For example, a medical doctor that is a recognized world expert would certainly assign different relevancy scores to the returned documents than say an intern. This means that information retrieval systems should be highly dynamic and consider the current expertise level of the searcher and/or creator/s at the time of the query.

[0010] In addition, the current lack of precision is at least partially caused by the treatment of documents as static entities. Current information retrieval techniques typically do not take into account the dynamic nature of documents. For example, after creation, documents may be commented on, printed, viewed, copied, *etc.* To this end, document relevancy should consider the activity around a document.

[0011] Therefore, a need exists in the art for a system and method for retrieving information that can yield a significant improvement in precision over that attainable through conventional information retrieval systems. Moreover, such a system and method should preferably personalize information retrieval based on user expertise.

BRIEF SUMMARY OF THE INVENTION

[0012] According to the invention there is provided a method for personalizing information retrieval. A search request is received at an information retrieval system from a searcher. The search request preferably contains at least one search term

and a user identifier. A plurality of objects are then searched based on the search term(s). Objects preferably include: content objects, such as documents, comments, or folders; source objects; people objects, such as experts, peers, or workgroups; or the like. At least one located object is found from the plurality of objects. Each located object is associated with the search term(s). An intrinsic score based on the search term(s) is subsequently calculated for each located object. The intrinsic score is based on the search term(s). The intrinsic score is then adjusted to an adjusted score based on the difference between a creator expertise of the creator of the located object and/or a contributor expertise of a contributor/s to the located object, and a searcher expertise of the searcher.

[0013] To make this adjustment, objects associated with the searcher are located using the unique user identifier. The searcher expertise is then ascertained based on the search terms and the objects associated with the searcher. The creator expertise and/or the contributor/s expertise is determined for each located object. The intrinsic score, with the exception of people related searches, is then raised to the adjusted score for each located object having a creator expertise higher than the searcher expertise. Alternatively, the intrinsic score, with the exception of people related searches, is then lowered to the adjusted score for each located object having a creator expertise that is lower than the searcher expertise. For people related searches, creator expertise significantly above or below the searcher expertise negatively affect the intrinsic score of the located object. A list is then transmitted to the searcher to be displayed, where the list is based on the search request and the adjusted scores.

[0014] Accordingly, at the time of the query, the expertise of the searcher is taken into consideration in relation to the expertise of both the creator/s and contributor/s. Therefore, the present invention yields a significant improvement in precision over that attainable through conventional information retrieval systems.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] Additional features of the invention will be more readily apparent from the following detailed description and appended claims when taken in conjunction with the drawings, in which:

[0016] Figure 1 is a block diagram of a system architecture for a system for personalizing information retrieval;

[0017] Figure 2 is a block diagram of a creator device, contributor device, or searcher device, as shown in Figure 1;

[0018] Figure 3 is a block diagram of the information retrieval system and Repository of Figure 1;

[0019] Figure 4 is a flow chart of document collection according to an embodiment of the invention; and

[0020] Figure 5 is a flow chart of a process for information retrieval according to an embodiment of the invention.

[0021] Like reference numerals refer to corresponding parts throughout the several views of the drawings. For ease of reference, the first number/s of any reference numeral generally indicates the number of the figure where the reference numeral can be found. For example, 112 can be found on Figure 1, and 324 can be found on Figure 3.

DETAILED DESCRIPTION OF THE INVENTION

[0022] Figure 1 is a block diagram of a system architecture 100 for a system for personalizing information retrieval. An information retrieval system 102 is coupled to a repository 104 and to a network 110. Also coupled to the network 110 are a searcher device 108, one or more creator device/s 106, and one or more contributor device/s 112. Searcher device 108, creator device/s 106, contributor device/s 112, and information retrieval system 102 are all computing devices, such as clients, servers, or the like. The network is preferably a Local Area Network (LAN), but alternatively may be any network, such as the Internet. It should be appreciated that although searcher device 108, creator device/s 106, contributor device/s 112, and information retrieval system 102 are shown as distinct entities, they may be combined into one or more devices. Further details of the searcher

device 108, creator device/s 106, contributor device/s 112, and information retrieval system 102 can be found below in relation to Figures 2-5.

[0023] The repository 104 is any storage device/s that is capable of storing data, such as a hard disk drive, magnetic media drive, or the like. The repository 104 is preferably contained within the information retrieval system 102, but is shown as a separate compnent for ease of explaination. Alternatively, the repository 104 may be dispersed throughout a network, and may even be located within the searcher device 108, creator device/s 106, and/or contributor device/s 112.

[0024] Each creator device 106 is a computing device operated by a creator who creates one or more documents. Each contributor device 112 is a computing device operated by a contributor who contributes to a document by, for example, adding to, commenting on, viewing, or otherwise accessing documents created by a creator/s. The searcher device 108 is a computing device operated by a searcher who is conducting a search for relevant documents created by the creator/s or contributed to by the contributor/s. The searcher, creator/s, and contributor/s are not limited to the above described roles and may take on any role at different times. Also, the searcher, creator/s, and contributor/s may browse the repository 104 without the use of the information retrieval system 102.

[0025] Figure 2 is a block diagram of a creator device 106, contributor device 112, or searcher device 108, as shown in Figure 1. The devices 106/108/112 preferably include the following cmponents: at least one data processor or central processing unit (CPU) 202; a memory 214; input and/or output devices 206, such as a monitor and keyboard; communications circuitry 204 for communicating with the network 110 (Figure 1) and information retrieval system 102 (Figure 1); and at least one bus 210 that interconnects these components.

[0026] Memory 214 preferably includes an operating system 216, such as but not limited to, VXWORKS™, LINUX™, or WINDOWS™ having instructions for processing, accessing, storing, or searching data, *etc.* Memory 214 also preferably includes communication procedures for communicating with the network 110 (Figure 1) and information retrieval system 102 (Figure 1); searching procedures 220, such as proprietary search software, a Web-browser, or the like; application programs 222, such as a word processor, email client, database, or the like; a unique user

identifier 224; and a cache 226 for temporarily storing data. The unique user identifier 224 may be supplied by the creator/searcher/contributor each time he or she performs a search, such as by supplying a username. Alternatively, the unique user identifier 224 may be the user's login username, Media Access Control (MAC) address, Internet Protocol (IP) address, or the like.

[0027] Figure 3 is a block diagram of the information retrieval system 102 and Repository 104 of Figure 1. As mentioned in relation to Figure 1, the repository 104 is preferably contained within the information retrieval system 102. The information retrieval system 102 preferably includes the following components: at least one data processor or central processing unit (CPU) 302; a memory 308; input and/or output devices 306, such as a monitor and keyboard; communications circuitry 304 for communicating with the network 110 (Figure 1), creator device/s 106 (Figure 1), contributor device/s 112 (Figure 1), and/or searcher device 108 (Figure 1); and at least one bus 310 that interconnects these components.

[0028] Memory 308 preferably includes an operating system 312, such as but not limited to, VXWORKS™, LINUX™, or WINDOWS™ having instructions for processing, accessing, storing, or searching data, etc. Memory 308 also preferably includes communication procedures 314 for communicating with the network 110 (Figure 1) , creator device/s 106 (Figure 1), contributor device/s 112 (Figure 1), and/or searcher device 108 (Figure 1); a collection engine 316 for receiving and storing documents; a search engine 324; expertise adjustment procedures 326; a repository 104, as shown in Figure 1; and a cache 338 for temporarily storing data.

[0029] The collection engine 316 comprises a keyword extractor or parser 318 that extracts text and/or keywords from any suitable document, such as an ASCII or XML file, Portable Document Format (PDF) file, word processing file, or the like. The collection engine 316 also preferably comprises a concept identifier 320. The concept identifier 320 is used to extract the document's important concepts. The concept identifier may be a semantic, synaptic, or linguistic engine, or the like. In a preferred embodiment the concept identifier 320 is a semantic engine, such as TEXTANALYST™ made by MEGAPUTER INTELLIGENCE™ Inc. Furthermore, the collection engine 316 also preferably comprises a metadata filter 322 for filtering and/or refining the concept/s identified by the concept identifier 320. Once the

metadata filter 322 has filtered and/or refined the concept, metadata about each document is stored in the repository 104. Further details of the processes performed by the collection engine 316 are discussed in relation to Figure 4. In addition to refined concepts, metadata includes any data, other than raw content, associated with a document.

[0030] The search engine 324 is any standard search engine, such as a keyword search engine, statistical search engine, semantic search engine, linguistic search engine, natural language search engine, or the like. In a preferred embodiment, the search engine 324 is a semantic search engine.

[0031] The expertise adjustment procedures 326 are used to adjust an object's intrinsic score to an adjusted score based on the expertise of the searcher, creator/s, and/or contributor/s. The expertise adjustment procedures 326 are described in further detail below in relation to figures 5.

[0032] A file collection 328(1)-(N) is created in the repository 104 for each object input into the system, such as a document or source. Each file collection 328(1)-(N) preferably contains: metadata 330(1)-(N), such as associations between keywords, concepts, or the like; content 332(1)-(N), which is preferably ASCII or XML text or the content's original format; and contributions 334(1)-(N), such as contributor comments or the like. At a minimum, each file collection contains content 332(1)-(N). The repository 104 also contains user profiles 336(1)-(N) for each user, *i.e.*, each searcher, creator, or contributor. Each user profile 336(1)-(N) includes associated user activity, such as which files a user has created, commented on, opened, printed, viewed, or the like, and links to various file collections 328(1)-(N)that the user has created or contributed to. Further details of use of the repository 104 are discussed in relation to Figure 5.

[0033] Figure 4 is a flow chart of document collection according to an embodiment of the invention. A creator supplies an object, such as a document or source, to the searching procedures 220 (Figure 2) at step 402. To supply a document, the creator may for example, supply any type of data file that contains text, such as an email, word processing document, text document, or the like. A document comes from a source of the document. Therefore, to supply a source, the creator may provide a link to a document, such as by providing a URL to a Web-

page on the Internet, or supply a directory that contains multiple documents. In a preferred embodiment, the creator also supplies his or her unique user identifier 224 (Figure 2), and any other data, such as privacy settings, or the like. The unique user identifier may be supplied without the creator's knowledge, such as by the creator device 106 (Figure 1) automatically supplying its IP or MAC address.

[0034] The document, source, and/or other data is then sent to the information retrieval system 102 (Figure 1) by the communication procedures 218 (Figure 2). The information retrieval system 102 (Figure 1) receives the document, source, and/or other data at step 403. When supplied with a document, the keyword extractor or parser 318 (Figure 3) parses the document and/or source into ASCII text at step 404, and thereafter extracts the important keywords at step 408. However, when supplied with a source, the keyword extractor or parser 318 (Figure 3) firstly obtains the document/s from the source before parsing the important keywords into text.

[0035] Extraction of important keywords is undertaken using any suitable technique. These keywords, document, source, and other data are then stored at step 406 as in the repository 104 as part of a file collection 328(1)-(N) (Figure 3). Also, the unique user identifier is used to associate or link each file collection 328(1)-(N) (Figure 3) created with a particular creator. This link between the creator and the file collection is stored in the creator's user profile 336(1)-(N) (Figure 3). The user profile data can be updated by the user him/herself or more preferably by a system administrator.

[0036] In a preferred embodiment, the concept identifier 320 (Figure 3) then identifies the important concept/s from the extracted keywords at step 410. Again, in a preferred embodiment, the metadata filter 322 (Figure 3) then refines the concept at step 412. The refined concept is then stored in the repository 104 as part of the metadata 330(1)-(N) (Figure 3) within a file collection 328(1)-(N) (Figure 3).

[0037] At any time, contributors can supply their contributions, at step 416, such as additional comments, threads, or other activity to be associated with the file collection 328(1)-(N). These contributions are received by the information retrieval engine at step 418 and stored in the repository at step 420, as contributions 334(1)-

(N). Alternatively, contributions may be received and treated in the same manner as a document/source, *i.e.*, steps 403-414.

[0038] Figure 5 is a flow chart of a process for information retrieval according to an embodiment of the invention. A searcher using a searcher device 108 (Figure 1) submits a search request to the information retrieval system 102 (Figure 1), at step 502. Submittal of this search occurs using searching procedures 220 (Figure 2) and communication procedures 218 (Figure 2) on the searcher device 108 (Figure 1). The search request preferably contains one or more search terms, and the unique user identifier 224 (Figure 2) of the searcher.

[0039] The search is preferably conducted to locate objects. Objects preferably include: content objects, such as documents, comments, or folders; source objects; people objects, such as experts, peers, or workgroups; or the like. A search for documents returns a list of relevant documents, and a search for experts returns a list of experts with expertise in the relevant field. A search for sources returns a list of sources from where relevant documents were obtained. For example, multiple relevant documents may be stored within a particular directory or website.

[0040] The search is received at step 504 by the information retrieval system 102 (Figure 1) using communications procedures 314 (Figure 3). The information retrieval system 102 (Figure 1) then searches the repository 104 for relevant objects at step 506. This search is undertaken by the search engine 324 (Figure 3), at step 506, using any known or yet to be discovered search techniques. In a preferred embodiment, the search undertakes a semantic analysis of each file collection 328(1)-(N) stored in the repository 104.

[0041] The search engine 324 (Figure 3) then locates relevant objects 328(1)-(N) at step 508 and calculates an intrinsic score at step 510 for each located object. By "located object," it is meant any part of a file collection that is found to be relevant, including the content, source, metadata, *etc.* Calculation of the intrinsic score is based on known, or yet to be discovered techniques for calculating relevancy of located objects based solely on the located objects themseves, the repository itself and the search terms. In its simplest form, such a search calculates the intrinsic score based on the number of times that a search term appears in the

content 332(1)-(N) (Figure 3) of located objects. However, in a preferred embodiment, this calculation is also based on a semantic analysis of the relationship between words in the content 332(1)-(N) (Figure 3).

[0042] The intrinsic score is then adjusted to an adjusted score by the expertise adjustment procedures 326, at step 512. This adjustment takes the expertise of the creator/s, searcher, and/or contributor/s into account, as described in further detail below.

[0043] Once the intrinsic score has been adjusted to an adjusted score, a list of the located objects is sorted at step 514. The list may be sorted by any field, such as by adjusted score, intrinsic score, source, most recently viewed, creator expertise, *etc.* The list, preferably containing a brief record for each located object, is then transmitted to the searcher device 108 (Figure 1) at step 516. Each record preferably contains the located object's adjusted score, creator, title, *etc.* The list is then received by the searcher device at step 518 and displayed to the searcher at step 520. In an alternative embodiment, sorting of the list is performed by the searching procedures 220 (Figure 2) on the searcher device 108 (Figure 1).

[0044] Preferred algorithms for adjusting the intrinsic score (step 512 of Figure 5) will now be described. It should be appreciated that these algorithms are merely exemplary and in no way limit the invention other than as claimed. Calculation of the adjusted score from the intrinsic score is dependent on the objects searched for, such as documents, comments, sources, experts, or peers.

EXPERTISE ADJUSTMENT WHEN SEARCHING FOR DOCUMENTS

[0045] Search term(s) entered by the searcher may or may not be extended to form a query. Such possible extensions, include but are not limited to, synonyms or stemming of search term(s). Once the intrinsic score has been calculated according to step 510 above, the adjusted score (RS_ADJ) for each located document is calculated as follows:

(1) RS_ADJ = Intrinsic Document Score + Expertise Adjustment
= IDS + EA

where the Intrinsic Document Score (IDS) is a weighted average between a Document Content Score (DCS) and a Comments Content Score (CCS).

(2) IDS = a*DCS + (1-a)*CCS

with "a" being a number between 0 and 1 and determining the importance of the content of a document relative to the content of its attached comments.

[0046] The DCS and CCS are calculated by any methodology or technique. Existing search engine algorithms may be used to fulfill this task. Also note that the DCS and CCS are not influenced by the searcher that entered the query. In this embodiment, the DCS and CCS can be any number between 2 and 100. The Expertise Adjustment (EA) is calculated as follows:

(3) EA = DCE + CCE

where DCE is the Document Creator Expertise adjustment and CCE is the Comments Contributors Expertise adjustment. The DCE adjustment takes into account all activity performed by a given user and is computed as follows:

(4) DCE = R1(DCS) * W1(RS_EXP_ABS)

where R1(DCS) determines the maximal amount of the expertise adjustment, or, in other words, the range for the alteration due to the expertise of the creator of the document. This depends on the level of the DCS. The range function is given by:

(5) $$R1(DCS) = 20 * \left(1 - \frac{|DCS - 50|}{100}\right)$$

[0047] Extreme intrinsic scores, *i.e.*, scores near 2 or 100, are less influenced than scores near the middle, *i.e.*, scores near 50. The maximum possible change in a score is 20 when DCS = 50 and linearly decreases to 10 when DCS = 100 or 2.

[0048] W1(RS_EXP_ABS) determines what percentage of available range R1(DCS), positively or negatively, is considered for adjusting the intrinsic score. It is given by:

(6) $$W1(RS\_EXP\_ABS) = \frac{RS\_EXP\_ABS(Creator) - RS\_EXP\_ABS(Searcher)}{100}$$

where RS-EXP-ABS denotes the absolute relevance score of a user, that is, the user expertise, be it searcher expertise, creator expertise, or contributor expertise. The calculation of RS-EXP-ABS occurs as follows:

(7) RS-EXP-ABS = 3*F(User contribution)*G(Company expertise)*H(Query specificity)

where F (User contribution) accounts for the relevancy of all contributions made by the user, considering all documents created, all comments contributed, and the user's definition of his or her folders within the software. These folders (private or public) constitute the user's personal taxonomy. G (Company expertise) accounts for the company expertise about the query, *i.e.*, whether a few or most employees in a company have produced something relevant to the query. H (Query specificity) accounts for the specificity of the query within the repository,, *i.e.*, whether many or just a few file collections were created.

[0049] In detail:

(8) $$F(\text{User cont.}) = \sum_{\substack{i:\text{allrelevant} \\ \text{documents}}} \left(2*\left(W_{i,max} + C_i\right)*\left(\frac{(DCS)_i}{100}\right)^2 \right) + \sum_{\substack{i:\text{allnonrelevant} \\ \text{documents}}} C_i + 2*\text{Taxonomy}$$

where the first sum is over all relevant documents and the second sum is over all non-relevant documents that possessed a relevant comment, *i.e.*, the comment was relevant but not the document. $(DCS)_i$ is the Intrinsic document relevancy score attained for the i-th relevant document. Also, $W_{i,max}$, is the user activity measure. C_i is calculated as follows:

(9) $$C_i = 0.1 * \left(1 - \mathrm{Exp}\left(-\frac{\#\text{relevant comments in } \mathrm{Doc}_i \text{ by this user}}{2}\right)\right)$$

and is the reward assigned to matching comments made on documents, relevant or not. A matched comment is not necessarily attached to a relevant document.

[0050] $W_{i,max}$ accounts for the type of contribution (such as but not limited to creation, commenting, or highlighting). In short, $W_{i,max}$ is the maximum of the following weights (if applicable).

$W_{i,edit}$	= 1	if the user created or edited i-th file collection,
$W_{i,comment}$	= 0.5 * Max*(0,7 - $\mathrm{Min}_{comments}$*(Level))/6	if the user commented on the i-th file collection. Since these comments are organized in a threaded discussion, the weight will also depend on how remote a comment is to the file collection itself. For example, a comment on a comment on a comment to the original file collection will receive a lesser weight than a comment on the original file collection. In the formula, Level measures how remote the comment is from the file collection. The least remote comment is taken into consideration as long as it is closer than six comments away from the parent file collection.
$W_{i,rename}$	= 0.8	if the user renamed i-th file collection.
$W_{i,highlight}$	= 0.8	if the user highlighted some subparts of i-th file collection.
$W_{i,link}$	= 0.8	if the user linked the file collection to another file collection or "external" URL.

$$Taxonomy = \begin{cases} 1 & \textit{If Query term found within user's taxonomy} \\ 0 & \textit{Otherwise} \end{cases}$$

[0051] The taxonomy in this preferred embodiment stands for folder names. Each user has built some part of the repository by naming folders, directories, or sub-directories. For example, creator 1 might have grouped his Hubble telescope pictures in a folder called "Space Images." Then term "Space Images" becomes part of the user's taxonomy.

[0052] Within an organization or enterprise, some of the taxonomy (folder structure) has been defined by the organization or enterprise itself and has "no owners." In this case, each folder has an administrator who bestows rights to users, such as the right to access the folder, the right to edit any documents within it, the right to edit only documents that the specific user created, or the right to view but not edit or contribute any document to the folder. Only the names of the folders that a user creates are part of his or her taxonomy.

$$(10)\quad G(\text{Company expertise}) = 1 + \text{Log}\left(\frac{P}{E}\right) = \text{IEF},$$

where Log is the logarithmic function base 10; P is the total number of users; and E is the number of relevant experts. The number of relevant experts is calculated by determining how many unique creators and contributors either created or contributed to the located documents. IEF stands for Inverse Expertise Frequency.

[0053] This adjustment raises the adjusted scores when there are few relevant experts within the company.

$$(11)\quad H(\text{Query specificity}) = 1 + \frac{1}{\text{Log}(NCO)}\text{Log}\left(\frac{NCO}{NCOR}\right) = 2 - \frac{\text{Log}(NCOR)}{\text{Log}(NCO)} = \text{IWCOF}$$

where Log is the logarithmic function base 10; NCO is the total number of content objects available in the database at the time of the query; and NCOR is the total number of relevant content objects for a given query. IWCOF stands for the Inverse Weighted Content Objects Frequency. Preferably, in this embodiment, NCO, NCOR and IWCOF are only calculated using non-confidential content objects.

[0054] IWCOF is similar to IEF as it adjusts the score by slightly raising the adjusted score when only a few relevant content objects are found in the database. Therefore, the absolute relevance score for a given user (or the user expertise) is:

$$(12)\quad \text{RS-EXP-ABS} = 3 * \left(1+ \text{Log}\left(\frac{P}{E}\right)\right)\left(2-\frac{\text{Log(NCOR)}}{\text{Log(NCO)}}\right)$$

$$* \left(\sum_{\substack{\text{i:allrelevant} \\ \text{documents}}} \left(2 * \left(W_{i,max} + C_i \right) * \left(\frac{(DCS)_i}{100} \right)^2 \right) + \sum_{\substack{\text{i:allnonrelevant} \\ \text{documents}}} C_i + 2 * \text{Taxonomy} \right)$$

$$= 3 * \text{IEF} * \text{IWCOF} *$$

$$\left(\sum_{\substack{\text{i:allrelevant} \\ \text{documents}}} \left(2 * \left(W_{i,max} + C_i \right) * \left(\frac{(DCS)_i}{100} \right)^2 \right) + \sum_{\substack{\text{i:allnonrelevant} \\ \text{documents}}} C_i + 2 * \text{Taxonomy} \right)$$

[0055] Using the above equations, the intrinsic score is increased to an adjusted score if the creator of the content objects is more knowledgeable about the searched subject matter than the person that entered the query, *i.e.*, if the creator expertise is higher than the searcher expertise. On the other hand, the intrinsic score is decreased to an adjusted score if the creator is less knowledgeable about the searched subject matter than the searcher, *i.e.*, if the creator expertise is lower than the searcher expertise.

[0056] To calculate the Comments Contributors Expertise Adjustment (CCE) the following equation is used:

$$(13)\quad \text{CCE} = 5 * \left(2 * \frac{\text{Exp(Dx)}}{1+\text{Exp(Dx)}} - 1\right)$$

where

(14)

$$\Delta x = \frac{1}{50} \sum_{DistinctContributors} (RS\_EXP\_ABS(Contributors) - RS\_EXP\_ABS(Searcher))$$

[0057] Once these adjustments have been computed, one has to ensure that the relevancy score from (1) is in the appropriate range and that it is preferably in this embodiment an integer. This is obtained as follows:

(15) RS_ADJ = Min(100, Max (1, Round(RS_ADJ)))

where Round(d) rounds the number d to its nearest integer.

EXPERTISE ADJUSTMENT WHEN SEARCHING FOR SOURCES

[0058] Once the intrinsic score has been calculated according to step 510 above, the adjusted score for sources (RSS_ADJ) for each source is calculated as follows:

(16) RSS_ADJ = intrinsic Source Content score + expertise adjustment

= SCS + R2(SCS)*W2(RS_EXP_ABS)

where SCS is the intrinsic Source Content Score computed, which is, preferably in this embodiment, defined here as the maximum of all the intrinsic Document Content Scores (DCS) that were created from each source, *i.e.*,

(17) SCS = MAX(DCS)

[0059] For example, multiple documents may have been saved as multiple file collections from a single Web-site.

[0060] R2(SCS) determines the maximal amount of the expertise adjustment, or, in other words, the range for the alteration due to the expertise of the creator of the document taken from the source, which depends on the level of the intrinsic source score, *i.e.*, SCS. The range function is given by:

$$(18)\quad R1(SCS) = 20 * \left(1 - \frac{|SCS - 50|}{100}\right)$$

[0061] Extreme scores are less influenced than scores in the middle. The maximum possible change in a score is 20 when SCS = 50 and linearly decreases to 10 when SCS = 100 or 2.

[0062] W2(RS_EXP_ABS) determines what percentage of available range for the expertise adjustment, R2(SCS), positively or negatively, is considered for building the scoring. It is given by:

(19)

$$W2(RS\_EXP\_ABS) = \frac{MAX(RS\_EXP\_ABS(Creator)) - RS\_EXP\_ABS(Searcher)}{100}$$

where RS_EXP_ABS is the absolute relevance score of the expert (as defined previously). MAX(RS _EXP _ABS(Creator)) is the maximum of absolute expertise scores over all creators that have created file collections from this source. RS _EXP _ABS(Searcher) is the absolute relevance score of the searcher. In other words, the intrinsic score for the source is adjusted upward to an adjusted score if the maximum creator expertise of all creators for a particular source exceeds the searcher expertise. On the other hand, the intrinsic score for the source is lowered to an adjusted score if the creator expertise of all creators for a particular source is lower than the searcher expertise.

[0063] Once this adjustment has been computed, one has to ensure that the relevancy score is in the appropriate range and that it is preferably in this embodiment an integer. This is obtained as follows:

(20) RSS_ADJ = Min(100, Max (1, Round(RSS_ADJ)))

where Round(d) rounds the number d to its nearest integer.

[0064] In this way, the adjusted score for each document (RS_ADJ) or the adjusted score for sources (RSS_ADJ) is calculated based on the expertise of the searcher, creator/s, and/or contributor/s. Such adjusted scores provide a significant improvement in precision over that attainable through conventional information retrieval systems.

EXPERTISE ADJUSTMENT WHEN SEARCHING FOR PEERS

[0065] When users are looking for peers rather than experts an adjusted relevancy score is calculated. Peers are other users that have a similar expertise or come from a similar, or the same, department as the searcher. The adjusted relevancy score uses the expertise values and adjusts them with respect to the searcher's expertise. This is the similar to resorting the list with respect to the searcher, but instead recalculates the values themselves.

[0066] Once the expertise for each user has been determined, they are adjusted with respect to the searcher expertise. The adjusted relative or personalized relevancy score for an expert is defined by:

(21)

$$\textbf{Adjusted Rel} = 100 - 10 * \left|\left(\sqrt{\text{RS- EXP- ABS}} - \sqrt{\text{RS- EXP- ABS}_{\text{searcher}}} + 10\right)\right|$$

[0067] The adjusted relevancy score is a measure of the difference between two levels of expertise. The square root maps the difference to a continuous and monotone measure while diminishing the importance of differences when two experts are far apart. It is also asymmetric in the sense that it favors expertise above the searcher expertise. Finally, recall that $|K|$ represents the absolute value of K (*i.e.*, the difference).

[0068] An example of a method for personalizing information retrieval using the above formulae will now be described. It should, however, be appreciated that

this example is described herein merely for ease of explanation, and in no way limits the invention to the scenario described. Table 1 sets out the environment in which a search is conducted. Furthermore, in this illustration, the factor a (from formula 2, determining the importance of the content of a document relative to its attached comments) has been arbitrarily set to 1.

TABLE 1		
Number of users	**# experts**	
100	10	
Total Number of File Collections	**# of relevant File Collections**	**# of relevant comments**
1000	10	10
Departments of experts	**Names**	
Marketing	Adam M. Bryan M. Christie M. David M.	
Engineering	Eric E. Fred E. Gail E.	
Finance	Hugo F. Henry F.	
Legal	Ivan L.	
File Collection number	**Creator**	**Contributors (total # of contributions, # of relevant contributions)**
11	Adam M.	Bryan M. (2,2) Christie M. (1,0)
101	Adam M.	
201	David M.	David M. (2) Hugo F. (3)
301	David M.	David M. (1)
401	Christie M.	Adam M. (1) Christie M. (3,1) David M. (1) Eric E. (2) Fred E. (2,2) Hugo F. (3) Ivan L. (5)
501	Gail E.	Eric E. (1,0) Fred E. (5,0) Gail E. (4,0)
601	Eric E.	
701	Henry F.	Henry F. (6,0) Hugo F. (7,1) Bryan M. (1,1)
801	Hugo F.	
901	Ivan L.	Henry F. (1,0)

999	John I.	Bryan M. (2,2) Fred E. (3,1)
File Collection Intrinsic score		**Attached comments intrinsic score, by author**
File Collection number	**DCS score**	**CCS scores**
11	85	Bryan M., 1 Bryan M., 1 Christie M., 0
101	85	
201	100	David M., 0 Hugo F. 0
301	50	David M., 0
401	75	Adam M., 0 Christie M., 1 David M., 0 Eric E., 0 Fred E., 1 Fred E. 1 Hugo F., 0 Ivan L., 0
501	80	Eric E., 0 Fred E., 0 Gail E., 0
601	80	
701	40	Henry F., 0 Hugo F., 1 Hugo F., 0 Bryan M., 1
801	60	
901	70	Henry F., 0
999	0	Byran M., 1 Bryan M., 1 Fred E., 1 Fred E., 0
Taxonomy matches		
Christie M.		
Bryan M.		
File Collection number	**Original source**	
11	cnn.com	The source name here is truncated to the "root level" for simplification purposes. In reality it is the entire url tag. For example, http://www.cnn.com/2002/WORLD/meast/03/26/arab.league/index.html
101	nytimes.com	
201	microsoft.com	
301	bbc.com	
401	nytimes.com	
501	cnn.com	
601	nytimes.com	
701	latimes.com	
801	bbc.com	
901	corporate intranet	

[0069] For this example, 100 users having a total number of 1000 file collections in the repository yields 10 experts and 10 relevant file collections. There are also 10 comments that are found to be relevant. The enterprise in which the example takes place has four departments, namely marketing, engineering, finance, and legal. For ease of explanation, each employee's last name begins with the department in which they work.

[0070] Once the repository 104 (Figure 1) has been searched (step 506 - Figure 5) and all relevant documents located (step 508 - Figure 5), an Intrinsic Document Score (IDS) is calculated for each located document. This score is a weighted average between a Document Content Score (DCS) and a Comment Content Score (CCS). The DCS and CCS are calculated using any standard search engine techniques. CCS is the Comment Content Score calculated by any means such as semantic engine, frequency of words, *etc.*

[0071] Using formulae 7-12 above, the expertise of each searcher, creator, and/or contributor is then calculated. The calculations for F(User contribution) yield the results in Table 2 below.

TABLE 2

F(user contribution) User	File Collection	W_i by File collection	C_i by File collection	First sum in formula Details	Value	Second sum in formula	Taxonomy match
Adam M.							
	11	1	0	2*1*(85/100)^2	1.445	0	
	101	1	0	2*1*(85/100)^2	1.445	0	0
F(Adam M.)		2.89					
Bryan M.							
	11	0.5	0.063	2*(0.5+0.063)*(.85)^2	0.814	0	
	701	0.5	0.039	2*(0.5+0.039)*(.4)^2	0.172	0	
	999	0.5	0.063	2*(0.5+0.063)*0	0	0.063	2
F(Bryan M.)		3.049					
Christie M.							
	401	1	0.039	2*(1+0.039)*(75/100)^2	1.169	0.039	2
F(Christie M.)		3.208					
David M.							
	201	1	0	2*1*1	2	0	
	301	1	0	2*1*(0.5)^2	0.5	0	0
F(David M.)		2.5					
Eric E.							
	601	1	0	2*1*(.8)^2	1.28	0	0
F(Eric E.)		1.28					
Fred E.							
	401	0.5	0.063	2*(0.5+0.063)*.75^2	0.633		
	999	0.5	0.039	2*(0.5+0.039)*0	0	0.039	0
F(Fred E.)		0.672					
Gail E.							
	501	1	0	2*1*.8^2	1.28	0	0
F(Gail E.)		1.28					
Hugo F.							
	801	1	0	2*1*.6^2	0.72	0	0
F(Hugo F.)		0.72					
Ivan L.							
	901	1	0	2*1*.7^2	0.98	0	0
F(Ivan L.)		0.98					

[0072] Using formulae 10 and 11, G(Company Expertise) is calculated to be 2, while H(Query Specificity) is calculated to be 1.667. These values and the values in Table 2 are plugged into formula 7 to arrive at the following expertise values:

TABLE 3	
Name	RS-EXP-ABS
Adam M.	29
Bryan M.	30
Christie M.	32
David M.	25
Eric E.	13
Fred E.	7
Gail E.	13
Hugo F.	7
Henry F.	3
Ivan L.	10

[0073] W1(RS_EXP_ABS) is then calculated using formula 6 (for different searcher expertises) to yield the following results:

TABLE 4										
W(RS_EXP_ABS)										
Searcher Expertise / Name	0	5	10	15	20	25	30	35	40	45
Adam M.	0.29	0.24	0.19	0.14	0.09	0.04	-0.01	-0.06	-0.11	-0.16
Bryan M.	0.3	0.25	0.2	0.15	0.1	0.05	0	-0.05	-0.1	-0.15
Christie M.	0.32	0.27	0.22	0.17	0.12	0.07	0.02	-0.03	-0.08	-0.13
David M.	0.25	0.2	0.15	0.1	0.05	0	-0.05	-0.1	-0.15	-0.2
Eric E.	0.13	0.08	0.03	-0.02	-0.07	-0.12	-0.17	-0.22	-0.27	-0.32
Fred E.	0.07	0.02	-0.03	-0.08	-0.13	-0.18	-0.23	-0.28	-0.33	-0.38
Gail E.	0.13	0.08	0.03	-0.02	-0.07	-0.12	-0.17	-0.22	-0.27	-0.32
Hugo F.	0.07	0.02	-0.03	-0.08	-0.13	-0.18	-0.23	-0.28	-0.33	-0.38
Henry F.	0.03	-0.02	-0.07	-0.12	-0.17	-0.22	-0.27	-0.32	-0.37	-0.42
Ivan L.	0.1	0.05	0	-0.05	-0.1	-0.15	-0.2	-0.25	-0.3	-0.35

[0074] DCE and CCE are then calculated using formulae 4, 5, 13, and 14 (for different searcher expertises) to yield the following results:

TABLE 5

DCE Calculations

File collection ID	R1	W(searcher exp =0)	DCE (0)	
11	13	0.29	3.77	
101	13	0.29	3.77	
201	10	0.25	2.5	
301	20	0.25	5	
401	15	0.32	4.8	
501	14	0.13	1.82	
601	14	0.13	1.82	
701	18	0.03	0.54	
801	18	0.07	1.26	
901	16	0.1	1.6	

File collection ID	R1	W(searcher exp =30)	DCE (30)	
11	13	-0.01	-0.13	
101	13	-0.01	-0.13	
201	10	-0.05	-0.5	
301	20	-0.05	-1	
401	15	0.02	0.3	
501	14	-0.17	-2.38	
601	14	-0.17	-2.38	
701	18	-0.27	-4.86	
801	18	-0.23	-4.14	
901	16	-0.2	-3.2	

CCE Calculations				
File collection ID	"Delta X" or Dx			
	Searcher Exp = 0	Searcher Exp = 30	CCE (0)	CCE(30)
11	1.24	0.04	2.76	0.1
101	0	0	0	0
201	0.64	-0.56	1.55	-1.36
301	0.5	-0.1	1.22	-0.25
401	2.46	-1.74	4.21	-3.51
501	0.66	-1.14	1.59	-2.58
601	0	0	0	0
701	0.8	-1	1.9	-2.31
801	0	0	0	0
901	0.06	-0.54	0.15	-1.32

[0075] The Expertise Adjustment (EA) is then calculated according to formula 3 to yield the following results for EA:

TABLE 6			
Expertise Adjustment (EA)	Values for DCE and CCE are from Table 5 above		
File collection ID	**Searcher expertise = 0**		**Searcher expertise = 30**
11	6.53		-0.03
101	3.77		-0.13
201	4.05		-1.86
301	6.22		-1.25
401	9.01		-3.21
501	3.41		-4.96
601	1.82		-2.38
701	2.44		-7.17
801	1.26		-4.14
901	1.75		-4.52
	This entry is DCE+CCE when the searcher expertise is 0		This entry is DCE+CCE when the searcher expertise is 30

[0076] Finally, the adjusted score (RS_ADJ) for each located document is calculated using formula 1 to yield the following results:

TABLE 7				
File collection ID	Document Intrinsic score		RS_ADJ Adjusted score Searcher exp = 0	RS_ADJ Adjusted score Searcher exp = 30
11	85		92	85
101	85		89	85
201	100		100	98
301	50		56	49
401	75		84	72
501	80		83	75
601	80		82	78
701	40		42	33
801	60		61	56
901	70		72	65

[0077] In a similar manner, the adjusted scores are calculated when searching for sources as per tables 8-12 below.

TABLE 8		
Source name	File collection created from source	Creators of File collections
cnn.com	11,501	Adam M., Gail E.
microsoft.com	201	David M.
nytimes.com	101,401,601	Adam M., Christie M., Eric E.
bbc.com	801	Hugo F.
latimes.com	701	Henry F.

corporate intranet	901	Ivan L.

TABLE 9
SCS calculations

Source	SCS
cnn.com	85
microsoft.com	100
nytimes.com	85
bbc.com	60
latimes.com	40
corporate intranet	70

TABLE 10
R2 calculations

cnn.com	13
microsoft.com	10
nytimes.com	13
bbc.com	18
latimes.com	18
corporate intranet	16

TABLE 11
W2 Calculations

	Searcher Expertise = 0	Searcher Expertise = 30
cnn.com	0.29	-0.01
microsoft.com	0.25	-0.05
nytimes.com	0.32	0.02
bbc.com	0.07	-0.23
latimes.com	0.03	-0.27
corporate intranet	0.1	-0.2

TABLE 12
Adjusted relevancy scores

Source name	RSS_ADJ Searcher Expertise = 0	RSS_ADJ Searcher Expertise = 30	SCS Intrinsic score
cnn.com	89	85	85
microsoft.com	100	100	100
nytimes.com	89	85	85
bbc.com	61	56	60
latimes.com	41	35	40
corporate intranet	72	67	70

[0078] As can be seen the intrinsic scores of each document and/or source is adjusted to an adjusted score based on the expertise of the users. In other words, a document and/or source that may have been less relevant, is adjusted so that it is

more relevant, or *visa versa*. In this way, the precision of document and/or source relevancy is improved.

[0079] While the foregoing description and drawings represent preferred embodiments of the present invention, it will be understood that various additions, modifications and substitutions may be made therein without departing from the spirit and scope of the present invention as defined in the accompanying claims. In particular, it will be clear to those skilled in the art that the present invention may be embodied in other specific forms, structures, arrangements, proportions, and with other elements, materials, and components, without departing from the spirit or essential characteristics thereof. The presently disclosed embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims, and not limited to the foregoing description. Furthermore, it should be noted that the order in which the process is performed may vary without substantially altering the outcome of the process.

WHAT IS CLAIMED IS:

1. A method for personalizing information retrieval, comprising:

receiving at an information retrieval system a search request from a searcher, where said search request contains at least one search term;

searching a plurality of objects based on said at least one search term;

finding at least one located object from said plurality of objects, where said at least one located object is associated with said at least one search term;

calculating an intrinsic score for each located object, where said intrinsic score is based on said at least one search term;

adjusting said intrinsic score to an adjusted score based on the difference between a creator expertise of a creator of said at least one located object and/or a contributor expertise of a contributor to said at least one located object, and a searcher expertise of said searcher; and

transmitting a list to said searcher, where said list is based on said search request and said adjusted scores.

2. The method of claim 1, wherein said search request identifies said searcher with a unique user identifier.

3. The method of claim 1, wherein said plurality of objects include content objects, source objects, or people objects.

4. The method of claim31, wherein said adjusting further comprises:

locating objects associated with said searcher using said unique user identifier;

ascertaining said searcher expertise based on said search terms and said objects associated with said searcher;

determining said creator expertise and said contributor expertise for each located object;

raising said intrinsic score to said adjusted score for each located content or source object having a creator expertise higher than said searcher expertise;

lowering said intrinsic score to said adjusted score for each located content or source object having a creator expertise that is lower than said searcher expertise; and

changing said intrinsic score to said adjusted score for each located people object according to said searcher and creator expertise.

5. The method of claim 1, wherein said searching is undertaken using a search technique selected from a group consisting of: semantic processing, syntactic processing, natural language processing, statistical processing, and any combination of the aforementioned techniques.

6. The method of claim 3, wherein said adjusting comprises:

ascertaining said searcher expertise based on said search terms and any objects associated with said searcher;

determining said creator expertise and/or said contributor expertise for each located content or source object;

raising said intrinsic score to said adjusted score for each located content or source object having a creator expertise higher than said searcher expertise;

lowering said intrinsic score to said adjusted score for each located object having a creator expertise that is lower than said searcher expertise; and

changing said intrinsic score to said adjusted score for each located people object according to said searcher and creator expertise.

7. The method of claim 1, wherein said searcher expertise is dependent on the sum of the intrinsic scores of all located objects created by said searcher.

8. The method of claim 1, wherein said searcher expertise is dependent on the sum of all non-relevant documents not created by said searcher but containing relevant comments of said searcher.

9. The method of claim 1, wherein said searcher expertise is dependent on the sum of all non-relevant documents created by said searcher but containing relevant comments of said searcher.

10. The method of claim 1, wherein said searcher expertise is dependent on the total number of users relative to the total number of relevant users.

11. The method of claim 1, wherein said searcher expertise is dependent on the total number of objects in the repository relative to the total number of relevant objects in the repository.

12. The method of claim 1, wherein said searcher expertise is dependent on the searcher's taxonomy.

13. The method of claim 1, wherein said searcher expertise is dependent on whether the searcher created, edited, commented on, renamed, highlighted, linked, printed, copied or any activity that is monitored in a log file to said at least one located object.

14. The method of claim 1, wherein said creator expertise is dependent on the sum of the intrinsic scores of all located objects created by said creator.

15. The method of claim 1, wherein said creator expertise is dependent on the sum of all non-relevant documents not created by said creator but containing relevant comments of said creator.

16. The method of claim 1, wherein said creator expertise is dependent on the sum of all non-relevant documents created by said creator but containing relevant comments of said creator.

17. The method of claim 1, wherein said creator expertise is dependent on the total number of users relative to the total number of relevant users.

18. The method of claim 1, wherein said creator expertise is dependent on the total number of objects in the repository relative to the total number of relevant objects in the repository.

19. The method of claim 1, wherein said creator expertise is dependent on a creator's taxonomy.

20. The method of claim 1, wherein said creator expertise is dependent on whether the creator created, edited, commented on, renamed, highlighted, or linked, printed, copied or any activity that is monitored in a log file to said at least one located object.

21. The method of claim 1, wherein said contributor expertise is dependent on the sum of the intrinsic scores of all located objects created by said contributor.

22. The method of claim 1, wherein said contributor expertise is dependent on the sum of all non-relevant documents not created by said contributor but containing relevant comments of said contributor.

23. The method of claim 1, wherein said contributor expertise is dependent on the sum of all non-relevant documents created by said contributor but containing relevant comments of said contributor.

24. The method of claim 1, wherein said contributor expertise is dependent on the total number of users relative to the total number of relevant users.

25. The method of claim 1, wherein said contributor expertise is dependent on the total number of objects in the repository relative to the total number of relevant objects in the repository.

26. The method of claim 1, wherein said contributor expertise is dependent on the contributor's taxonomy.

27. The method of claim 1, wherein said contributor expertise is dependent on the whether the contributor created, edited, commented on, renamed, highlighted, linked, printed, copied or any activity that is monitored in a log file to said at least one located object.

28. The method of claim 1, further comprising, prior to said transmitting, sorting said list of located objects based on each object's adjusted score.

29. The method of claim 1, wherein said transmitting comprises sending a list of said located objects and their associated adjusted scores to said searcher.

30. The method of claim 1, wherein said transmitting comprises sending a list of people objects and their associated adjusted scores to said searcher.

31. The method of claim 1, wherein said transmitting comprises sending a list of source objects and their associated adjusted scores to said searcher.

32. A computer program product for personalizing information retrieval, the computer program product comprising a computer readable storage and a computer program stored therein, the computer program comprising:

instructions for receiving at an information retrieval system a search request from a searcher, where said search request contains at least one search term;

instructions for searching a plurality of objects based on said at least one search term;

instructions for finding at least one located object from said plurality of objects, where said at least one located object is associated with said at least one search term;

instructions for calculating an intrinsic score for each located object, where said intrinsic score is based on said at least one search term;

instructions for adjusting said intrinsic score to an adjusted score based on the difference between a creator expertise of a creator of said at least one located object and/or a contributor expertise of a contributor to said at least one located object, and a searcher expertise of said searcher; and

instructions for transmitting a list of said located objects and their associated adjusted scores to said searcher.

33. The computer program product of claim 31, wherein said computer program further comprises:

instructions for ascertaining said searcher expertise based on said search terms and any objects associated with said searcher;

instructions for determining said creator expertise and/or said contributor expertise for each located object;

instructions for raising said intrinsic score to said adjusted score for each located content or source object having a creator expertise higher than said searcher expertise;

instructions for lowering said intrinsic score to said adjusted score for each located content or source object having a creator expertise that is lower than said searcher expertise; and

instructions for changing said intrinsic score to said adjusted score for each located people object according to said searcher and creator expertise.

34. A system for personalizing information retrieval comprising:

at least one searcher device, creator device, and contributor device coupled to a network;

a repository containing one or more objects; and

an information retrieval system comprising:

a Central Processing Unit (CPU); and

a memory comprising:

instructions for receiving at an information retrieval system a search request from a searcher, where said search request contains at least one search term;

instructions for searching a plurality of objects based on said at least one search term;

instructions for finding at least one located object from said plurality of objects, where said at least one located object is associated with said at least one search term;

instructions for calculating an intrinsic score for each located object, where said intrinsic score is based on said at least one search term;

instructions for adjusting said intrinsic score to an adjusted score based on the difference between a creator expertise of a creator of said at least one located object and/or a contributor expertise of a contributor to said at least one located object, and a searcher expertise of said searcher; and

instructions for transmitting a list of said located objects and their associated adjusted scores to said searcher.

35. The system of claim 33, wherein said memory further comprises:

instructions for ascertaining said searcher expertise based on said search terms and any objects associated with said searcher;

instructions for determining said creator expertise and/or said contributor expertise for each located object;

instructions for raising said intrinsic score to said adjusted score for each located content or source object having a creator expertise higher than said searcher expertise;

instructions for lowering said intrinsic score to said adjusted score for each located content or source object having a creator expertise that is lower than said searcher expertise; and

instructions for changing said intrinsic score to said adjusted score for each located people object according to said searcher and creator expertise.

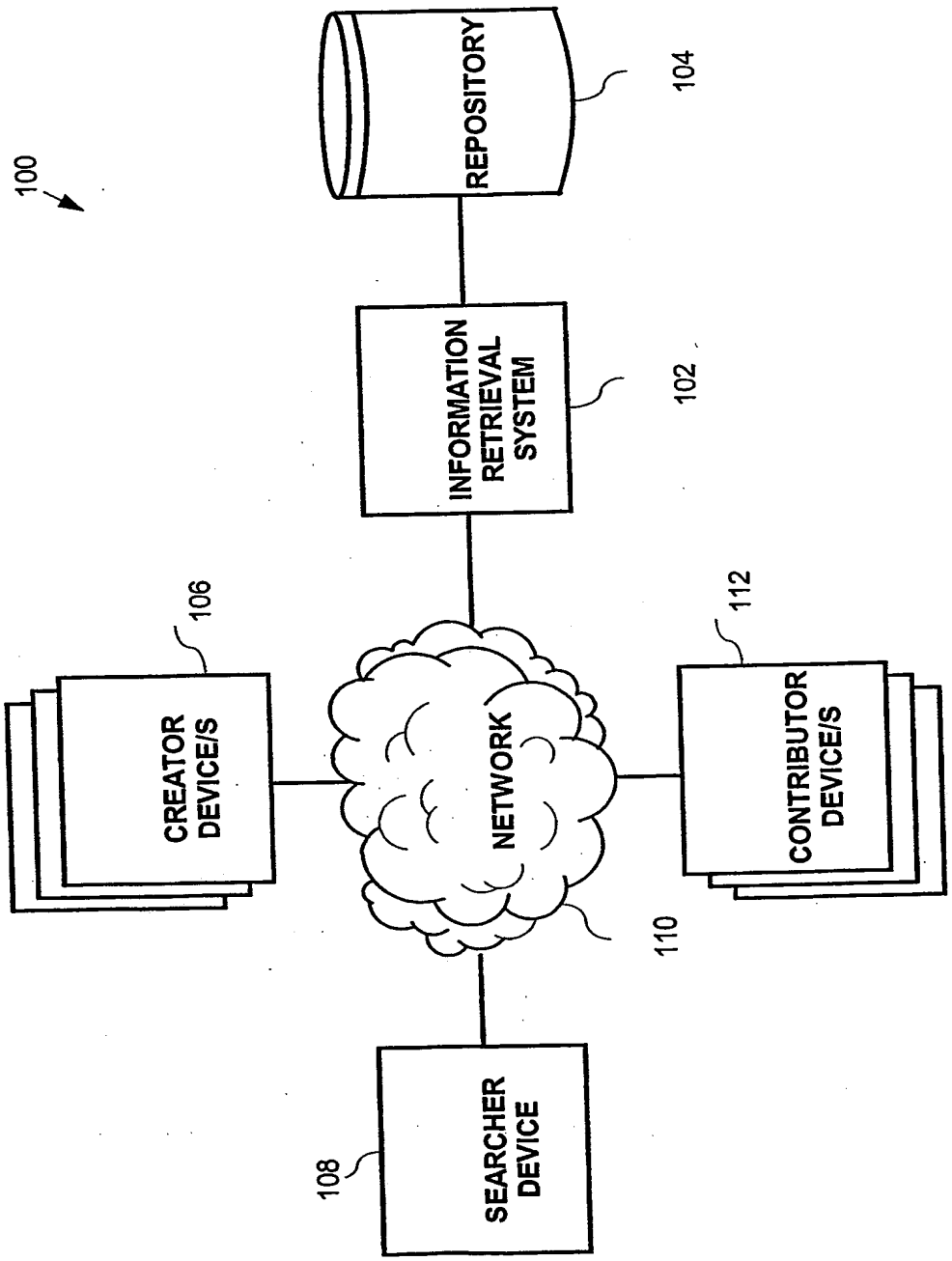

100
REPOSITORY
104
INFORMATION RETRIEVAL SYSTEM
102
106
CREATOR DEVICE/S
NETWORK
110
112
CONTRIBUTOR DEVICE/S
108
SEARCHER DEVICE

FIG. 1

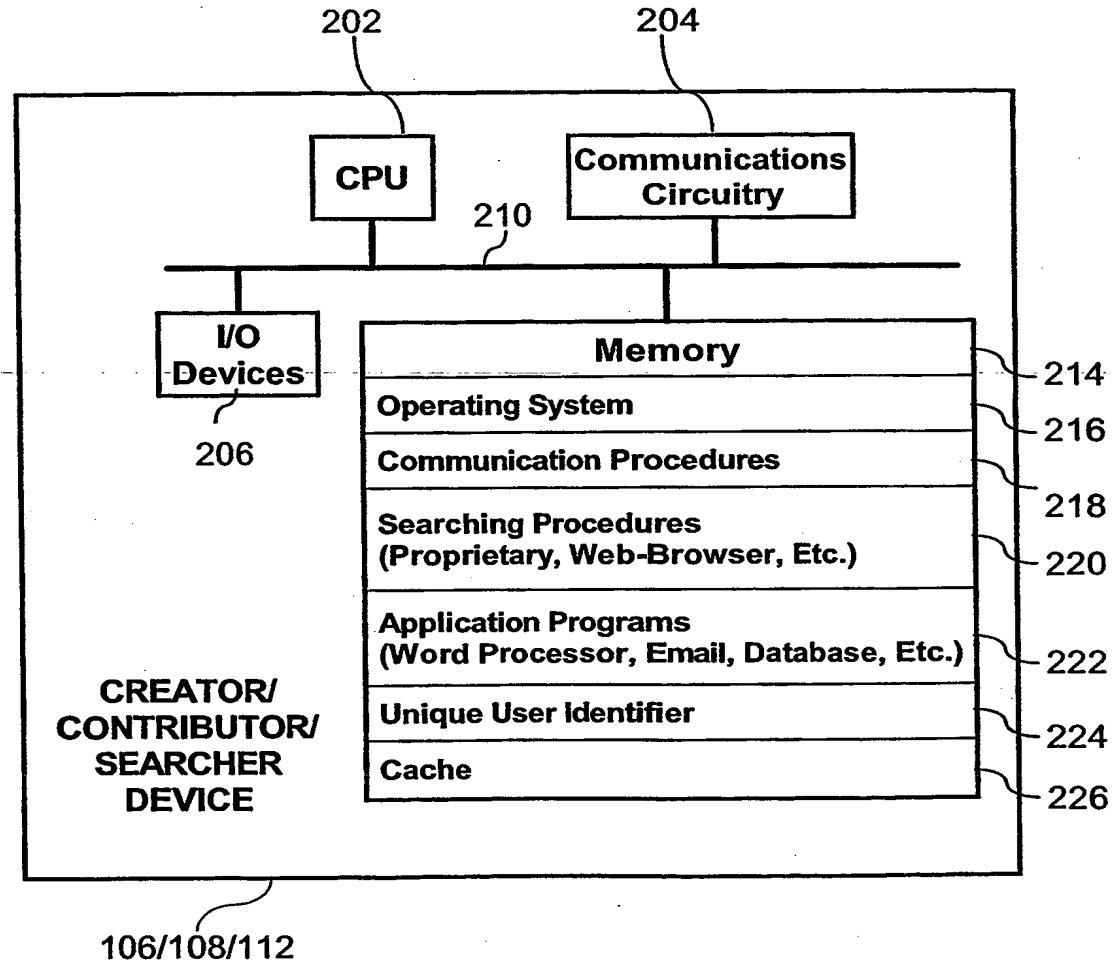


FIG. 2

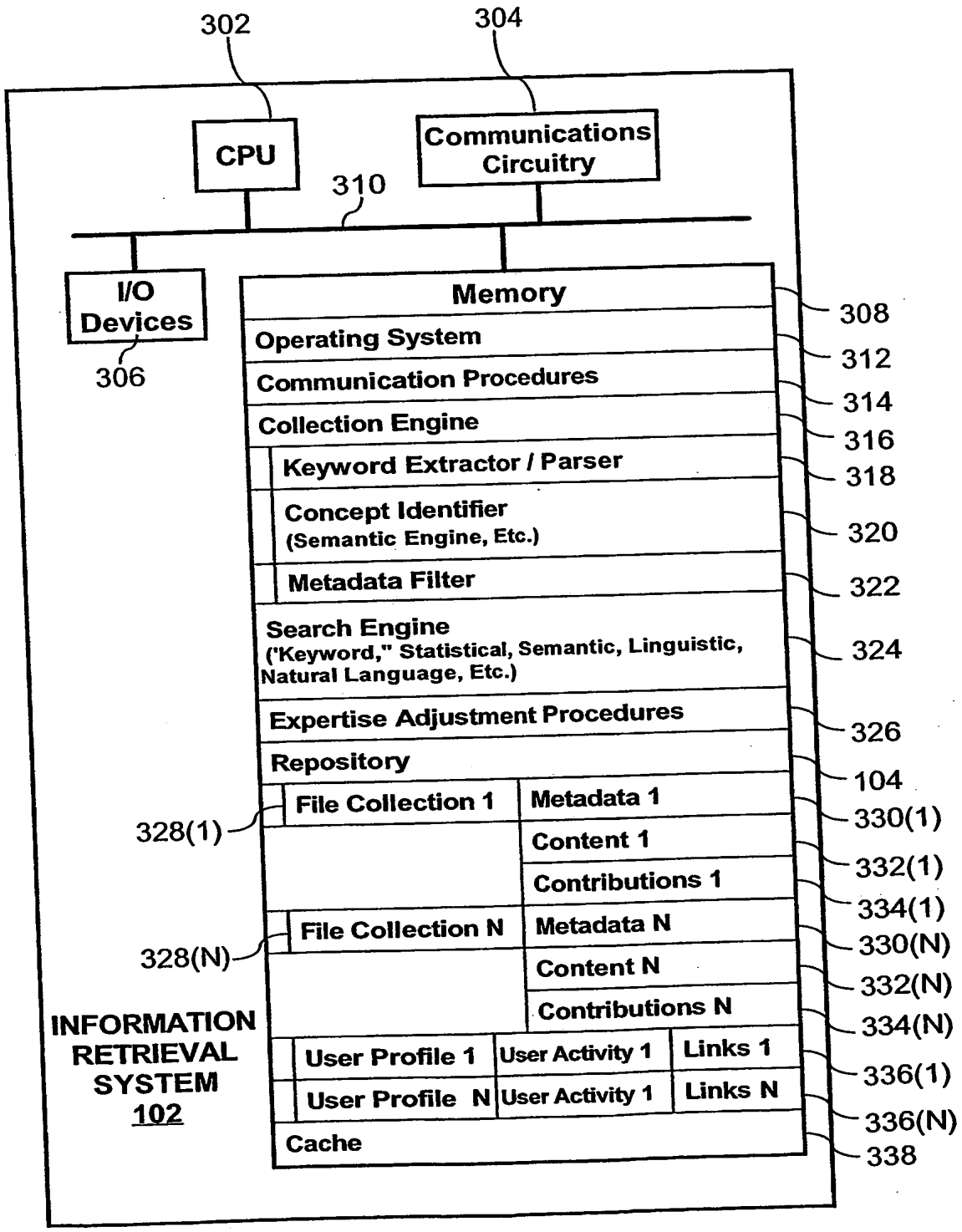

302
304
CPU
Communications Circuitry
310
I/O Devices
306
Memory
308
Operating System
312
Communication Procedures
314
Collection Engine
316
Keyword Extractor / Parser
318
Concept Identifier (Semantic Engine, Etc.)
320
Metadata Filter
322
Search Engine ('Keyword," Statistical, Semantic, Linguistic, Natural Language, Etc.)
324
Expertise Adjustment Procedures
326
Repository
104
328(1)
File Collection 1
Metadata 1
330(1)
Content 1
332(1)
Contributions 1
334(1)
328(N)
File Collection N
Metadata N
330(N)
Content N
332(N)
Contributions N
334(N)
INFORMATION RETRIEVAL SYSTEM 102
User Profile 1
User Activity 1
Links 1
336(1)
User Profile N
User Activity 1
Links N
336(N)
Cache
338

FIG. 3

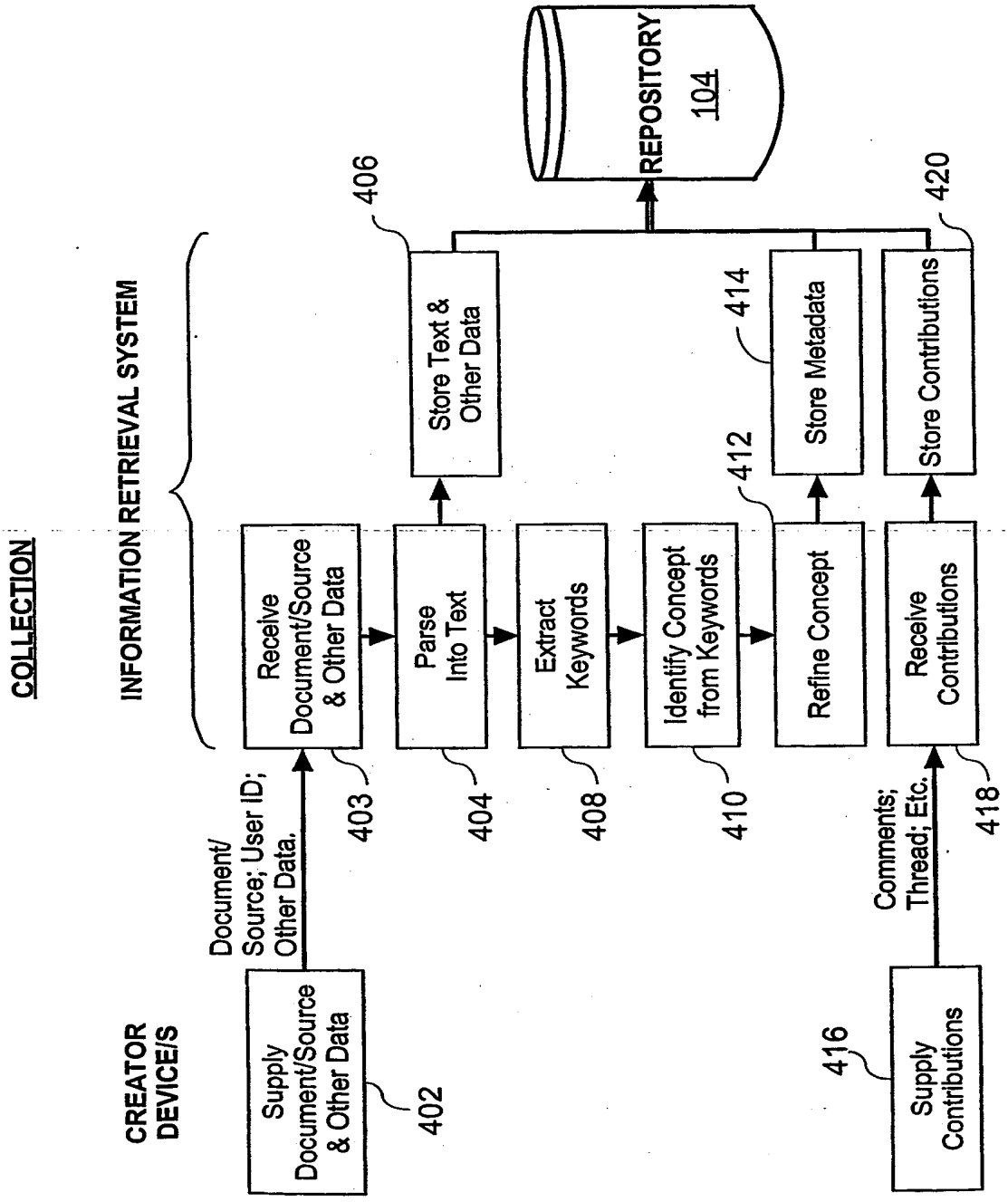


FIG. 4

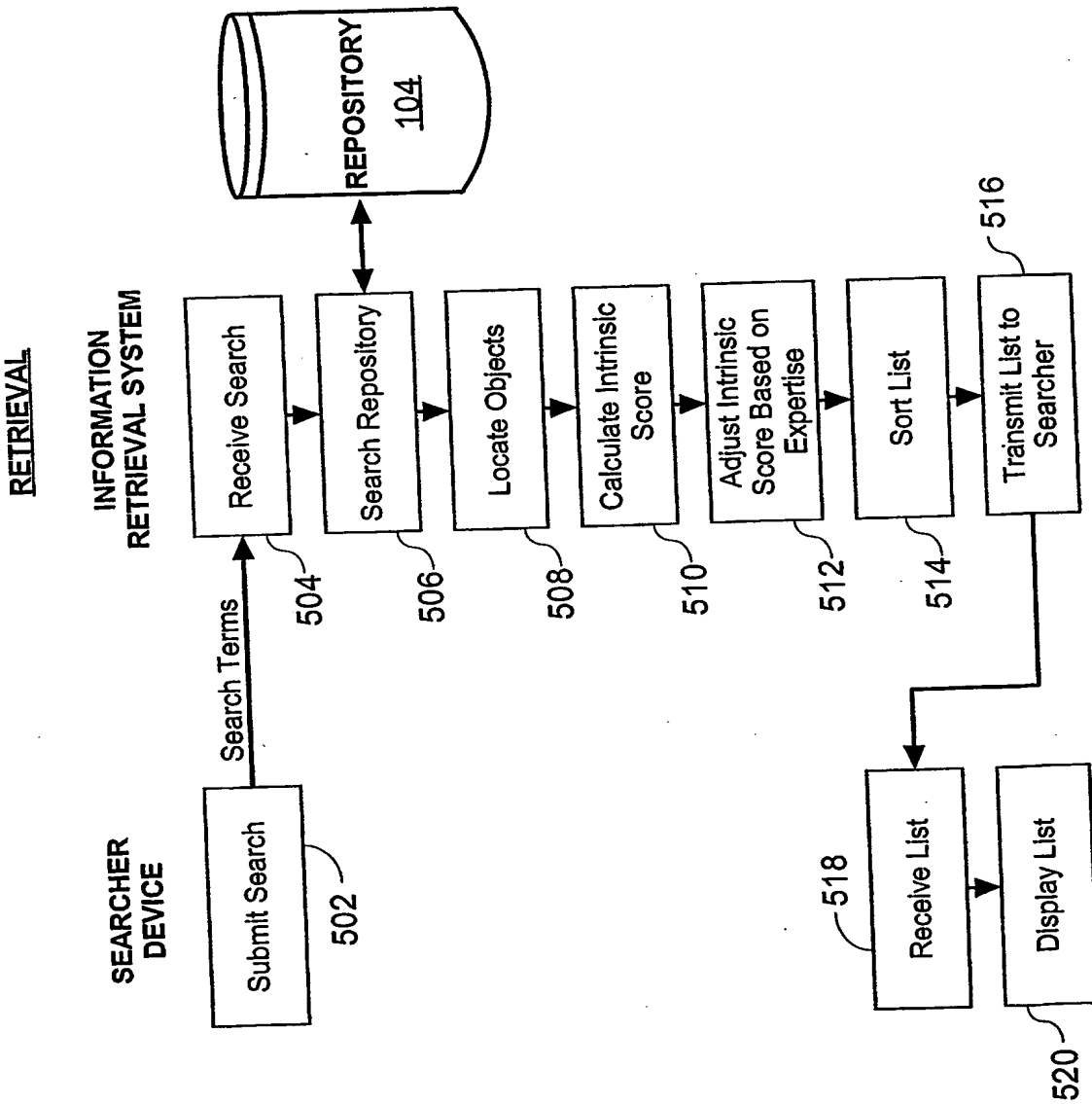

RETRIEVAL
INFORMATION RETRIEVAL SYSTEM
SEARCHER DEVICE
REPOSITORY 104
Submit Search
502
Search Terms
Receive Search
504
Search Repository
506
Locate Objects
508
Calculate Intrinsic Score
510
Adjust Intrinsic Score Based on Expertise
512
Sort List
514
Transmit List to Searcher
516
Receive List
518
Display List
520

FIG. 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2003 (24.12.2003)

PCT

(10) International Publication Number
WO 2003/107127 A3

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number:
PCT/US2003/018685

(22) International Filing Date: 12 June 2003 (12.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/172,165 14 June 2002 (14.06.2002) US

(71) Applicant: ENTOPIA, INC. [US/US]; 1301 Shoreway Road, Suite 302, Belmont, CA 94002 (US).

(72) Inventors: PERISIC, Igor; 739 West Capistrano Way, San Mateo, CA 94402-2012 (US). POSSE, Christian; 2833 21st Avenue W., Seattle, WA 98199-2910 (US).

(74) Agents: MORRIS, Francis, E. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

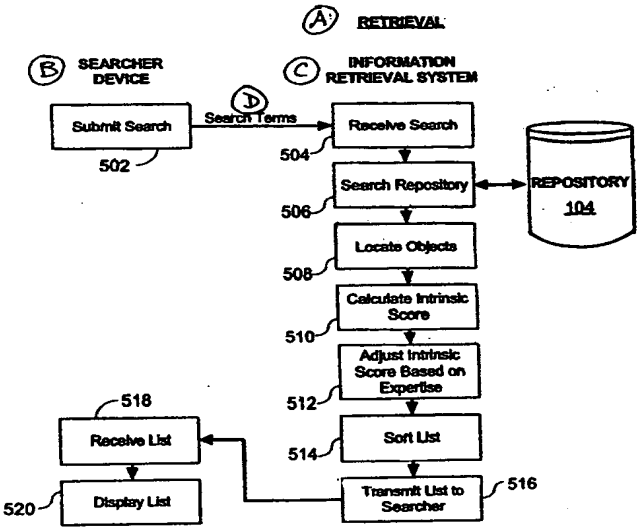
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
8 April 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM AND METHOD FOR PERSONALIZED INFORMATION RETRIEVAL BASED ON USER EXPERTISE



(57) Abstract: A search request is received at an information retrieval system from a searcher (Fig. 5, 504). The search request preferably contains at least one search term and a user identifier (502, 504). A plurality of objects (506) are then searched based on the at least one search term. At least one located object is found from the plurality of objects (508). The at least one located object (508) is associated with the search term(s). An intrinsic score (510) based on the search term(s) is subsequently calculated for each located object. The intrinsic score is then adjusted (512) to an adjusted score based on the difference between a creator expertise of a creator of the at least one located object and/or a contributor expertise of a contributor to the at least one located object, and a searcher expertise of the searcher (512).

WO 2003/107127 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/18685

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : G06F 17/30
US CL : 707/3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 707/3;707/4;707/5;707/6;707/10;707/104.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,974,412 (HAZLEHURST ET AL.) 26 October 1999 (26.10.1999), column 1, lines 1-67; column 2, lines 1-67	1-35
A	US 4,689,743 (CHIU) 25 August 1987 (25.08.1987), column 1, lines 1-67; column 2, lines 1-67; column 3, lines 1-67	1-35

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
05 January 2004 (05.01.2004)	21 JAN 2004
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Apu Mofiz Telephone No. 7036054240

Form PCT/ISA/210 (second sheet) (July 1998)

07-11-05

IFW

Express Mail No. EV 534 877 657 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Stephen R. Lawrence	Confirmation No.:	8147
Serial No.:	10/676,711	Art Unit:	2171
Filed:	September 30, 2003	Examiner:	To be determined
For:	*Personalization of Web Search*	Attorney Docket No.:	60963-0014-US

INFORMATION DISCLOSURE STATEMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the duty of disclosure provisions of 37 C.F.R. §1.56, there is hereby provided certain information which the Examiner may consider material to the examination of the subject U.S. patent application. It is requested that the Examiner make this information of record if it is deemed material to the examination of the application.

1. Enclosures accompanying this Information Disclosure Statement are:

1a. ☒ A list of all patents, publications, applications, or other information submitted for consideration by the office.

1b. A legible copy of :

☐ Each U.S. patent application publication and U.S. and foreign patent;

☐ Each publication or that portion which caused it to be listed on the PTO-1449;

☐ For each cited pending U.S. application, the application specification including the claims, and any drawing of the application, or portion of the application which caused it to be listed on the PTO-1449 including any claims directed to that portion;

☐ all other information or portion which caused it to be listed on the PTO-1449.

1c. ☐ An English language copy of search report(s) from a counterpart foreign application or PCT International Search Report.

1d. ☐ Explanations of relevancy (ATTACHMENT 1(d), hereto) or English language abstracts of the non-English language publications.

2. ☒ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(b):

☐ Within three months of the filing date of a national application other than a continued prosecution application under §1.53(d);

☐ Within three months of the date of entry of the national stage as set forth in §1.491 in an international application;

☒ Before the mailing of the first Office action on the merits;

☐ Before the mailing of a first Office action after the filing of a request for continued examination under §1.114.

3. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(c) after the period specified in 37 C.F.R §1.97(b), but before the mailing date of any of a final action under 37 C.F.R. §1.113, a notice of allowance under 37 C.F.R. §1.311 or an action that otherwise closes prosecution in the application.

(Check either Item 3a or 3b)

3a. ☐ The Certification Statement in Item 5 below is applicable. Accordingly, no fee is required.

3b. ☐ The $180.00 fee set forth in 37 C.F.R. §1.17(p) in accordance with 37 C.F.R. §1.97(c) is:
☐ enclosed
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

(Item 3b to be checked if any reference known for more than 3 months)

4. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(d) after the period specified in 37 C.F.R. §1.97(c), but on or before the date of payment of the issue fee.

(Check either Item 4a or 4b)

4a. ☐ The Certification Statement in Item 5 below is applicable.

4.b ☐ The $180.00 fee set forth in 37 C.F.R. §1.17(p) is:
☐ enclosed.
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

5. ☐ Certification Statement (applicable if Item 3a or Item 4a is checked)

(Check either Item 5a, 5b or 5c)

5a. ☐ In accordance with 37 C.F.R. §1.97(e)(1), it is certified that each item of information contained in this Information Disclosure Statement was first cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of this Information Disclosure Statement.

5b. ☐ Each item of information contained in this information disclosure statement was cited in a communication from a foreign patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

5c. ☐ Pursuant to 37 C.F.R. §1.704(d), each item of information contained in this information disclosure statement was cited in a communication from a foreign

patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

6. ☒ Copies of each cited U.S. patent and each U.S. patent application publication are not enclosed pursuant to the USPTO OG Notice dated 05 August 2003 waiving the requirement under 37 C.F.R. 1.98(a)(2)(i) for U.S. patent applications filed after June 30, 2003.

7. ☐ This application is a continuation application under 37 C.F.R. §1.53(b) or (d).

(Check appropriate Items 7a, 7b and/or 7c)

7a. ☐ A Petition to Withdraw from issue under 37 C.F.R. §1.313(b)(5) is concurrently filed herewith.

7b. ☐ Copies of publications listed on Form PTO-1449 from prior application Serial No. , filed on , of which this application claims priority under 35 U.S.C. §120, are not being submitted pursuant to 37 C.F.R. §1.98(d).

7c. ☐ Copies of the publications listed on Form PTO-1449 were not previously cited in prior application Serial No. , filed on , and are provided herewith.

8. ☐ This is a Supplemental Information Disclosure Statement. (Check Item 8a)

8a. ☐ This Supplemental Information Disclosure Statement under 37 C.F.R. §1.97(f) supplements the Information Disclosure Statement filed on . A bona fide attempt was made to comply with 37 C.F.R. §1.98, but inadvertent omissions were made. These omissions have been corrected herein. Accordingly, additional time is requested so that this Supplemental Information Disclosure Statement can be considered as if properly filed on .

9. ☐ In accordance with 37 C.F.R. §1.98, a concise explanation of what is presently understood to be the relevance of each non-English language publication is:

(Check Item 9a, 9b, or 9c)

9a. ☐ satisfied because all non-English language publications were cited on the enclosed English language copy of the PCT International Search Report or the search report from a counterpart foreign application indicating the degree of relevance found by the foreign office.

9b. ☐ set forth in the application.

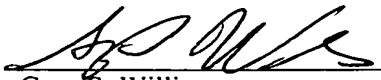
9c. ☐ enclosed as an attachment hereto.

10. ☒ The Commissioner is authorized to charge any additional fee required or credit any overpayment for this Information Disclosure Statement and/or Petition to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no. 60963-0014-US).

11. ☒ No admission is made that the information cited in this Statement is, or is considered to be, material to patentability nor a representation that a search has been made (other than a search report of a foreign counterpart application or PCT International Search Report if submitted herewith). 37 C.F.R. §§1.97(g) and (h).

Respectfully submitted,

Date: July 8, 2005



31,066

Gary S. Williams (Reg. No.)
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, CA 94306
(650) 843-4000

INFORMATION DISCLOSURE CITATION PTO-1449				*Complete If Known*	
				Application Number	10/676,711
				Filing Date	September 30, 2003
				First Named Inventor	Stephen R. Lawrence
				Art Unit	2171
				Examiner Name	To be determined
Sheet	1	of		Attorney Docket Number	60963-0014-US

U.S. PATENT DOCUMENTS							
Examiner Initials	Cite No.	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate
		6,912,505 B2	06/28/2005	Linden et al.	705	14	

FOREIGN PATENT DOCUMENTS								
Examiner Initials	Cite No.	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation	
							Yes	No

OTHER NON PATENT LITERATURE DOCUMENTS		
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

01-03-06

IFW

Express Mail No. EV 533 716 335 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Stephen R. Lawrence	Confirmation No.:	8147
Serial No.:	10/676,711	Art Unit:	2171
Filed:	September 30, 2003	Examiner:	To be determined
For:	*Personalization of Web Search*	Attorney Docket No.:	60963-0014-US

INFORMATION DISCLOSURE STATEMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the duty of disclosure provisions of 37 C.F.R. §1.56, there is hereby provided certain information which the Examiner may consider material to the examination of the subject U.S. patent application. It is requested that the Examiner make this information of record if it is deemed material to the examination of the application.

1. Enclosures accompanying this Information Disclosure Statement are:

1a. ☒ A list of all patents, publications, applications, or other information submitted for consideration by the office.

1b. A legible copy of :

☐ Each U.S. patent application publication and U.S. and foreign patent;

☐ Each publication or that portion which caused it to be listed on the PTO-1449;

☐ For each cited pending U.S. application, the application specification including the claims, and any drawing of the application, or portion of the application which caused it to be listed on the PTO-1449 including any claims directed to that portion;

☐ all other information or portion which caused it to be listed on the PTO-1449.

1c. ☒ An English language copy of search report(s) from a counterpart foreign application or PCT International Search Report.

1d. ☐ Explanations of relevancy (ATTACHMENT 1(d), hereto) or English language abstracts of the non-English language publications.

2. ☒ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(b):

☐ Within three months of the filing date of a national application other than a continued prosecution application under §1.53(d);

☐ Within three months of the date of entry of the national stage as set forth in §1.491 in an international application;

☒ Before the mailing of the first Office action on the merits;

☐ Before the mailing of a first Office action after the filing of a request for continued examination under §1.114.

3. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(c) after the period specified in 37 C.F.R §1.97(b), but before the mailing date of any of a final action under 37 C.F.R. §1.113, a notice of allowance under 37 C.F.R. §1.311 or an action that otherwise closes prosecution in the application.

(Check either Item 3a or 3b)

3a. ☐ The Certification Statement in Item 5 below is applicable. Accordingly, no fee is required.

3b. ☐ The $180.00 fee set forth in 37 C.F.R. §1.17(p) in accordance with 37 C.F.R. §1.97(c) is:
☐ enclosed
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

(Item 3b to be checked if any reference known for more than 3 months)

4. ☐ This Information Disclosure Statement is filed under 37 C.F.R. §1.97(d) after the period specified in 37 C.F.R. §1.97(c), but on or before the date of payment of the issue fee.

(Check either Item 4a or 4b)

4a. ☐ The Certification Statement in Item 5 below is applicable.

4.b ☐ The $180.00 fee set forth in 37 C.F.R. §1.17(p) is:
☐ enclosed.
☐ to be charged to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no.).

5. ☐ Certification Statement (applicable if Item 3a or Item 4a is checked)

(Check either Item 5a, 5b or 5c)

5a. ☐ In accordance with 37 C.F.R. §1.97(e)(1), it is certified that each item of information contained in this Information Disclosure Statement was first cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of this Information Disclosure Statement.

5b. ☐ Each item of information contained in this information disclosure statement was cited in a communication from a foreign patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

5c. ☐ Pursuant to 37 C.F.R. §1.704(d), each item of information contained in this information disclosure statement was cited in a communication from a foreign

patent office in a counterpart application, and the communication was not **received** by any individual designated in 37 C.F.R. §1.56(c) more than thirty days prior to the filing of this information disclosure statement.

6. ☒ Copies of each cited U.S. patent and each U.S. patent application publication are not enclosed pursuant to the USPTO OG Notice dated 05 August 2003 waiving the requirement under 37 C.F.R. 1.98(a)(2)(i) for U.S. patent applications filed after June 30, 2003.

7. ☐ This application is a continuation application under 37 C.F.R. §1.53(b) or (d).

(Check appropriate Items 7a, 7b and/or 7c)

7a. ☐ A Petition to Withdraw from issue under 37 C.F.R. §1.313(b)(5) is concurrently filed herewith.

7b. ☐ Copies of publications listed on Form PTO-1449 from prior application Serial No. , filed on , of which this application claims priority under 35 U.S.C. §120, are not being submitted pursuant to 37 C.F.R. §1.98(d).

7c. ☐ Copies of the publications listed on Form PTO-1449 were not previously cited in prior application Serial No. , filed on , and are provided herewith.

8. ☐ This is a Supplemental Information Disclosure Statement. (Check Item 8a)

8a. ☐ This Supplemental Information Disclosure Statement under 37 C.F.R. §1.97(f) supplements the Information Disclosure Statement filed on . A bona fide attempt was made to comply with 37 C.F.R. §1.98, but inadvertent omissions were made. These omissions have been corrected herein. Accordingly, additional time is requested so that this Supplemental Information Disclosure Statement can be considered as if properly filed on .

9. ☐ In accordance with 37 C.F.R. §1.98, a concise explanation of what is presently understood to be the relevance of each non-English language publication is:

(Check Item 9a, 9b, or 9c)

9a. ☐ satisfied because all non-English language publications were cited on the enclosed English language copy of the PCT International Search Report or the search report from a counterpart foreign application indicating the degree of relevance found by the foreign office.

9b. ☐ set forth in the application.

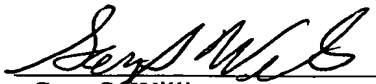
9c. ☐ enclosed as an attachment hereto.

10. ☒ The Commissioner is authorized to charge any additional fee required or credit any overpayment for this Information Disclosure Statement and/or Petition to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no. 60963-0014-US).

11. ☒ No admission is made that the information cited in this Statement is, or is considered to be, material to patentability nor a representation that a search has been made (other than a search report of a foreign counterpart application or PCT International Search Report if submitted herewith). 37 C.F.R. §§1.97(g) and (h).

Respectfully submitted,

Date: December 29, 2005



31,066 (Reg. No.)

Gary S. Williams
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, CA 94306
(650) 843-4000

INFORMATION DISCLOSURE CITATION PTO-1449			*Complete If Known*	
			Application Number	10/676,711
			Filing Date	September 20, 2003
			First Named Inventor	Stephen R. Lawrence
			Art Unit	2171
			Examiner Name	To be determined
Sheet 1	of	1	Attorney Docket Number	60963-0014-US

U.S. PATENT DOCUMENTS							
Examiner Initials	Cite No.	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate
		5,724,567	03/03/1998	Rose et al.	395	602	
		5,754,939	05/19/1998	Herz et al.	455	4.200	
		6,385,619 B1	05/07/2002	Eichstaedt et al.	707	1	
		2002/044571 A1	03/04/2004	Bronniman	705	14	

FOREIGN PATENT DOCUMENTS								
Examiner Initials	Cite No.	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation	
							Yes	No

OTHER NON PATENT LITERATURE DOCUMENTS		
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published
		International Search Report for International Application No. PCT/US2005/025081, mailed December 2, 2005

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	1	("6981040".pn. or "6006218".pn.) and distance	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 16:12
L2	36	url same distance same (weigh$4 scor$4) and (search adj (result$2 hit$2))	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 13:36
L3	125	user adj profile$2 same demographic same geographic	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 13:47
L4	1	user adj profile$2 same demographic same geographic same search adj result$2	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 13:36
L5	230	user adj profile$2 same demographic same (location geographic)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 13:47
L6	1	user adj profile$2 same demographic same (location geographic) same search adj (hit$2 result$2)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 13:47
L7	51	user adj profile$2 same demographic same (location geographic) and search adj (hit$2 result$2)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 13:47
L8	2	("6981040".pn. or "6006218".pn.) and demographic	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 13:59
L9	1	("6981040".pn. or "6006218".pn.) and URL	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 17:02
L10	1	("6981040".pn. or "6006218".pn.) and negative	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 17:15

EAST Search History

L11	2	("6981040".pn. or "6006218".pn.) and group	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 17:15
S1	1	"676711".apn.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 17:07
S2	62	("20020082901" or "20020049635" or "20040267725" or "20050080772" or "20030149937" or "20040267806" or "20020044571" or "20020024532" or "20020198882" or "20030233345" or "5754939" or "5933827" or "5999975" or "6006218" or "6134532" or "6182050" or "6366956" or "6493702" or "6711585" or "6721713" or "6981040" or "7031961" or "6144944" or "6606619" or "6535888" or "6868525" or "5724567" or "5754939" or "6385619" or "6912505" or "6285999" or "6327590").pn.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:00
S3	6	"6892198"	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:00
S4	2	"6892198".pn.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:00
S5	8	search adj result$2 same promot$4 same profil$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:44
S6	10	generic with (scor$4 rank$4 weight$4) with search adj result	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:34
S12	363	search adj result$2 same (re$rank$4 re$scor$4 re$order$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 16:06
S13	53	search adj result$2 same (re$rank$4 re$scor$4 re$order$4) same (profile interest)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:58
S14	51	S13 not google.as.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:48

EAST Search History

S15	34	S12 same weight$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:58
S16	33	S12 same weight$4 not google.as.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 15:58
S17	350	search adj result$2 same (re$rank$4 re$scor$4 re$order$4) not google.as.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 16:07
S18	139	search adj result$2 same (re$rank$4 re$scor$4) not google.as.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/16 16:26
S21	1	"6981040".pn. and (scor$4 rank$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/20 11:21
S22	1	"676711".apn. and quer$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/19 15:42
S23	723868	"6006218".pn. and search adj information or entries	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/19 15:42
S24	2	"6006218".pn. and (search adj information or entries)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/19 15:43
S25	1	"6006218".pn. and (seen viewed)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/19 15:44
S26	1	"6981040".pn. and link$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:41

EAST Search History

S27	3	"6981040".pn. or "6006218".pn. and (demographic$$ geographic$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:41
S28	3	"6981040".pn. or "6006218".pn. and (demographic$4 geographic$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:42
S29	2	("6981040".pn. or "6006218".pn.) and (demographic$4 geographic$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:42
S30	0	("6981040".pn. or "6006218".pn.) and (demographic$4) and (region$4 geographic$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:42
S31	2	("6981040".pn. or "6006218".pn.) and (demographic$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:43
S32	1	("6981040".pn. or "6006218".pn.) and (demographic$4) and (location$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:45
S33	1	("6981040".pn. or "6006218".pn.) and (probabili$6 same frequenc$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:46
S34	1	("6981040".pn. or "6006218".pn.) and (site adj probabilit$6)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 13:47
S35	3	("6981040".pn. or "6006218".pn.) and (interest$4 same probabilit$6)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 15:25
S36	1	("6981040".pn. or "6006218".pn.) and (model$4 same client$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/22 10:34

EAST Search History

S37	2	("6981040".pn. or "6006218".pn.) and (group$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 15:39
S38	1	"676711".apn. and interleav$6	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 17:23
S39	78	search adj result$2 same (interleav$6 intermix$6)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 17:09
S40	69	interleav$6 same search adj (result$2 hit$2)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 17:24
S41	94	(interleav$6 blend$4) same search adj (result$2 hit$2)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 17:25
S42	94	S41 not google.as.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/03/21 17:25

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.
10/676,711	09/30/2003	Stephen R. Lawrence	60963-0014-US	8147

24341 7590 04/02/2007
MORGAN, LEWIS & BOCKIUS, LLP.
2 PALO ALTO SQUARE
3000 EL CAMINO REAL
PALO ALTO, CA 94306

EXAMINER	
LU, CHARLES EDWARD	
ART UNIT	PAPER NUMBER
2161	

SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE
3 MONTHS	04/02/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Office Action Summary	Application No.	Applicant(s)	
	10/676,711	LAWRENCE, STEPHEN R.	
	Examiner	Art Unit	
	Charles E. Lu	2161	

-- 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) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, 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 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 _____.

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) *1-58* 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) *1-58* 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 *30 September 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/SB/08) Paper No(s)/Mail Date *12/29/5;7/8/5;3/3/5;2/24/4*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-58 are pending.

2. Claims 1-58 are rejected.

Drawings

3. The drawings are objected to because of the following informalities:

As to **fig. 4B**, elements **#342-1...342-n** should be shown in the specification.

As to **fig. 9A**, element **#945** should be shown in the specification.

Corrected drawing sheets in compliance with 37 CFR 1.121(d) are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. The figure or figure number of an amended drawing should not be labeled as "amended." If a drawing figure is to be canceled, the appropriate figure must be removed from the replacement sheet, and where necessary, the remaining figures must be renumbered and appropriate changes made to the brief description of the several views of the drawings for consistency. Additional replacement sheets may be necessary to show the renumbering of the remaining figures. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). If the changes are not accepted by the examiner,

the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

Appropriate corrections are required.

Specification

4. The disclosure is objected to because of the following informalities:

In **para. 0047**, the element numbers do not appear to be shown in the drawings.

In the pseudo code of **para. 0061**, seventh line from bottom, the word "value" appears to be misspelled.

The title is not sufficiently descriptive of the claimed invention.

The following title is suggested:

PERSONALIZATION OF WEB SEARCH RESULTS USING TERM, CATEGORY, AND LINK-BASED USER PROFILES

Appropriate corrections are required.

Claim Objections

5. Claim 25 is objected to because of the following informalities:

As to claim 25, the occurrence of "scored" should be changed to scores.

Appropriate corrections are required.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

6. Claims 1-58 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

As to claim 1, the claim contains an abstract idea (e.g., "accessing, receiving, identifying, assigning, ranking" etc.). Therefore, the claim must be drawn to a practical application of the abstract idea, which may be established through either a physical transformation or a useful, concrete, and tangible result. See MPEP 2106.

The claim does not cause a physical transformation. For example, the steps of "accessing, receiving, identifying, assigning, ranking" are reasonably understood to one of ordinary skill in the art as merely data manipulation without actually producing any physical transformation.

The claim does not produce a useful, concrete, and tangible result. Merely "accessing, receiving, identifying, assigning, ranking" is believed to be an abstract manipulation, failing to enable the "useful, concrete, and tangible" to be realized. The claimed invention as a whole must produce a "useful, concrete and tangible result." Emphasis added. *State Street, 149 F.3d at 1373-74, 47 USPQ2d at 1601-02*. See MPEP 2106.

This discussion also applies for all other pending claims.

As to claims 43-58, the claims are reasonably interpreted to be software per se, which is non-statutory. See MPEP 2106.

Art rejection of the above claims is applied in anticipation of Applicant amending the claims to overcome the rejection under 35 U.S.C. 101, discussed above.

Claim Rejections - 35 USC § 102

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.

7. Claims 1, 27, and 43 are rejected under 35 U.S.C. 102(b) as being anticipated by Breese et al (U.S. Patent 6,006,218).

As to claim 1, Breese teaches the following claimed subject matter:

A method of personalizing search results of a search engine, comprising: accessing a user profile for a user based on information about the user (fig. 2B, #224, fig. 5, col. 5, ll. 20-45),

The user information including information derived from a set of documents (col. 5, ll. 20-45),

The set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user (col. 5, ll. 20-45);

Receiving a search query from the user (col. 6, ll. 60-65);

Identifying a set of search result documents that match the search query (fig. 2C, #230-231);

Assigning a generic score to each of at least a plurality of the search result documents (col. 7, ll. 18-45);

Assigning a personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the user profile (col. 7, ll. 18-45, details on col. 8-17); and

Ranking the set of search result documents according to their personalized scores (col. 7, ll. 18-45, details on col. 8-17).

Claims 27 and 43 are drawn to substantially the same subject matter as claim 1, discussed above.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

8. Claims 2-3, 18, 22, 28-29 and 44-45 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese et al (U.S. Patent 6,006,218).

As to claims 2 and 3, Breese does not expressly teach wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have/have not been viewed by the user.

However, Breese teaches that the user information includes previous search information (col. 5, ll. 30-33, col. 16, l. 40) and that the search information may include

information on the entries that were presented to the user as a result of the search (i.e. search results). Furthermore, Breese states, "it may be assumed that the user is aware of these entries, or at least the highest ranked entries (col. 16, ll. 34-50)." As discussed above, the user information includes information on previous Internet site access operations (col. 5, ll. 30-35).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that "wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have/have not been viewed by the user" is implemented. The motivation would have been to enhance the effectiveness of the retrieval result adjustor, because data regarding actual document views would be used.

As to claim 18, Breese teaches the following claimed subject matter:

A method of personalizing search results of a search engine, comprising: creating a plurality of user profiles for a plurality of users, each user profile including at least a user's identification number and information derived from documents visited by the user (col. 5, ll. 20-45);

Receiving a search query from a user of the plurality of users, the search query including at least one query term (e.g., col. 8, ll. 62-66).

Selecting a set of documents from the Internet, assigning to each document in the set a generic score that characterizes the relevance of the document to the at least one query term (discussed above);

Retrieving the user's user profile and assigning to each of the set of documents a profile score based on the user profile (discussed above); and

Ranking the set of documents according to their generic and profile scores (discussed above).

Breese does not expressly teach the search query including the user's identification number.

However, Breese teaches that a user has a unique identification number for storing user attributes in a user database (col. 5, ll. 20-45), and that information regarding the user and the search to be performed is obtained at the input step 222 (col. 8, ll. 15-20, #224).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that the search query includes the user's identification number. The motivation would have been to adapt to specific user requirements in setting up the search engine.

As to claim 22, Breese as applied above further teaches wherein the documents visited by the user from which information is derived for use in a particular user's user profile is selected based on the user's activities when visiting the documents (e.g., col. 5, ll. 20-45).

Claims 28-29 and 44-45 are drawn to substantially the same subject matter as claims 2-3, discussed above.

9. Claim 21 is rejected under 35 U.S.C. 103(a) as being unpatentable over Breese et al (U.S. Patent 6,006,218) in view of Gerace (U.S. Patent 5,848,396).

As to claim 21, Breese teaches wherein the user profile includes demographic information provided by the user (fig. 5).

Breese does not expressly teach geographic information.

However, Gerace teaches a user profile including both demographic and geographic information (col. 5, l. 63 – col. 6, l. 15).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that geographic information is additionally stored with the user profile. The motivation would have been to store more information about the user to facilitate better decisions by the information retrieval system.

10. Claim 4-7, 9-17, 19-20, 23-24, 30-33, 35-42, 46-49, and 51-58 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese et al (U.S. Patent 6,006,218) in view of Konig et al (U.S. Patent 6,981,040).

As to claims 4-5 and 14, Breese does not expressly teach updating the user profile by updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile; and updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

However, Konig teaches updating a term-based, and category-based profile as claimed (col. fig. 4A, fig. 4C, col. 10, l. 51, col. 12, l. 55). Furthermore, Konig teaches analyzing links and locations in documents for updating a user model (user topic distribution, user site distribution, and see fig. 13, fig. 4, fig. 15A, col. 17, l. 47, col. 15, ll. 8-15, col. 18, ll. 17-25, col. 22, l. 64 – col. 23, l. 10).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that the claimed updating of the term-based, category-based, and link-based profiles is implemented with appropriate weights associated with each item (see e.g., fig. 4). The motivation would have been to represent user interest in a document or product independently of any specific user information need, as taught by Konig (col. 4, ll. 1-7). This would further enhance search results when combined with Breese.

As to claim 6, Konig as applied above further teaches wherein the link-based profile includes information about URLs or portions of URLs (fig. 4).

As to claim 7, Konig as applied above further teaches or suggests wherein the link-based profile comprises a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host (col. 12, ll. 28-54, col. 23, ll. 1-10).

As to claim 9, Konig as applied above further teaches wherein a term in the term-based profile is an expression comprising at least one word and a weight (fig. 4A).

As to claim 10, Konig as applied above further teaches wherein the weight is a weight associated with occurrences of the term in the set of documents (fig. 4, col. 10, l. 52 – col. 12, l. 55).

As to claim 11, Konig as applied above further teaches wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents (col. 10, l. 52 – col. 11, l. 20).

As to claim 12, Konig as applied above further teaches wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories (fig. 4, 7, 8, col. 15, ll. 7-32).

As to claim 13, Konig as applied above further teaches wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin (e.g., col. 15, ll. 7-15 and see above).

As to claim 15, Breese discloses a user profile and a search engine (fig. 1), but does not expressly teach wherein the user profile is stored on a server of the search engine.

However, Konig teaches wherein the user profile is stored on a server of the search engine (fig. 1).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that the user profile is stored on a server of the search engine. The motivation would have been to adapt to the requirements of the user in setting up the search system, or to provide personalized services for simultaneous clients, as taught by Konig (col. 7, ll. 20-25).

As to claim 16, Breese and Konig both disclose a user profile and a client associated with the user (Breese and Konig, fig. 1), but do not expressly teach wherein the user profile is stored on the client.

However, it has been held that rearrangement of parts is obvious. *In re Japikse*, 181 F.2d 1019, 86 USPQ 70. It has also been held that making integral and/or separable is obvious. *In re Larson*, 340 F.2d 965, 968, 144 USPQ 347, 349 (CCPA 1965), *In re Dulberg*, 289 F.2d 522, 523, 129 USPQ 348, 349 (CCPA 1961). Furthermore, as taught by Konig, each user has his/her own user profile (Konig, col. 7, ll. 27-28).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that each user profile is stored on the user's (client) computer, and operation of Breese otherwise remains the same. The motivation would have been to adapt to the requirements of the user in setting up the search system. For example, one may desire to store personal user data on the client computer, rather than a public server.

As to claim 17, Breese does not expressly teach wherein the user is a group of users.

However, Konig teaches wherein the user is a group of users (col. 20, ll. 24-28, col. 9, ll. 47-52).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that the user is a group of users. The motivation would have been to represent the interest level of a group of users in a document independently of any specific information need, as taught by Konig (col. 9, ll. 47-52).

Claims 19-20 are drawn to substantially the same subject matter as claims 4-5, discussed above, in addition to creating, which must happen in Konig in order to store the relevant data (see e.g., fig. 4).

As to claim 23, the "storing" limitation is addressed with respect to claim 15 above. Breese, as applied above, further teaches the retrieving including the user's user profile based on an identification number associated with the user and the user's profile (col. 5, ll. 23-30). Note that Breese must retrieve the data in order to process it.

Claim 24 is drawn to substantially the same subject matter as claim 16, discussed above.

Claims 30-33, 35-42, 46-49, and 51-58 are drawn to substantially the same subject matter as claims 4-7 and 9-17 discussed above.

11. Claims 8, 34, and 50 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese et al (U.S. Patent 6,006,218) in view of Konig et al (U.S. Patent 6,981,040) further in view of Gabriel et al (U.S. Patent 6,584,468).

As to claim 8, Breese and Konig do not expressly teach wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

However, Gabriel teaches wherein URLs include URLs that have not been visited by a user but are related to URLs visited by a user, and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited (col. 7, l. 37 – col. 9, l. 10).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese and Konig with the above teachings, such that the above claimed subject matter is implemented. The motivation would have been to facilitate indexing relevant information, as taught throughout Gabriel (e.g., Abstract, col. 7, ll. 37-40, col. 2, ll. 34-46).

Claims 34 and 50 are drawn to substantially the same subject matter as claim 8, discussed above.

12. Claims 25-26 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese et al (U.S. Patent 6,006,218) in view of Konig et al (U.S. Patent 6,981,040) further in view of Dumais et al (US 2004/0267700).

As to claims 25-26, Breese as applied above further teaches wherein the ranked set of documents comprises a personalized set of documents ordered by personalized scores generated by combining the document's generic and profile scores

(col. 7, ll. 33-36, fig. 2C). Furthermore, Breese teaches a set of documents ordered by their generic scores (see above).

Breese and Konig do not expressly teach the ranked set of documents comprising the above two sets of documents, and interleaving the two sets to form the ranked set of documents.

However, Dumais teaches interleaving results from a personal search engine and other search results for presenting to the user (para. 0029).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese and Konig with the above, such that the ranked set of documents comprises the above two sets of documents, and the two sets are interleaved to form the ranked set of documents. The motivation would have been to create a personal browsing system to be a portal to all of a user's content, including personal information as well as more general resources, as taught by Dumais (para. 0029).

Conclusion

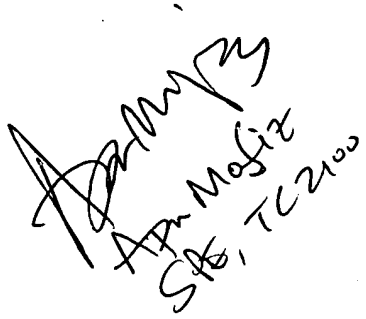
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Charles E. Lu whose telephone number is (571) 272-8594. The examiner can normally be reached on 8:30 - 5:00; M-F.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Apu Mofiz can be reached at (571) 272-4080. 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).

CL
Assistant Examiner
AU 2161
3/26/2007

/CDL/



INFORMATION DISCLOSURE CITATION PTO-1449				Complete If Known	
				Application Number	10/676,711
				Filing Date	September 30, 2003
				First Named Inventor	Lawrence
				Art Unit	~~2171~~ 2161
				Examiner Name	~~To be determined~~ C.Li
Sheet	1	of	1	Attorney Docket Number	60963-0014 (formerly 11378-0014-999)

U.S. PATENT DOCUMENTS						
Examiner Initials	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate

FOREIGN PATENT DOCUMENTS							
Examiner Initials	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation Yes	Translation No
						X	

OTHER PRIOR ART - NON PATENT LITERATURE DOCUMENTS	
Examiner Initials	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published
C.Li	Haveliwala, Taher H., "Topic-Sensitive PageRank", *Proceedings of the Eleventh International World Wide Web Conference*, Honolulu, Hawaii, May 2002. 11 pp
	Jeh, Glen, et al., "Scaling Personalized Web Search", Stanford University Technical Report, 2002. 24 pp
	Pretschner, Alexander, et al., "Ontology Based Personalized Search", *Proc. 11th IEEE Int. Conf. on Tools with Artificial Intelligence*, Chicago, Nov. 1999, pgs. 391-398.

Examiner Signature		Date Considered	3/22/07

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

INFORMATION DISCLOSURE CITATION

PTO-1449

Sheet	1	of	1

Complete If Known	
Application Number	10/676,711
Filing Date	September 30, 2003
First Named Inventor	Stephen R. Lawrence
Art Unit	~~2171~~ 2161
Examiner Name	~~To be determined~~ C.Lu
Attorney Docket Number	60963-0014-US

U.S. PATENT DOCUMENTS

Examiner Initials	Cite No.	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate
C.Lu		6,285,999	09-04-2001	Page	707	5	
		6,327,590	12-04-2001	Chidlovskii et al	707	5	
		2002/0024532 A1	02-28-2002	Fables et al.	345	700	
		2002/0198882 A2	12-26-2002	Linden et al.	707	10	
		2003/0233345 A1	12-18-2003	Perisic et al.	707	3	

FOREIGN PATENT DOCUMENTS

Examiner Initials	Cite No.	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation Yes	Translation No
		EP 1050830A2	11-08-2000	Xerox Corp.				
		WO 03/107127A2	12-24-2003	Entopia, Inc.				

OTHER NON PATENT LITERATURE DOCUMENTS

Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published
		Sergey Brin et al., "The Anatomy of a Large-scale Hypertextual Web Search Engine", *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, April 1998, pgs. 107-117.
		Junghoo Cho et al., "Efficient crawling through URL ordering", *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, April 1998, pgs. 161-171.

Examiner Signature		Date Considered	3/22/07

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

INFORMATION DISCLOSURE CITATION PTO-1449				*Complete If Known*	
				Application Number	10/676,711
				Filing Date	September 30, 2003
				First Named Inventor	Stephen R. Lawrence
				Art Unit	~~2171~~ 2161
				Examiner Name	~~To be determined~~ C. Lu
Sheet	1	of		Attorney Docket Number	60963-0014-US

U.S. PATENT DOCUMENTS							
Examiner Initials	Cite No.	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate
CyLu		6,912,505 B2	06/28/2005	Linden et al.	705	14	

FOREIGN PATENT DOCUMENTS								
Examiner Initials	Cite No.	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation	
							Yes	No

OTHER NON PATENT LITERATURE DOCUMENTS		
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published

Examiner Signature		Date Considered	3/22/07

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

INFORMATION DISCLOSURE CITATION PTO-1449			Complete If Known	
			Application Number	10/676,711
			Filing Date	September 20, 2003
			First Named Inventor	Stephen R. Lawrence
			Art Unit	~~2171~~ 2161
			Examiner Name	~~To be determined~~ C. Lu
Sheet 1	of	1	Attorney Docket Number	60963-0014-US

U.S. PATENT DOCUMENTS							
Examiner Initials	Cite No.	Document Number Number - Kind Code[1]	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Filing Date if Appropriate
CL		5,724,567	03/03/1998	Rose et al.	395	602	
		5,754,939	05/19/1998	Herz et al.	455	4.200	
		6,385,619 B1	05/07/2002	Eichstaedt et al.	707	1	
		2002/044571 A1	03/04/2004	Bronniman	705	14	

FOREIGN PATENT DOCUMENTS								
Examiner Initials	Cite No.	Foreign Patent Document Country Code[2] - Number[3] - Kind Code[4] *(if known)*	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Class	Subclass	Translation Yes	Translation No

OTHER NON PATENT LITERATURE DOCUMENTS		
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published
		International Search Report for International Application No. PCT/US2005/025081, mailed December 2, 2005

Examiner Signature		Date Considered	3/22/07

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [2] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [5] Applicant is to place a check mark here if English language Translation is attached.

Burden Hour Statement: This form is estimated to take 2.0 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Notice of References Cited	Application/Control No. 10/676,711	Applicant(s)/Patent Under Reexamination LAWRENCE, STEPHEN R.	
	Examiner Charles E. Lu	Art Unit 2161	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-6,006,218	12-1999	Breese et al.	707/3
*	B	US-6,981,040	12-2005	Konig et al.	709/224
*	C	US-6,584,468	06-2003	Gabriel et al.	707/10
*	D	US-5,848,396	12-1998	Gerace, Thomas A.	705/10
*	E	US-2004/0267700	12-2004	Dumais et al.	707/002
	F	US-			
	G	US-			
	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
10/676,711	LAWRENCE, STEPHEN R.
Examiner	**Art Unit**
Charles E. Lu	2161

√	Rejected	–	(Through numeral) Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

Claim		Date
Final	Original	01/23/07
	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	01/23/07
	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	
10/676,711	LAWRENCE, STEPHEN R.	
Examiner	**Art Unit**	
Charles E. Lu	2161	

SEARCHED			
Class	Subclass	Date	Examiner

INTERFERENCE SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	DATE	EXMR
Consulted C. Dune Ly (707)	3/22/2007	CL
EAST Search See Attached	3/22/2007	CL

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. 8147

SERIAL NUMBER	FILING OR 371(c) DATE / RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
10/676,711	09/30/2003 RULE	707	2161	60963-0014-US

APPLICANTS

Stephen R. Lawrence, Mountain View, CA;

** CONTINUING DATA ********** None Verified C.L.

** FOREIGN APPLICATIONS ********** None C.L.

IF REQUIRED, FOREIGN FILING LICENSE GRANTED
** 12/19/2003

Foreign Priority claimed	35 USC 119 (a-d) conditions met	Verified and Acknowledged	STATE OR COUNTRY	SHEETS DRAWING	TOTAL CLAIMS	INDEPENDENT CLAIMS
☐ yes ☒ no	☐ yes ☒ no ☐ Met after Allowance	Examiner's Signature 3/22/07 C.L. Initials	CA	12	58	4

ADDRESS
24341

TITLE

Personalization of web search

FILING FEE RECEIVED	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
1518		

Electronically filed August 2, 2007

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Stephen R. Lawrence	Confirmation No.:	8147
Serial No.:	10/676,711	Art Unit:	2161
Filed:	September 30, 2003	Examiner:	Lu, Charles Edward
For:	*Personalization of Web Search Results Using Term, Category, and Link-Based User Profiles (As Amended)*	Attorney Docket No.:	60963-0014-US
		Date:	August 2, 2007

AMENDMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

The enclosed Amendment is in response to the Office Action dated April 2, 2007 for the above identified patent application.

Petition for Extension of Time under 37 CFR 1.136. It is respectfully requested that the time for response to the Office Action dated April 2, 2007 be extended for a period of one (1) month from July 2, 2007 to August 2, 2007.

The Commissioner is hereby authorized to charge any required fee(s) to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no. 60963-0014-US). A copy of this sheet is enclosed for such purpose.

IN THE SPECIFICATION:

Please revise the specification in the Published Patent Application No. 2005/0071328 as follows:

Please amend the title to read as follows:

PERSONALIZATION OF WEB SEARCH RESULTS USING TERM, CATEGORY, AND LINK-BASED USER PROFILES

Please amend the following paragraphs of the specification:

[0044] A user's specific interests may be associated with multiple categories at various levels, each of which may have a weight indicating the degree of relevance between the category and the user's interest. In one embodiment, a category-based profile may be implemented using a Hash table data structure as shown in Fig. 4B. A category-based profile table 450 includes a table [[455]] that comprises a plurality of records [[460]] 341-1, 342-2, …, 342-n, each record including a USER_ID and a pointer pointing to another data structure, such as a table, e.g., one of the tables shown on the right side of Fig. 4B.~~table 460-1.~~ This table ~~Table 460-1~~ may include two columns, a CATEGORY_ID column [[470]] and a WEIGHT column [[480]]. The CATEGORY_ID column [[470]] contains a category's identification number as shown in Fig. 4A, suggesting that this category is relevant to the user's interests and the value in the WEIGHT column [[480]] indicates the degree of relevance of the category to the user's interests.

[0070] In some embodiments that employ a server-side implementation, the user's identification number is embedded in the search query. Based on the user's identification number, the user profile server 108 identifies the user's user profile at step 925. Starting from step 930, the user profile server 108 or the search engine 104 analyzes each document identified at step 920 to determine its relevance to the user's profile, creates a profile score for the identified document at step 935 and then assigns the document a personalized score that is a function of the document's generic and profile scores at step 940. At step 942, the user profile server 108 or the search engine 104 checks whether this the last one in the list of identified documents. If no, the system processes the next document in the list. Otherwise,

the list of documents are re-ordered according to their personalized scores at step 945 and then sent to the corresponding client from which the user submitted the search query.

IN THE CLAIMS:

Rewrite the pending claims and add new claims as follows:

1. (Currently Amended) A method of personalizing search results of a search engine, comprising:

accessing a first user profile for a first user based on information about the first user, the first user information including information derived from a first set of documents, the first set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the first user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the first user;

receiving a search query from the first user;

identifying a set of generic search result documents that match the search query;

assigning a generic score to each document of at least a plurality of the search result documents;

assigning a first personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the first user profile; [[and]]

ranking the set of search result documents into a first order according to their first personalized scores;

providing the ranked set of search result documents in the first order to the first user; and

receiving the search query from a second user that is different from the first user;

accessing a second user profile associated with the second user based on information about the second user, the second user information including information derived from a second set of documents, the second set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the second user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the second user;

assigning a second personalized score to each document of the plurality of identified search result documents in accordance with the generic score assigned to the document and the second user profile;

ranking the set of search result documents into a second order according to their second personalized scores; and

providing the ranked set of search result documents in the second order to the second user.

2. (Currently Amended) The method of claim 1, wherein the first set of documents ~~include~~ includes a plurality of documents that have been identified by search results from the search engine and that have been viewed by the first user.

3. (Currently Amended) The method of claim 1, wherein the first set of documents ~~include~~ includes a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the first user.

4. (Currently Amended) The method of claim 1, including updating the first user profile by:

updating a term-based profile of the first user profile by identifying a set of terms from a document in the first set of documents, and adding information about the identified set of terms to the term-based profile; and

updating a category-based profile of the first user profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

5. (Currently Amended) The method of claim 4, including updating the first user profile by:

updating a link-based profile of the first user profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

6. (Original) The method of claim 5, wherein the link-based profile includes information about URLs or portions of URLs.

7. (Currently Amended) The method of claim 6, wherein the link-based profile of the first user profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the first user visits the URL, time the first user has spent viewing a document associated with the URL and quantity of the first user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the first user's visits to the host.

8. (Currently Amended) The method of claim 7, wherein the URLs further include URLs that have not been visited by the first user, but are related to the URLs that have been visited by the first user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

9. (Original) The method of claim 4, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

10. (Currently Amended) The method of claim 9, wherein the weight is a weight associated with occurrences of the term in the first set of documents.

11. (Currently Amended) The method of claim 9, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said first set of documents.

12. (Original) The method of claim 4, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

13. (Original) The method of claim 12, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

14. (Currently Amended) The method of claim 1, including updating the first user profile by:

updating a category-based profile of the first user profile by classifying a document in the first set of documents into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

15. (Currently Amended) The method of claim 1, wherein the first and second user ~~profile is~~ profiles are stored on a server of the search engine.

16. (Currently Amended) The method of claim 1, wherein the first user profile is stored on a first client associated with the first user and the second user profile is stored on a second client associated with the second user.

17. (Currently Amended) The method of claim 1, wherein ~~each of the user is a~~ the first user profile corresponds to a respective group of users.

18. (Currently Amended) A method of personalizing search results of a search engine, comprising:

creating a plurality of user profiles for a plurality of users, each user profile including at least a user's identification number and information derived from documents visited by the user;

receiving a search query from a user of the plurality of users, the search query including at least one query term and the user's identification number;

retrieving a user profile that matches the user's identification number;

generating a personalized query strategy from the search query and the user profile;

selecting a personalized set of documents from the Internet according to the personalized query strategy, each document having a generic ranking score based at least in part on the relevance of the document to the search query;

assigning to each document in the set a ~~generic~~ personalized ranking score ~~that characterizes the relevance of the document to the at least one query term~~ based at least in part on the user profile and the document's generic ranking score;

~~retrieving the user's user profile and assigning to each of the set of documents a profile score based on the user profile; and~~

ranking the set of documents according to their generic and ~~profile~~ personalized ranking scores; and

providing the ranked set of search result documents to the user.

19. (Original) The method of claim 18, wherein the step of creating a user's user profile further comprises:

creating a term-based profile by extracting a set of terms from documents visited by the user and associating a weight with each extracted term; and

creating a category-based profile by determining a plurality of categories associated with documents visited by the user and associating a weight with each determined category.

20. (Original) The method of claim 18, wherein the step of creating a user's user profile further comprises:

creating a link-based profile by analyzing links in documents visited by the user and associating weights with the link.

21. (Original) The method of claim 18, wherein the user profile for a particular user includes demographic and geographic information provided by the user.

22. (Original) The method of claim 18, wherein the documents visited by the user from which information is derived for use in a particular user's user profile are selected based on the user's activities when visiting the documents.

23. (Currently Amended) The method of claim 18, including storing the plurality of user profiles on a server of the search engine; and

the retrieving including identifying the user's user profile based on [[a]] the user identification number associated with both the user and the user's user profile.

24. (Original) The method of claim 18, including storing the plurality of user profiles on client computers associated with the plurality of users.

25. (Currently amended) The method of claim 18, wherein the ranked set of documents comprises two subsets of documents, one subset of documents ordered by their generic ranking scores and the other subset of documents ordered by personalized ~~scored~~ ranking scores ~~generated by combining the document's generic and profile scores~~.

26. (Currently Amended) The method of claim 25, including interleaving or intermixing the two subsets of documents to form the ranked set of documents.

27. (Currently Amended) A search engine system, comprising:

one or more central processing units for executing programs;

an interface for receiving ~~event messages~~ information; and

a search engine module executable by the one or more central processing units, the module comprising:

instructions for accessing a user profile for a user, the user profile based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set

consisting of documents identified by search results from the search engine system, documents accessed by the user, documents linked to the documents identified by search results from the search engine system, and documents linked to the documents accessed by the user;

instructions for receiving a search query from [[the]] a first user and the same search query from a second user that is different from the first user;

instructions for identifying a set of generic search result documents that match the search query;

instructions for assigning a generic score to each document of at least a plurality of the search result documents;

instructions for assigning [[a]] first and second personalized [[score]] scores to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the first user's user profile and the second user's user profile, respectively; [[and]]

instructions for ranking at least the plurality of the search result documents into first and second orders, respectively, according to their first and second personalized scores; and

instructions for providing the ranked set of search result documents in the first order to the first user and the ranked set of search result documents in the second order to the second user.

28. (Original) The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

29. (Original) The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

30. (Original) The system of claim 27, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

31. (Original) The system of claim 30, further including:

instructions for updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

32. (Original) The system of claim 31, wherein the link-based profile includes information about URLs or portions of URLs.

33. (Original) The system of claim 32, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

34. (Original) The system of claim 33, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

35. (Original) The system of claim 30, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

36. (Original) The system of claim 35, wherein the weight is a weight associated with occurrences of the term in the set of documents.

37. (Original) The system of claim 35, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

38. (Original) The system of claim 30, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

39. (Original) The system of claim 38, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

40. (Original) The system of claim 27, wherein the user profile is stored on a server of the search engine.

41. (Original) The system of claim 27, wherein the user profile is stored on a client associated with the user.

42. (Currently Amended) The system of claim 27, wherein the first user's user [[is]] profile corresponds to a group of users.

43. (Currently Amended) A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for accessing a user profile for a user based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from a search engine, documents accessed by the user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user;

instructions for receiving a search query from [[the]] a first user and the same search query from a second user that is different from the first user;

instructions for identifying a set of generic search result documents that match the search query;

instructions for assigning a generic score to each document of at least a plurality of the search result documents;

instructions for assigning [[a]] first and second personalized [[score]] scores to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the first user's user profile and the second user's user profile, respectively; [[and]]

instructions for ranking at least the plurality of the search result documents into first and second orders, respectively, according to their first and second personalized scores; and

instructions for providing the ranked set of search result documents in the first order to the first user and the ranked set of search result documents in the second order to the second user.

44. (Original) The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

45. (Original) The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

46. (Original) The computer program product of claim 43, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

47. (Original) The computer program product of claim 46, further including:

instructions for updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

48. (Original) The computer program product of claim 46, wherein the link-based profile includes information about URLs or portions of URLs.

49. (Original) The computer program product of claim 48, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

50. (Original) The computer program product of claim 49, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

51. (Original) The computer program product of claim 46, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

52. (Original) The computer program product of claim 48, wherein the weight is a weight associated with occurrences of the term in the set of documents.

53. (Original) The computer program product of claim 48, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

54. (Original) The computer program product of claim 46, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

55. (Original) The computer program product of claim 54, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

56. (Original) The computer program product of claim 43, wherein the user profile is stored on a server of the search engine.

57. (Original) The computer program product of claim 43, wherein the user profile is stored on a client associated with the user.

58. (Currently Amended) The computer program product of claim 43, wherein the first user's user [[is]] profile corresponds to a group of users.

REMARKS

This amendment responds to the office action mailed April 2, 2007. In the office action the Examiner:

- objected to claim 25 for containing an informality;
- objected to the specification, and in particular paragraphs 0047 and 0061;
- objected to the drawings, and in particular Figure 4B and Figure 9A;
- rejected claims 1-58 under 35 U.S.C. §101 as being directed to non-statutory subject matter;
- rejected claims 1, 27 and 43 under 35 U.S.C. §102(b) as being anticipated by Breese *et al.* (US 6,006,218);
- rejected claims 2-3, 18, 22, 28-29 and 44-45 under 35 U.S.C. §103(a) as being unpatentable over Breese *et al* (US 6,006,218)
- rejected claim 21 under 35 U.S.C. §103(a) as being unpatentable over Breese *et al.* (US 6,006,218) in view of Gerace (US 5,848,396);
- rejected claims 4-7, 9-17, 19-20, 23-24, 30-33, 35-42, 46-49, and 51-58 under 35 U.S.C. §103(a) as being unpatentable over Breese *et al.* (US 6,006,218) in view of Konig *et al.* (US 6,981,040);
- rejected claims 8, 34, and 50 under 35 U.S.C. §103(a) as being unpatentable over Breese *et al.* (US 6,006,218) in view of Konig *et al.* (US 6,981,040) further in view of Gabriel *et al.* (US 6,584,468); and
- rejected claims 25-26 under 35 U.S.C. §103(a) as being unpatentable over Breese *et al* (US 6,006,218) in view of Konig *et al.* (US 6,981,040) further in view of Dumais *et al.* (US 2004/0267700).

After entry of this amendment, the pending claims are: claims 1-58.

Claim Amendments

Independent claims 1, 18, 27, and 43 (as well as some of their associated dependent claims) have been amended. Support for these amendments may be found, for example, in Figures 9A and 9B and their related text in the specification. No new matter has been added.

Objections to the Drawings and Specification

Applicant notes that the Examiner's objections to the specification are drawn to paragraphs numbered according to the published application (US 2005/0071328). The

paragraph numbering in the published application differs from that of the specification as filed.

Applicant has amended paragraph 0044 of the application as filed to correctly refer to the elements of Figure 4B. This amendment is believed to address the Examiner's objection to Figure 4B as well, since all numbered elements in Figure 4B are now recited in the specification.

With regard to the Examiner's objection to the word "value" in the pseudo code in paragraph 0061, Applicant respectfully notes that this word was correctly spelled in paragraph 0058 of the originally submitted specification.

Paragraph 0070 of the application as filed has been amended to recite the numbered element 945 shown in Figure 9A. Applicant believes this amendment overcomes the Examiner's objection to Figure 9A.

The Examiner has objected to the title of the application as not being sufficiently descriptive. Applicant in this Amendment has requested that the title of the application be amended to that suggested by the Examiner. The Examiner's approval of these amendments and corresponding withdrawal of objections to the figures and specification is respectfully requested.

Objection to the Claims

The Examiner has objected to claim 25, specifically the occurrence of "scored" which should be changed to "scores." This change has been made and it is respectfully requested that the objections be withdrawn.

Claim Rejections under 35 U.S.C. §101

The Examiner rejected claims 1-58 under 35 U.S.C. §101 as being directed to non-statutory subject matter. Independent claim 18 has been amended to further recite "providing the ranked set of search result documents to the user," and a similar element has been added to claim 1, and thus the claimed method generates a useful, concrete and tangible result. Independent claims 27 and 43 have been amended to recite similar limitations.

The Examiner also rejected claims 43-58 as being software *per se* citing MPEP §2106. The Applicant submits the following excerpt from MPEP §2106.01:

> In contrast, **a claimed computer-readable medium encoded with a computer program** is a computer element which defines structural and functional interrelationships between the computer program and the rest of the computer which permit the computer program's functionality to be realized, and **is thus statutory.** (*emphasis added*)

The preamble of claim 43 reads as follows:

> A computer program product for use in conjunction with a computer system, the computer program product comprising **a computer readable storage medium and a computer program mechanism embedded therein** … (*emphasis added*)

The highlighted portion of the preamble clearly meets MPEP's requirement of statutory subject matter. Thus, Applicant submits that the Examiner withdraw the rejection under 35 U.S.C. 101.

Claimed Rejections under 35 U.S.C. §102(b)

The Examiner rejected claims 1, 27 and 43 as being anticipated by Breese. Applicant respectfully disagrees and traverses.

Claim 1, as amended, reads as follows:

> "A method of personalizing search results of a search engine, comprising:
> …
> identifying a set of **generic** search result documents that match the search query [submitted by a first user];
> …
> receiving **the search query from a second user that is different from the first user;**
> …
> assigning a second personalized score to each document of **the plurality of identified search result documents** in accordance with the generic score assigned to the document and **the second user profile**;
> ranking the set of search result documents into **a second order** according to their second personalized scores; and
> providing the ranked set of search result documents **in the second order to the second user**." (*emphasis added*)

Breese teaches methods and apparatus for retrieving information and/or processing retrieved information as a function of a user's estimated knowledge. To retrieve information database entries relevant to a particular user, the search engine 130 uses the **user-specific** information stored in the user database 104:

> In step 228, the information obtained from the user input device 170, **the user database 104** and the information database 106 is supplied to **the search engine 130** and the search result post-processor 231 as illustrated in FIG. 2A by lines 105 and 107. Col. 8, line 57-61 (*emphasis added*)

Breese does not teach or suggest the feature of identifying (e.g., in a database) a set of **generic** search result documents in response to a search query and then returning the same set of generic search result documents in **different orders** to different users who submit the **same search query** in accordance with their respective user profiles.

Claims 27 and 43 have been amended in substantially the same manner. Therefore, for at least the reasons above, claims 1, 27, and 43 are not anticipated by Breese.

Claimed Rejections under 35 U.S.C. §103(a)

The Examiner rejected other pending claims over Breese in view of other references under 35 U.S.C. §103(a). Applicant respectfully disagrees and traverses.

Claims 2-17, 28-42, and 44-58

Claims 2-3, 28-29, and 44-45 are dependent from claims 1, 27, and 43, respectively, and are therefore patentable over Breese for at least the reasons discussed above.

Applicant further notes that none of the other references discloses all the claimed features as discussed in the previous section.

Konig teaches a method of locating a set of search results from "the set containing all documents D and **user documents obtained during personal crawling**" (col. 27, lines 44-45). According to Konig, different users may have different sets of user documents, each set of user documents being of interest to that particular user. In response to the same search query from two different users, Konig requires two distinct search operations against two different corpuses of documents. The two search operations could produce two different sets of search results because the two users may not have the same set of user documents. Thus, it is impossible for Konig to identify a set of **generic** search results and re-order it in accordance with different users' user profile as recited in claims 1, 27, and 43.

Gabriel was cited by the Examiner to reject claims 8, 34, and 50. It does not disclose the features as recited in claims 1, 27, and 50. Nor does the Examiner claim so.

Thus, dependent claims 4-17, 30-42, and 46-58 are all patentable over the cited references for at least the reasons discussed above.

Claims 18-26

Independent claim 18, as amended, is directed to a personalized search method that includes the limitations "generating **a personalized query strategy** from the search query and the user profile" and "selecting a **personalized** set of **documents** from the Internet according to the personalized query strategy."

Breese does not teach or suggest a method of generating a personalized query strategy. Nor does Breese disclose selecting a personalized set of documents from the Internet using a personalized query strategy.

Further, Applicant is unable to find evidence that any of the other references discloses all the features recited in claim 18. Gerace discloses a method of creating user profiles, which was cited to reject claim 21. Gerace does not teach or suggest a method of generating a personalized search strategy. Similarly, Dumais teaches relates to systems and methods providing content-access-based information retrieval, which was cited to reject claims 25-26. According to Konig (Figure 19), a set of search results is first located from the set containing all documents D and user documents obtained during personal crawling (192) and then personalized using a user's User Model (194). There is no personalized search strategy involved in the step 192.

In sum, claims 18-26 are patentable over the references cited by the Examiner for at least the reasons mentioned above.

In light of the above amendments and remarks, the Applicant respectfully requests that the Examiner reconsider this application with a view towards allowance. The Examiner is invited to call the undersigned attorney at (650) 843-4000, if a telephone call could help resolve any remaining items.

Respectfully submitted,

Date: August 2, 2007

/ Gary S. Williams / 31,066

Gary S. Williams (Reg. No.)
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, CA 94306
(650) 843-4000

Electronic Patent Application Fee Transmittal

Application Number:	10676711
Filing Date:	30-Sep-2003
Title of Invention:	Personalization of web search
First Named Inventor/Applicant Name:	Stephen R. Lawrence
Filer:	Gary Scott Williams/Beverly Gemello
Attorney Docket Number:	60963-0014-US

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:				
Extension - 1 month with $0 paid	1251	1	120	120

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
			Total in USD ($)	**120**

Electronic Acknowledgement Receipt

EFS ID:	2043493
Application Number:	10676711
International Application Number:	
Confirmation Number:	8147
Title of Invention:	Personalization of web search
First Named Inventor/Applicant Name:	Stephen R. Lawrence
Customer Number:	24341
Filer:	Gary Scott Williams/Beverly Gemello
Filer Authorized By:	Gary Scott Williams
Attorney Docket Number:	60963-0014-US
Receipt Date:	02-AUG-2007
Filing Date:	30-SEP-2003
Time Stamp:	20:10:45
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment was successfully received in RAM	$120
RAM confirmation Number	3134
Deposit Account	500310

File Listing:

Document Number	Document Description	File Name	File Size(Bytes) /Message Digest	Multi Part /.zip	Pages (if appl.)

1		60963-0014_Amendment.pdf	169716 3fe292475547e991cc1deab44fede5f30422935f	yes	18

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Amendment - After Non-Final Rejection	1	1
Specification	2	3
Claims	4	13
Applicant Arguments/Remarks Made in an Amendment	14	18

Warnings:

Information:

2	Fee Worksheet (PTO-06)	fee-info.pdf	8141 73f39c15774c2c9547bba980c07778375cd89fc8	no	2

Warnings:

Information:

Total Files Size (in bytes):	177857

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.

PATENT APPLICATION FEE DETERMINATION RECORD
Effective January 1, 2003

Application or Docket Number: 10/676,711

CLAIMS AS FILED - PART I

	(Column 1)	(Column 2)
TOTAL CLAIMS	53	
FOR	NUMBER FILED	NUMBER EXTRA
TOTAL CHARGEABLE CLAIMS	53 minus 20=	* 33
INDEPENDENT CLAIMS	4 minus 3 =	* 1
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	375.00	OR	BASIC FEE	750.00
X$ 9=		OR	X$18=	594.
X42=		OR	X84=	84
+140=		OR	+280=	
TOTAL		OR	TOTAL	1518

CLAIMS AS AMENDED - PART II

8-2-07

AMENDMENT A

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA
Total	* 58	Minus	** 58	= 0
Independent	* 4	Minus	*** 4	= 0
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM				☐

SMALL ENTITY		OR	OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE		RATE	ADDITIONAL FEE
X$ 9=		OR	X$18=	0
X42=		OR	X84=	0
+140=		OR	+280=	0
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	0

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				☐

SMALL ENTITY		OR	OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE		RATE	ADDITIONAL FEE
X$ 9=		OR	X$18=	
X42=		OR	X84=	
+140=		OR	+280=	
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				☐

SMALL ENTITY		OR	OTHER THAN SMALL ENTITY	
RATE	ADDITIONAL FEE		RATE	ADDITIONAL FEE
X$ 9=		OR	X$18=	
X42=		OR	X84=	
+140=		OR	+280=	
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.

EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	8914	707/3.ccls.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/10/09 11:20
L2	1597	l1 and profil$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/10/09 11:21
L3	181	l2 and term with profile	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/10/09 11:21
L4	146	l3 and previous$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/10/09 11:21

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.
10/676,711	09/30/2003	Stephen R. Lawrence	60963-0014-US	8147

24341 7590 10/11/2007

MORGAN, LEWIS & BOCKIUS, LLP.
2 PALO ALTO SQUARE
3000 EL CAMINO REAL
PALO ALTO, CA 94306

EXAMINER	
LU, CHARLES EDWARD	

ART UNIT	PAPER NUMBER
2161	

MAIL DATE	DELIVERY MODE
10/11/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/676,711	LAWRENCE, STEPHEN R.	
	Examiner	Art Unit	
	Charles E. Lu	2161	

-- 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) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, 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 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 *02 August 2007*.

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) *1-58* 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) *1-58* 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 _____ 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/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This Action is in response to the Amendment dated 8/2/2007. Claims 1-58 are pending and rejected.

Response to Arguments/Response to Amendments

2. Amendments to the specification and claims were noted. The objection to the title, drawings, specification, and claims are withdrawn.

3. The 35 U.S.C. 101 rejection of the claims is withdrawn.

4. Applicant's arguments regarding the prior art rejections were fully considered but were not persuasive. Applicants argue the claims as amended.

Applicant argues on p. 17 of the Amendment that Breese does not teach or suggest "identifying a set of generic search results in response to a search query and then returning the same set of generic search results in different orders to different users who submit the same search query in accordance with their respective user profiles." The examiner respectfully disagrees.

Breese is drawn to performing a "conventional search operation" and then performing a "search result post processing" (col. 6, ll. 60-65) on the search results. The results from the "conventional search operation" (fig. 2A, #131) meets the claimed limitation "generic search results in response to a search query" because the conventional search results are not yet "personalized" according to the user. Breese's "post processing" is the step that "personalizes" the order of the set of "generic" results

based on user information to produce a personalized search result (e.g., fig. 2C, #231, col. 6, ll. 65-67, col. 7, ll. 30-35, col. 9, ll. 1-4, col. 12, ll. 32-42). The added limitation of "different users...same query" does not overcome the prior art because Breese deals with multiple users (fig. 5), each with his own user profile (#224, #500). Thus, when a different user enters the same search query, Breese will still process a generic set of results as explained above, and then post-process using that user's profile (#234) to re-rank the search results. This interpretation will be used to apply Breese to the amended claims.

Applicant argues on p. 17 that neither Breese nor the other cited prior art teach the above features. The examiner respectfully disagrees because Breese teaches the features.

Applicant argues on p. 18 that neither Breese nor the other cited prior art teach or suggest, "generating a personalized query strategy," and "selecting a personalized set of documents according to the personalized query strategy." The examiner respectfully disagrees for the same reasons as explained above. The personalized post-processing of Breese is a personalized query strategy that selects a personalized set of documents.

The new grounds of rejection presented below are necessitated by Applicant's amendment.

Claim Rejections - 35 USC § 102

The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

5. Claims 1, 16, 27, 41, 43, and 57 are rejected under 35 U.S.C. 102(b) as being anticipated by Breese et al (U.S. Patent 6,006,218), hereafter "Breese."

As to claim 1, Breese teaches the following claimed subject matter:

A method of personalizing search results of a search engine, comprising:

accessing a first user profile for a first user based on information about the first user (fig. 2B, #224, fig. 5, #500, col. 5, ll. 20-45),

The first user information including information derived from a first set of documents (col. 5, ll. 20-45),

The first set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the first user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the first user (col. 5, ll. 20-45);

Receiving a search query from the first user (col. 6, ll. 60-65);

Identifying a set of generic search result documents that match the search query (fig. 2C, #230-231);

Assigning a generic score to each document of at least a plurality of the search result documents (col. 7, ll. 18-45);

Assigning a first personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the first user profile (col. 7, ll. 18-45, details on col. 8-17);

Ranking the set of search result documents into a first order according to their first personalized scores (col. 7, ll. 18-45, details on col. 8-17).

Providing the ranked set of search result documents into a first order to the first user (e.g., fig. 2C, #236).

As to "receiving a search query from a second user that is different from the first user...accessing a second user profile...assigning a second personalized score...ranking the set of search results...and providing the ranked set of search results," Breese teaches the claimed subject matter because Breese deals with multiple users (fig. 5, #500), each with his own user profile (e.g., #224, #500). Thus, when a different user enters the same search query, Breese will still process a generic set of results as explained above, and then post-process using that user's profile (fig. 2C, #234) to re-rank the search results. See above.

As to claim 16, Breese further teaches wherein the first user profile is stored on a first client associated with the first user and the second user profile is stored on a second client associated with the second user (col. 4, l. 62, col. 5, ll. 1-2).

Claims 27, 41, 43 and 57 are rejected on the same basis as the above claims.

Claim Rejections - 35 USC § 103

The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

6. Claims 2-3, 18, 22, 24, 28-29 and 44-45 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese.

As to claims 2 and 3, Breese does not expressly teach wherein the first set of documents includes a plurality of documents that have been identified by search results from the search engine and that have/have not been viewed by the first user.

However, Breese teaches that the user information includes previous search information (col. 5, ll. 30-33, col. 16, l. 40) and that the search information may include information on the entries that were presented to the user as a result of the search (i.e. search results). Furthermore, Breese states, "it may be assumed that the user is aware of these entries, or at least the highest ranked entries (col. 16, ll. 34-50)." The user information includes information on previous Internet site access operations (col. 5, ll. 30-35). Thus, Breese suggests that the user may have actually viewed the information because the information was presented to the user.

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese with the above, such that "wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have/have not been viewed by the user" is implemented. The motivation would have been to enhance the effectiveness of the retrieval result adjustor, because data regarding actual document views would be used.

As to claim 18, Breese teaches the following claimed subject matter:

A method of personalizing search results of a search engine, comprising: creating a plurality of user profiles for a plurality of users, each user profile including at least a user's identification number and information derived from documents visited by the user (col. 5, ll. 20-45);

Receiving a search query from a user of the plurality of users, the search query including at least one query term (e.g., col. 8, ll. 62-66).

Retrieving a user profile that matches the user's identification number (e.g., col. 5, ll. 25-30, col. 8, ll. 29-31);

Selecting a personalized set of documents from the Internet, according to the personalized query strategy, each document having a generic ranking score based at least on part on the relevance of the document to the search query, assigning to each document in the set a personalized ranking score based at least in part on the user profile and the document's generic ranking score (discussed above);

Ranking the set of documents according to their generic and personalized ranking scores and providing the ranked set of search result documents to the user (see above).

Breese does not expressly teach the search query including the user's identification number.

However, Breese teaches that a user has a unique identification number for storing user attributes in a user database (col. 5, ll. 20-45), and that information regarding the user and the search to be performed is obtained at the input step 222 (col. 8, ll. 15-20, #224).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese, such that the search query includes the user's identification number in the input step. The motivation would have been to adapt to specific user requirements in setting up the search engine. For example, one may send the identification with the query to facilitate efficient processing.

As to claim 22, Breese as applied above further teaches wherein the documents visited by the user from which information is derived for use in a particular user's user profile is selected based on the user's activities when visiting the documents (e.g., col. 5, ll. 20-45).

As to claim 24, see above for claim 16.

Claims 28-29 and 44-45 are rejected on the same basis as the above claims.

7. Claim 21 is rejected under 35 U.S.C. 103(a) as being unpatentable over Breese, in view of Gerace (U.S. Patent 5,848,396).

As to claim 21, Breese teaches wherein the user profile includes demographic information provided by the user (fig. 5).

Breese does not expressly teach geographic information.

However, Gerace teaches a user profile including both demographic and geographic information (col. 5, l. 63 – col. 6, l. 15).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese, such that geographic information is additionally stored with the user profile. The motivation would have been to store more

information about the user to facilitate better decisions by the information retrieval system.

8. Claim 4-7, 9-15, 17, 19-20, 23, 30-33, 35-40, 42, 46-49, 51-56 and 58 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese in view of Konig et al (U.S. Patent 6,981,040), hereafter Konig.

As to claims 4-5 and 14, Breese does not expressly teach updating the user profile by updating a term-based profile of the first user profile by identifying a set of terms from a document in the first set of documents, and adding information about the identified set of terms to the term-based profile; and updating a category-based profile of the first user profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile; and updating a link-based profile of the first user profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

However, Konig teaches updating a term-based, and category-based profile for a user as claimed (col. fig. 4A, fig. 4C, col. 10, l. 51, col. 12, l. 55). Furthermore, Konig teaches analyzing links and locations in documents for updating a user model (user topic distribution, user site distribution, and see fig. 13, fig. 4, fig. 15A, col. 17, l. 47, col. 15, ll. 8-15, col. 18, ll. 17-25, col. 22, l. 64 – col. 23, l. 10).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese, such that the claimed updating of the term-based, category-based, and link-based profiles is implemented with appropriate weights associated with each item (see e.g., fig. 4). The motivation would have been to

represent user interest in a document or product independently of any specific user information need, as taught by Konig (col. 4, ll. 1-7). This would further enhance search results when combined with Breese.

As to claim 6, Konig as applied above further teaches wherein the link-based profile includes information about URLs or portions of URLs (fig. 4).

As to claim 7, Konig as applied above further teaches or suggests wherein the link-based profile of the first user profile comprises a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the first user visits the URL, time the first user has spent viewing a document associated with the URL and quantity of the first user's scrolling activity at the document; and a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the first user's visits to the host (col. 12, ll. 28-54, col. 23, ll. 1-10).

As to claim 9, Konig as applied above further teaches wherein a term in the term-based profile is an expression comprising at least one word and a weight (fig. 4A).

As to claim 10, Konig as applied above further teaches wherein the weight is a weight associated with occurrences of the term in the first set of documents (fig. 4, col. 10, l. 52 – col. 12, l. 55).

As to claim 11, Konig as applied above further teaches wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said first set of documents (col. 10, l. 52 – col. 11, l. 20).

As to claim 12, Konig as applied above further teaches wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories (fig. 4, 7, 8, col. 15, ll. 7-32).

As to claim 13, Konig as applied above further teaches wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin (e.g., col. 15, ll. 7-15 and see above).

As to claim 15, Breese discloses a first and second user profile and a search engine (fig. 1, 5), but does not expressly teach wherein the user profiles are stored on a server of the search engine.

However, Konig teaches wherein user profiles are stored on a server of the search engine (fig. 1).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese, such that the user profiles are stored on a server of the search engine. The motivation would have been to adapt to the requirements of the user in setting up the search system, or to provide personalized services for simultaneous clients, as taught by Konig (col. 7, ll. 20-25).

As to claim 17, Breese does not expressly teach wherein the first user profile corresponds to a respective a group of users.

However, Konig teaches wherein a user is a group of users (col. 20, ll. 24-28, col. 9, ll. 47-52).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese, such that the first user is a group of users. The motivation would have been to represent the interest level of a group of users in a document independently of any specific information need, as taught by Konig (col. 9, ll. 47-52).

Claims 19-20 are drawn to substantially the same subject matter as claims 4-5, discussed above, in addition to creating, which must happen in Konig in order to store the relevant data (see e.g., fig. 4).

As to claim 23, the "storing" limitation is addressed with respect to claim 15 above. Breese, as applied above, further teaches the retrieving including the user's user profile based on an identification number associated with the user and the user's profile (col. 5, ll. 23-30). Note that Breese must retrieve the data in order to process it.

Claims 30-33, 35-40, 42, 46-49, and 51-56 and 58 are rejected on the same basis as the above claims.

9. Claims 8, 34, and 50 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese, in view of Konig, further in view of Gabriel et al (U.S. Patent 6,584,468), hereafter "Gabriel."

As to claim 8, Breese and Konig do not expressly teach wherein the URLs further include URLs that have not been visited by the first user, but are related to the URLs that have been visited by the first user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

However, Gabriel teaches wherein URLs include URLs that have not been visited by a user but are related to URLs visited by a user, and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited (col. 7, l. 37 – col. 9, l. 10).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese and Konig, such that the above claimed subject matter is implemented. The motivation would have been to facilitate indexing relevant information, as taught throughout Gabriel (e.g., Abstract, col. 7, ll. 37-40, col. 2, ll. 34-46).

Claims 34 and 50 are rejected on the same basis as claim 8, discussed above.

10. Claims 25-26 are rejected under 35 U.S.C. 103(a) as being unpatentable over Breese, in view of Konig, further in view of Dumais et al (US 2004/0267700), hereafter "Dumais."

As to claims 25-26, Breese as applied above further teaches wherein the ranked set of documents comprises a personalized subset of documents ordered by personalized scores and the other subset ordered by the generic ranking scores (col. 7, ll. 33-36, fig. 2C). Furthermore, Breese teaches a set of documents ordered by their generic scores (see above).

Breese and Konig do not expressly teach the ranked set of documents comprising the above two sets of documents, and interleaving the two sets to form the ranked set of documents.

However, Dumais teaches interleaving results from a personal search engine and other search results for presenting to the user (para. 0029).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Breese and Konig with the above, such that the ranked set of documents comprises the above two sets of documents, and the two sets are interleaved to form the ranked set of documents. The motivation would have been to create a personal browsing system to be a portal to all of a user's content, including personal information as well as more general resources, as taught by Dumais (para. 0029).

Conclusion

Applicant's arguments were fully considered but were not persuasive. Furthermore, Applicant's amendment necessitates new grounds of rejection. Accordingly, **THIS ACTION IS MADE FINAL.** See MPEP 706.07(a). 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.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Charles E. Lu whose telephone number is (571) 272-8594. The examiner can normally be reached on 8:30 - 5:00; M-F.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Apu Mofiz can be reached at (571) 272-4080. The fax phone number for the organization where this application or proceeding is assigned 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).

APU MOFIZ
SUPERVISORY PATENT EXAMINER

/CL/
Assistant Examiner
AU 2161
10/9/2007

Index of Claims

Application/Control No.	Applicant(s)/Patent under Reexamination
10/676,711	LAWRENCE, STEPHEN R.
Examiner	**Art Unit**
Charles E. Lu	2161

√	Rejected	–	(Through numeral) Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

Claim		Date									
Final	Original	3/22/07	10/1/07								
	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	3/22/07	10/1/07								
	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
10/676,711	LAWRENCE, STEPHEN R.
Examiner	**Art Unit**
Charles E. Lu	2161

SEARCHED			
Class	Subclass	Date	Examiner
707	3	10/9/2007	CL

INTERFERENCE SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	DATE	EXMR
Consulted Apu Mofiz	10/9/2007	CL
EAST	10/9/2007	CL

PTO/SB/30 (04/2005)
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 displays a valid OMB control number

REQUEST FOR CONTINUED EXAMINATION (RCE) TRANSMITTAL

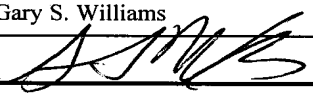
Address to:
Mail Stop RCE
Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Electronically filed	December 12, 2007
Application Number	10/676,711
Filing Date	September 30, 2003
First Named Inventor	Stephen R. Lawrence
Art Unit	2161
Examiner Name	Lu, Charles Edward
Attorney Docket No.	60963-0014-US

This is a Request for Continued Examination (RCE) under 37 C.F.R. § 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.

1. Submission required under 37 C.F.R. § 1.114
   - a. ☐ Previously submitted
     - i. ☐ Consider the amendment(s)/reply under 37 C.F.R. § 1.116 previously filed on _____ (Any unentered amendment(s) referred to above will be entered).
     - ii. ☐ Consider the arguments in the Appeal Brief or Reply Brief previously filed _____
     - iii. ☐ 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 C.F.R. § 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 C.F.R. § 1.17(i) required)
   - b. ☐ Other _____
3. Fees — The RCE fee under 37 C.F.R. § 1.17(e) is required by 37 C.F.R. § 1.114 when the RCE is filed
   - a. ☒ The Director is hereby authorized to charge the following fees, or credit any overpayments, to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no. 60963-0014-US). I have enclosed a duplicate copy of this sheet.
     - i. ☒ RCE fee required under 37 C.F.R. § 1.17(e), estimated to be $ 810 (large entity) or $405 (small entity)
     - ii. ☐ Extension of time fee required under 37 C.F.R. §§ 1.136 and 1.17, estimated to be $ for a month(s) extension, the request for which is being made herewith
     - iii. ☐ Other _____
   - b. ☐ Check in the amount of $ _____ enclosed
   - c. ☐ Payment by credit card (Form PTO-2038 enclosed)

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED

Name (Print/Type)	Gary S. Williams	*Registration No. (Attorney/Agent)*	31,066
Signature		*Date*	December 12, 2007

CERTIFICATE OF MAILING OR TRANSMISSION

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, or by facsimile transmitted to fax no. 1-703- to the U.S. Patent and Trademark Office on the date indicated below.

Name (Print/Type)		*Registration No. (Attorney/Agent)*	
Signature		*Date*	

This collection of information is required by 37 CFR 1.114. This 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 1.11 and 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 needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND Fees and Completed Forms to the following address: Mail Stop RCE, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Electronically filed December 12, 2007

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Stephen R. Lawrence	Confirmation No.:	8147
Serial No.:	10/676,711	Art Unit:	2161
Filed:	September 30, 2003	Examiner:	Lu, Charles Edward
For:	*Personalization of Web Search Results Using Term, Category, and Link-Based User Profiles (As Amended)*	Attorney Docket No.:	60963-0014-US
		Date:	December 12, 2007

RESPONSE TO FINAL OFFICE ACTION

FILED WITH REQUEST FOR CONTINUED EXAMINATION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

The enclosed Amendment is in response to the Final Office Action dated October 11, 2007 for the above identified patent application.

The enclosed Amendment is being submitted with a "Request for Continued Examination" (RCE).

The Commissioner is hereby authorized to charge any required fee(s) to Morgan, Lewis & Bockius LLP Deposit Account No. 50-0310 (order no. 60963-0014-US). A copy of this sheet is enclosed for such purpose.

IN THE CLAIMS:

Rewrite the pending claims as follows:

1. (Currently Amended) A method of personalizing search results of a search engine, comprising:

accessing a first user profile for a first user based on information about the first user, the first user information including information derived from a first set of documents, the first set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the first user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the first user;

receiving a search query from the first user;

identifying a set of generic search result documents that match the search query;

assigning a generic score to each document of at least a plurality of the search result documents;

assigning a first personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the first user profile;

ranking the set of search result documents into a first order according to their first personalized scores;

providing the ranked set of search result documents in the first order to the first user;

updating the first user profile based on a document selected by the first user from the set of search result documents, including analyzing links within the document and adding information derived from the analyzed links to the first user profile; and

receiving the search query from a second user that is different from the first user;

accessing a second user profile associated with the second user based on information about the second user, the second user information including information derived from a second set of documents, the second set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the second user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the second user;

assigning a second personalized score to each document of the plurality of identified search result documents in accordance with the generic score assigned to the document and the second user profile;

ranking the set of search result documents into a second order according to their second personalized scores; and

providing the ranked set of search result documents in the second order to the second user.

2. (Previously presented) The method of claim 1, wherein the first set of documents includes a plurality of documents that have been identified by search results from the search engine and that have been viewed by the first user.

3. (Previously presented) The method of claim 1, wherein the first set of documents includes a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the first user.

4. (Currently Amended) The method of claim 1, including updating the first user profile by:

updating a term-based profile of the first user profile by identifying a set of terms from a document in the first set of documents, and adding information about the identified set of terms to the term-based profile~~; and~~

~~updating a category-based profile of the first user profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile~~.

5. (Canceled)

6. (Currently Amended) The method of claim 1 [[5]], wherein the information derived from the analyzed links that is added to the first user profile is added to a link-based profile and includes information about URLs or portions of URLs.

7. (Previously presented) The method of claim 6, wherein the link-based profile of the first user profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the

first user visits the URL, time the first user has spent viewing a document associated with the URL and quantity of the first user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the first user's visits to the host.

8. (Previously presented) The method of claim 7, wherein the URLs further include URLs that have not been visited by the first user, but are related to the URLs that have been visited by the first user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

9. (Previously presented) The method of claim 4, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

10. (Previously presented) The method of claim 9, wherein the weight is a weight associated with occurrences of the term in the first set of documents.

11. (Previously presented) The method of claim 9, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said first set of documents.

12. (Currently Amended) The method of claim 14 [[4]], wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

13. (Original) The method of claim 12, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

14. (Previously presented) The method of claim 1, including updating the first user profile by:

updating a category-based profile of the first user profile by classifying a document in the first set of documents into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

15. (Previously presented) The method of claim 1, wherein the first and second user profiles are stored on a server of the search engine.

16. (Previously presented) The method of claim 1, wherein the first user profile is stored on a first client associated with the first user and the second user profile is stored on a second client associated with the second user.

17. (Previously presented) The method of claim 1, wherein the first user profile corresponds to a respective group of users.

18. (Currently Amended) A method of personalizing search results of a search engine, comprising:

creating a plurality of user profiles for a plurality of users, each user profile including at least a user's identification number and information derived from documents visited by the user;

receiving a search query from a user of the plurality of users, the search query including at least one query term and the user's identification number;

retrieving a user profile that matches the user's identification number;

generating a personalized query strategy from the search query and the user profile;

selecting a personalized set of documents from the Internet according to the personalized query strategy, each document having a generic ranking score based at least in part on the relevance of the document to the search query;

assigning to each document in the set a personalized ranking score based at least in part on the user profile and the document's generic ranking score;

ranking the set of documents according to their generic and personalized ranking scores; [[and]]

providing the ranked set of search result documents to the user; and

updating the first user profile based on a document selected by the first user from the set of search result documents, including analyzing links within the document and adding information derived from the analyzed links to the first user profile.

19. (Original) The method of claim 18, wherein the step of creating a user's user profile further comprises:

creating a term-based profile by extracting a set of terms from documents visited by the user and associating a weight with each extracted term; and

creating a category-based profile by determining a plurality of categories associated with documents visited by the user and associating a weight with each determined category.

20. (Original) The method of claim 18, wherein the step of creating a user's user profile further comprises:

creating a link-based profile by analyzing links in documents visited by the user and associating weights with the link.

21. (Original) The method of claim 18, wherein the user profile for a particular user includes demographic and geographic information provided by the user.

22. (Original) The method of claim 18, wherein the documents visited by the user from which information is derived for use in a particular user's user profile are selected based on the user's activities when visiting the documents.

23. (Previously presented) The method of claim 18, including storing the plurality of user profiles on a server of the search engine; and

the retrieving including identifying the user's user profile based on the user identification number associated with both the user and the user's user profile.

24. (Original) The method of claim 18, including storing the plurality of user profiles on client computers associated with the plurality of users.

25. (Previously presented) The method of claim 18, wherein the ranked set of documents comprises two subsets of documents, one subset of documents ordered by their generic ranking scores and the other subset of documents ordered by personalized ranking scores.

26. (Previously presented) The method of claim 25, including interleaving or intermixing the two subsets of documents to form the ranked set of documents.

27. (Currently Amended) A search engine system, comprising:

one or more central processing units for executing programs;

an interface for receiving information; and

a search engine module executable by the one or more central processing units, the module comprising:

instructions for accessing a user profile for a user, the user profile based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine system,

documents accessed by the user, documents linked to the documents identified by search results from the search engine system, and documents linked to the documents accessed by the user;

instructions for receiving a search query from a first user and the same search query from a second user that is different from the first user;

instructions for identifying a set of generic search result documents that match the search query;

instructions for assigning a generic score to each document of at least a plurality of the search result documents;

instructions for assigning first and second personalized scores to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the first user's user profile and the second user's user profile, respectively;

instructions for ranking at least the plurality of the search result documents into first and second orders, respectively, according to their first and second personalized scores; [[and]]

instructions for providing the ranked set of search result documents in the first order to the first user and the ranked set of search result documents in the second order to the second user; and

instructions for updating the first user profile based on a document selected by the first user from the set of search result documents, including analyzing links within the document and adding information derived from the analyzed links to the first user profile.

28. (Original) The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

29. (Original) The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

30. (Original) The system of claim 27, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

31. (Cancelled)

32. (Currently Amended) The system of claim 27 [[31]], wherein the information derived from the analyzed links that is added to the first user profile is added to a link-based profile and includes information about URLs or portions of URLs.

33. (Original) The system of claim 32, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

34. (Original) The system of claim 33, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

35. (Original) The system of claim 30, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

36. (Original) The system of claim 35, wherein the weight is a weight associated with occurrences of the term in the set of documents.

37. (Original) The system of claim 35, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

38. (Original) The system of claim 30, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

39. (Original) The system of claim 38, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

40. (Original) The system of claim 27, wherein the user profile is stored on a server of the search engine.

41. (Original) The system of claim 27, wherein the user profile is stored on a client associated with the user.

42. (Previously presented) The system of claim 27, wherein the first user's user profile corresponds to a group of users.

43. (Currently Amended) A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for accessing a user profile for a user based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from a search engine, documents accessed by the user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user;

instructions for receiving a search query from a first user and the same search query from a second user that is different from the first user;

instructions for identifying a set of generic search result documents that match the search query;

instructions for assigning a generic score to each document of at least a plurality of the search result documents;

instructions for assigning first and second personalized scores to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the first user's user profile and the second user's user profile, respectively;

instructions for ranking at least the plurality of the search result documents into first and second orders, respectively, according to their first and second personalized scores; [[and]]

instructions for providing the ranked set of search result documents in the first order to the first user and the ranked set of search result documents in the second order to the second user; and

instructions for updating the first user profile based on a document selected by the first user from the set of search result documents, including analyzing links within the document and adding information derived from the analyzed links to the first user profile.

44. (Original) The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

45. (Original) The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

46. (Original) The computer program product of claim 43, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

47. (Cancelled)

48. (Currently Amended) The computer program product of claim 43 [[46]], wherein the information derived from the analyzed links that is added to the first user profile is added to a link-based profile and includes information about URLs or portions of URLs.

49. (Original) The computer program product of claim 48, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

50. (Original) The computer program product of claim 49, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

51. (Original) The computer program product of claim 46, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

52. (Currently Amended) The computer program product of claim [[48]] 51, wherein the weight is a weight associated with occurrences of the term in the set of documents.

53. (Currently Amended) The computer program product of claim [[48]] 51, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

54. (Original) The computer program product of claim 46, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

55. (Original) The computer program product of claim 54, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

56. (Original) The computer program product of claim 43, wherein the user profile is stored on a server of the search engine.

57. (Original) The computer program product of claim 43, wherein the user profile is stored on a client associated with the user.

58. (Previously presented) The computer program product of claim 43, wherein the first user's user profile corresponds to a group of users.

REMARKS

This amendment responds to the final office action mailed October 11, 2007. In the final office action the Examiner:

- rejected claims 1, 16, 27, 41, 43 and 57 under 35 U.S.C. §102(b) as being anticipated by Breese et al. (US 6,006,218) (hereinafter "Breese");
- rejected claims 2-3, 18, 22, 24, 28-29 and 44-45 under 35 U.S.C. §103(a) as being unpatentable over Breese;
- rejected claim 21 under 35 U.S.C. §103(a) as being unpatentable over Breese in view of Gerace (US 5,848,396) (hereinafter "Gerace");
- rejected claims 4-7, 9-15, 17, 19-20, 23, 30-33, 35-40, 42, 46-49, 51-56 and 58 under 35 U.S.C. §103(a) as being unpatentable over Breese in view of Konig et al. (US 6,981,040) (hereinafter "Konig");
- rejected claims 8, 34, and 50 under 35 U.S.C. §103(a) as being unpatentable over Breese in view of Konig further in view of Gabriel et al. (US 6,584,468) (hereinafter "Gabriel"); and
- rejected claims 25-26 under 35 U.S.C. §103(a) as being unpatentable over Breese in view of Konig further in view of Dumais et al. (US 2004/0267700) (hereinafter "Dumais").

Claim Amendments

In this Amendment, claims 1, 4, 6, 12, 18, 27, 32, 43, 48, 52 and 53 have been amended. Claims 5, 31 and 47 have been cancelled without prejudice. No new matter has been added. Support for the amendments can be found in previously pending claim 5. After entry of this amendment, the pending claims are: claims 1-4, 6-30, 32-46 and 48-58.

Claim Rejections under 35 USC 102 and 35 USC 103

Applicants respectfully submit that the cited references, either individually or in combination, do not teach or suggest each and every limitation of independent claims 1, 18, 27 and 43, as amended. For instance, as amended, claim 1 recites:

> updating the first user profile based on a document selected by the first user from the set of search result documents, including
>
> analyzing links within the document and adding information derived from the analyzed links to the first user profile....

Breese teaches methods and apparatus for retrieving information and/or processing retrieved information as a function of a user's estimated knowledge. Breese discloses:

> After system initialization, operation progresses from step 220 to step 222. In step 222 *input relating to, e.g., the search to be performed, user attributes, user preferences and/or the user's existing knowledge about items included in the information database to be searched, is obtained*, e.g., via the user input device 170, or from the Internet browser application 131. For example, the Internet browser application 131 may provide information on which sites a user likes based on a user's frequent access of certain sites and information on a user's actual knowledge by keeping a list of sites previously visited by the user.... In step 224, the user database 104 is accessed to obtain therefrom stored information about the user's attributes, preferences, actual knowledge and/or previous searches. This information is used to supplement the information obtained in step 222. In addition to *obtaining information from the user database 104, the information in the user database is updated* in step 224 to reflect the most recent data obtained in step 222. (Breese, col. 8, lines 15-35); Emphasis Added.

Thus, Breese only discloses updating a user database based on the search to be performed, user attributes, user preferences and/or the user's existing knowledge about items included in an information database to be searched. Breese does not teach or suggest "updating the first user profile based on a *document selected by the first user from the set of search result documents, including analyzing links within the document* and adding information derived from the analyzed links to the first user profile," as claimed.

Konig does not supply the missing limitations. Konig discloses:

> The User Model is a dynamic entity that is refined and updated based on all user actions. (Konig, col. 21, lines 63-64).
>
> *Information about each document that the user views is stored in a recently accessed buffer for subsequent analysis*. The recently accessed buffer includes information about the document itself and information about the user's interaction with the document. (Konig, col. 22, lines 27-31); Emphasis Added.
>
> *After a document stored in the recently accessed buffer is parsed, parsed portions are stored in candidate tables*. For example, FIGS. 15A and 15B illustrate a user site candidate table and user word candidate table. The user site candidate table holds sites that are candidates to move into the user site distribution of FIG. 4B. *The site candidate table stores the site*

> *name, i.e., the URL* until the first backslash, except for special cases; the number of site accesses; and the time of last access. (Konig, col. 22, line 64 to col. 23, line 5); *See* also Figure 15A; Emphasis Added.

Thus, Konig discloses storing (in a site candidate table) a URL for a document that is viewed by the user. However, Konig does not teach or suggest *analyzing links **within** the document* and adding information derived from the analyzed links to the first user profile, as claimed.

Further, none of Gerace, Gabriel and Dumais supplies the missing limitations. Gerace discloses a method of profiling a user by obtaining a pattern of the user's viewing actions or viewing habits of agate information. (Gerace, col. 4, lines 12-15). Gabriel discloses identifying links within web pages and evaluating the links to determine the link's relevancy to a given technical field. (Gabriel, col. 3, lines 50-52). Dumais discloses automatically determining whether a user has contemplated a data source selected from two disparate data sources and indexing the data source. (Dumais, col. 9, claim 28). None of Gerace, Gabriel and Dumais teaches or suggests "updating the first user profile based on a *document selected by the first user from the set of search result documents, including analyzing links within the document* and adding information derived from the analyzed links to the first user profile," as claimed.

As none of the cited references teach each and every limitation of claim 1, claim 1 (and associated dependent claims) are patentable over the cited references. Independent claims 18, 27 and 43, as amended, recite similar limitations as claim 1. Therefore, claims 18, 27 and 43 and associated dependent claims are also patentable over the cited references.

In light of the above amendments and remarks, the Applicant respectfully requests that the Examiner reconsider this application with a view towards allowance. The Examiner is invited to call the undersigned attorney at (650) 843-4000, if a telephone call could help resolve any remaining items.

Respectfully submitted,

Date: December 12, 2007

/ Gary S. Williams / 31,066

Gary S. Williams (Reg. No.)
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, CA 94306
(650) 843-4000

Electronic Patent Application Fee Transmittal

Application Number:	10676711
Filing Date:	30-Sep-2003
Title of Invention:	Personalization of web search
First Named Inventor/Applicant Name:	Stephen R. Lawrence
Filer:	Gary Scott Williams/Beverly Gemello
Attorney Docket Number:	60963-0014-US

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($)
Miscellaneous:				
Request for continued examination	1801	1	810	810
		Total in USD ($)		**810**