Exhibit A

US006199067B1

(12) **United States Patent**
Geller

(10) **Patent No.: US 6,199,067 B1**
(45) **Date of Patent: Mar. 6, 2001**

(54) **SYSTEM AND METHOD FOR GENERATING PERSONALIZED USER PROFILES AND FOR UTILIZING THE GENERATED USER PROFILES TO PERFORM ADAPTIVE INTERNET SEARCHES**

(75) Inventor: **Ilya Geller**, Brooklyn, NY (US)

(73) Assignee: **Mightiest Logicon Unisearch, Inc.**, Brooklyn, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/422,286**

(22) Filed: **Oct. 21, 1999**

Related U.S. Application Data

(60) Provisional application No. 60/116,582, filed on Jan. 20, 1999.

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

(52) **U.S. Cl.** **707/10**; 707/3; 707/5; 707/100; 704/4; 704/9; 704/247; 700/17

(58) **Field of Search** 707/1–5, 100, 707/10, 200; 704/1, 4, 9, 270.1, 247, 250; 382/181, 190, 209; 700/11, 17, 56, 83, 86; 369/13

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,914,590	*	4/1990	Loatman et al.	704/8
5,483,650	*	1/1996	Pedersen et al.	707/2
5,696,962	*	12/1997	Kupice	707/4

(List continued on next page.)

OTHER PUBLICATIONS

Helm, Richanr et al., "Integrating Information Retrieval and Domain Specific Approaches for Browsing and Retrieval in Object–Oriented Class Libraries," Conference proceedings on Object–Oriented programming systems, languages, and applications, May 1991. p.*

Maass, Henning, "Location–aware mobile applications based on dictionary services," Mobile Networks and Applications, vol. 3, Issue 2, Aug. 1998, pp. 157–173.*

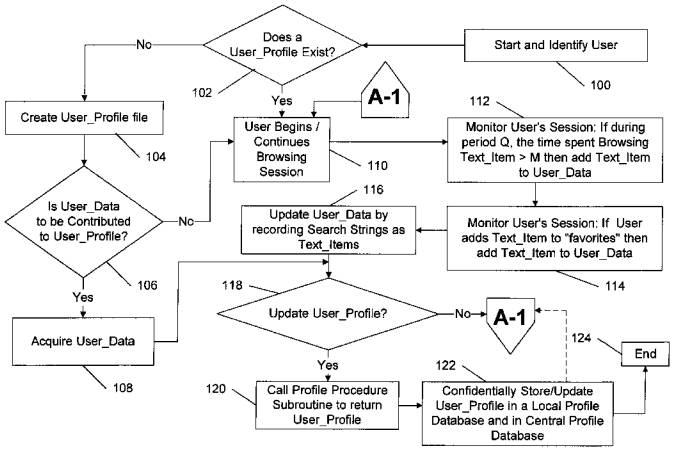
Xu, Yaowu et al., "Hierarchical Content Description and Object Formation by Learning," Proceedings of the IEEE Workshop on Content–Based Access of Image and Video Libraries, (CBAIVL '99), Jun. 22, 1999, pp. 84–88, May 1991.*

Primary Examiner—Hosain T. Alam
Assistant Examiner—Shahid Alam
(74) *Attorney, Agent, or Firm*—Edward Etkin, Esq.

(57) **ABSTRACT**

A system and method for automatically generating personalized user profiles and for utilizing the generated profiles to perform adaptive Internet or computer data searches is provided. In accordance with the present invention, particular linguistic patterns and their frequency of recurrence are extracted from personal texts provided by the users of the system of the present invention and stored in a user profile data file such that the user profile data file is representative of the user's overall linguistic patterns and the frequencies of recurrence thereof. All documents in a remote computer system, such as the Internet, are likewise analyzed and their linguistic patterns and pattern frequencies are also extracted and stored in corresponding document profiles. When a search for particular data is initiated by the user, linguistic patterns are also extracted from a search string provided by the user into a search profile. The user profile is then cross matched with the search profile and the document profiles to determine whether any linguistic patterns match in all three profiles and to determine the magnitude of the match based on summation of respective frequencies of recurrence of the matching patterns. The documents with document profiles having the highest matching magnitudes are presented to the user as not only matching the subject of the search string, but also as corresponding to the user's cultural, educational, and social backgrounds as well as the user's psychological profile.

62 Claims, 8 Drawing Sheets



Profiles with the highest FINAL_VALUEs and presents a list of the M Data_Item_Addresses to the user in order of descending magnitude of their corresponding FINAL_VALUEs. The number M of the Data_Item_Addresses presented may be selected as a matter of design choice. For example, M may be set to 10 or 20. At an optional step **440**, the RCS control unit **34** automatically retrieves and opens, for the user, the Data_Item corresponding to the Data_Item_Profile with the highest FINAL_VALUE.

Thus, while there have shown and described and pointed out fundamental novel features of the invention as applied to preferred embodiments thereof, it will be understood that various omissions and substitutions and changes in the form and details of the devices and methods illustrated, and in their operation, may be made by those skilled in the art without departing from the spirit of the invention. For example, it is expressly intended that all combinations of those elements and/or method steps which perform substantially the same function in substantially the same way to achieve the same results are within the scope of the invention. It is the intention, therefore, to be limited only as indicated by the scope of the claims appended hereto.

I claim:

1. A data processing method for enabling a user utilizing a local computer system having a local data storage system to locate desired data from a plurality of data items stored in a remote data storage system in a remote computer system, the remote computer system being linked to the local computer system by a telecommunication link, the method comprising the steps of:

(a) extracting, by one of the local computer system and the remote computer system, a user profile from user linguistic data previously provided by the user, said user data profile being representative of a first linguistic pattern of the said user linguistic data;

(b) constructing, by the remote computer system, a plurality of data item profiles, each plural data item profile corresponding to a different one of each plural data item stored in the remote data storage system, each of said plural data item profiles being representative of a second linguistic pattern of a corresponding plural data item, each said plural second linguistic pattern being substantially unique to each corresponding plural data item;

(c) providing, by the user to the local computer system, search request data representative of the user's expressed desire to locate data substantially pertaining to said search request data;

(d) extracting, by one of the local computer system and the remote computer system, a search request profile from said search request data, said search request profile being representative of a third linguistic pattern of said search request data;

(e) determining, by one of the local computer system and the remote computer system, a first similarity factor representative of a first correlation between said search request profile and said user profile by comparing said search request profile to said user profile;

(f) determining, by one of the local computer system and the remote computer system, a plurality of second similarity factors, each said plural second similarity factor being representative of a second correlation between said search request profile and a different one of said plural data item profiles, by comparing said search request profile to each of said plural data item profiles;

(g) calculating, by one of the local computer system and the remote computer system, a final match factor for each of said plural data item profiles, by adding said first similarity factor to at least one of said plural second similarity factors in accordance with at least one intersection between said first correlation and said second correlation;

(h) selecting, by one of the local computer system and the remote computer system, one of said plural data items corresponding to a plural data item profile having a highest final match factor; and

(i) retrieving, by one of the local computer system and the remote computer system from the remote data storage system, said selected data item for display to the user, such that the user is presented with a data item having linguistic characteristics that substantially correspond to linguistic characteristics of the linguistic data generated by the user, whereby the linguistic characteristics of the data item correspond to the user's social, cultural, educational, economic background as well as to the user's psychological profile.

2. The method of claim **1**, further comprising the step of:

(j) prior to said step (a), automatically adding, by one of the local computer system and the remote computer system, textual data generated by the user during utilization of the local computer system to said user linguistic data.

3. The method of claim **1**, wherein said user linguistic data comprises at least one of: personal textual data generated by the user and favorite textual data generated by a source other than the user and that the user has adopted as being favorite.

4. The method of claim **1**, wherein said user linguistic data comprises at least one text item, each said at least one text item comprising at least one sentence.

5. The method of claim **3**, further comprising the step of:

(k) prior to said step (a), selecting, by the user at least one of said personal textual data and said favorite textual data, from textual data stored in one of the local data storage system and the remote data storage system.

6. The method of claim **1**, further comprising the step of:

(l) prior to said step (a), determining, by one of the local computer system and the remote computer system, whether an existing user data profile is stored in one of the local data storage system and the remote data storage system, and:

1) when an existing user data profile is stored in one of the local data storage system and the remote data storage system, retrieving said existing user data profile and proceeding to said step (b); and

2) when an existing user data profile is not stored in one of the local data storage system and the remote data storage system, proceeding to said step (a).

7. The method of claim **4**, wherein said step (a) comprises the steps of:

(m) generating, by one of the local computer system and the remote computer system, a user data profile;

(n) retrieving, by one of the local computer system and the remote computer system, a text item from said user linguistic data;

(o) separating, by one of the local computer system and the remote computer system, said text item into at least one sentence;

(p) extracting, from each of said at least one sentence, by one of the local computer system and the remote computer system, at least one segment representative of a linguistic pattern of each sentence of said at least one sentence;

34. The method of claim 33, further comprising the step of:

(iii) after said step (ggg), determining whether each word may serve as an additional part of speech, and when a word may serve as an additional part of speech, adding an additional tag to said word to identify said word as said additional part of speech.

35. The method of claim 33, wherein said predetermined plurality of different parts of speech comprises at least one of: noun, pronoun, verb, adverb, adjective, gerund, proposition, conjunction and interjection.

36. The method of claim 33, wherein said predetermined plurality of different parts of speech comprises a noun, a verb and an adjective, wherein said predetermined number is three, and wherein said predetermined order is noun, verb, adjective.

37. The method of claim 33, wherein said step (hhh) further comprises the step of:

(jjj) when one of said predetermined plural different parts of speech is missing from said sentence, inserting a blank mark into said segment instead of said missing predetermined part of speech.

38. The method of claim 33, further comprising the steps of:

(kkk) determining, by one of the local computer system and the remote computer system, at least one synonym for each word in each segment;

(lll) composing, by one of the local computer system and the remote computer system, a plurality of alternate search segments for each segment utilizing said synonyms, wherein said alternate search segments are composed in accordance with said predetermined order of said predetermined plural different parts of speech; and

(mmm) recording, by one of the local computer system and the remote computer system, said plural alternate search segments in said search profile.

39. The method of claim 30, wherein said step (e) comprises the steps of:

(nnn) retrieving, by one of the local computer system and the remote computer system, said user data profile from one of the local data storage system and the remote data storage system; and

(ooo) comparing, by one of the local computer system and the remote computer system, said at least one user segment group to said at least one search segment, and recording said user segment counts of each user segment group of said at least one user segment group that matches a corresponding search segment of said at least one search segment, said user segment counts being representative of said first similarity factor.

40. The method of claim 39, wherein said step (f) comprises the steps of:

(ppp) for each plural data item, retrieving, by one of the local computer system and the remote computer system, a corresponding data item profile from the remote data storage system; and

(qqq) for each plural data item profile, comparing, by one of the local computer system and the remote computer system, said at least one data segment group to said at least one search segment, and recording said data segment counts of each data segment group of said at least one data segment group that matches a corresponding search segment of said at least one search segment, said data segment counts being representative of said plural second similarity factor.

41. The method of claim 40, wherein said step (g) comprises the steps of:

(rrr) for each said plural data item profile, determining a least one match value, by one of the local computer system and the remote computer system, by first identifying a data segment group in the plural data item profile that matches both a corresponding search segment and a corresponding user segment group and then adding said user segment count of said corresponding user segment group to said data segment count of said identified data segment group, wherein when no matches are identified, said at least one match value is set to null; and

(sss) for each said plural data item profile, determining a final match factor, by one of the local computer system and the remote computer system, by adding together all said at least one match values determined for said plural data item profile at said step (rrr).

42. The method of claim 40, wherein said step (ppp) comprises the steps of:

(ttt) applying, by the remote computer system, said search request data to a conventional data search engine, implemented in the remote computer system, to return a list of at least one data item address of at least one preliminary matching data item that potentially corresponds to said search request data; and

(uuu) retrieving from the remote storage system, by one of the local computer system and the remote computer system, at least one data item profile corresponding to said at least one preliminary matching data item in said list.

43. The method of claim 1, wherein said step (h) comprises the steps of:

(vvv) selecting, by one of the local computer system and the remote computer system, a portion of said plural data items corresponding to a predetermined number of plural data item profiles having highest final match factors; and

wherein said step (i) comprises the step of:

(www) retrieving, by one of the local computer system and the remote computer system from the remote data storage system, said selected data items for display to the user, such that the user is presented with a group of data items having linguistic characteristics that substantially correspond to linguistic characteristics of the linguistic data generated by the user, whereby the linguistic characteristics of the data items correspond to the user's social, cultural, educational, economic background as well as to the user's psychological profile.

44. A data processing method for enabling a user, utilizing a computer system having a data storage system, to locate desired data from a plurality of data items stored in the data storage system, the method comprising the steps of:

(a) extracting, by the local computer system, a user profile from user linguistic data previously provided by the user, said user data profile being representative of a first linguistic pattern of the said user linguistic data;

(b) constructing, by the computer system, a plurality of data item profiles, each plural data item profile corresponding to a different one of each plural data item stored in the data storage system, each of said plural data item profiles being representative of a second linguistic pattern of a corresponding plural data item, each said plural second linguistic pattern being substantially unique to each corresponding plural data item;

second extracting means, connected to said first input means, for extracting a search request profile from said acquired search request data, said search request profile being representative of a third linguistic pattern of said search request data;

second control means, connected to said first extracting means and said second extracting means, for determining a first similarity factor representative of a first correlation between said search request profile and said user profile by comparing said search request profile to said user profile;

third control means, connected to said first control means and said second extracting means, for determining a plurality of second similarity factors, each said plural second similarity factor being representative of a second correlation between said search request profile and a different one of said plural data item profiles, by comparing said search request profile to each of said plural data item profiles;

fourth control means, connected to said second an said third control means, for calculating a final match factor for each of said plural data item profiles, by adding said first similarity factor to at least one of said plural second similarity factors in accordance with at least one intersection between said first correlation and said second correlation;

first selection means, connected to said fourth control means, for selecting one of said plural data items corresponding to a plural data item profile having a highest final match factor; and

first retrieving means, connected to said first selection means, for retrieving, from the data storage system, said selected data item for display to the user, such that the user is presented with a data item having linguistic characteristics that substantially correspond to linguistic characteristics of the linguistic data generated by the user, whereby the linguistic characteristics of the data item correspond to the user's social, cultural, educational, economic background as well as to the user's psychological profile.

61. The method of claim **1**, wherein the remote computer system comprises a plurality of computer systems connected to the Internet and the World Wide Web.

62. The system of claim **59**, wherein the remote computer system comprises a plurality of computer systems connected to the Internet and the World Wide Web.

* * * * *