EXHIBIT 1

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.*

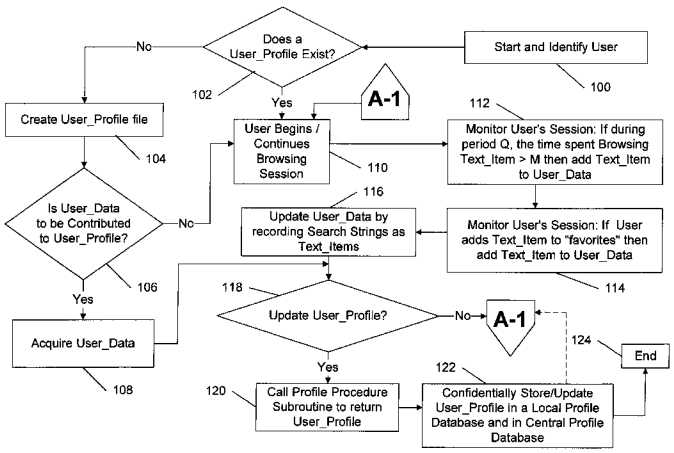
Xu, Yaowu et al., "Hierarchical Content Description and Object Formation by Learning," Proceedings of the IEEE Workshop on Content–Based Access of Image and Video Libraries, (CBAIVL '99), Jun. 22, 1999, pp. 84–88, May 1991.*

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



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

RELATED APPLICATIONS

This application claims priority from U.S. Provisional Patent Application Ser. No. 60/116,582, entitled "Internet Search Vehicles" which was filed on Jan. 20, 1999.

FIELD OF THE INVENTION

The present invention relates generally to the computer data searches and more particularly to a system and method for automatically generating personalized user profiles and for utilizing the generated profiles to perform adaptive Internet or computer data searches.

BACKGROUND OF THE INVENTION

In recent years, computers have taken the world by storm. Today, most businesses entirely rely on computers to conduct daily operations. In the academic world, computers have become essential tools for learning, teaching and research. In homes, computers are used to perform daily tasks ranging from paying bills to playing games. The one unifying requirement for all computer applications is the ability of a user to utilize a computer to locate particular information or data desired by the user.

During the past few years, the quantity and diversity of information and services available over the public (e.g. Internet) and private (e.g. Intranet) local and wide area networks has grown substantially. In particular, the variety of information accessible through Internet-based services is growing rapidly both in terms of scope and depth. In simple terms, the Internet is a massive collection of individual computer networks operated by government, industry, academia, and private parties that are linked together to exchange information. While originally, the Internet was used mostly by scientists, the advent of the World Wide Web has brought the Internet into mainstream use. The World Wide Web (hereafter "WWW") is an international, virtual-network-based information service composed of Internet host computers that provide on-line information in a specific hypertext format. WWW host servers provide hypertext metalanguage (HTML) formatted documents using a hypertext transfer protocol (HTTP). Information on the WWW is accessed with a hypertext browser, such as the Netscape navigator or Microsoft Explorer. Web sites are collections of interconnected WWW documents.

Typically, users communicate with the Internet through a communication gateway that may be implemented and controlled by an Internet service provider (i.e. an ISP)—a company that offers a user access to the Internet and the WWW through a software application that controls communication between the user's computer and the communication gateway. The role of the ISP may also be taken directly by a particular organization that allows internet access to its employees or members. The user can access and navigate the WWW using a hypertext browser application residing on, and executed by, the user's computer.

No hierarchy exists in the WWW, and the same information may be found by many different approaches. Hypertext links in WWW HTML documents allow readers to move from one place in a document to another (or even between documents) as they want to. One of the advantages of

WWW, is that there is no predetermined order that must be followed in navigating through various WWW documents. Readers can explore new sources of information by following links from place to place. Following links has been made as easy as clicking a mouse button on the link related to the subject a user wants to access. Each WWW document also has a unique uniform resource locator ("URL") that serves as an "address" that, when followed, leads the user to the document or file location on the WWW. Using the browser, the user can also mark and store "favorites"—URLs of particular WWW documents that interest the user such that the user can quickly and easily return to these documents in the future by selecting them from the favorites list in the browser.

Because of the vastness of the Internet and the WWW, locating specific information desired by the user can be very difficult. To facilitate search for information a number of "search engines" have been developed and implemented. A search engine is a software application that searches the Internet for web sites containing information on the subject in which the user is interested. These searches are accomplished in a variety of ways—all well-known in the art. Typically, a user first inputs a "search string" to the hypertext browser containing key words representative of the information desired by the user. The search engine then applies the search string to a previously constructed index of a multitude of web sites to locate a certain number of web sites having content that matches the user's search string.

The located web site URLs are then presented to the user in the order of relevance to the key words in the user's search string. For example, a user providing the key word PLANT would obtain an exhaustive list of all registered sites that refer to plants. This list, however would be so large that the user would want to limit this search. Depending on the search engine used, the user could limit the search by entering a combination of key words such as the following: PLANT AND FLOWER AND GARDEN. This would limit the search to only Internet sites that contain all three words. In addition, users could further limit the search by entering PLANT AND FLOWER AND GARDEN NOT TREE NOT ORCHID. The results from this search would be further limited to exclude sites in which trees and orchids are listed as keywords.

A number of approaches have been developed to improve the performance and accuracy of typical key word searches. For example, U.S. Pat. No. 5,845,278, issued to Kirsch, et. al, teaches approaches to establishing a quantitative basis for selecting client database sets (i.e. Internet documents or web sites) that include the use of comprehensive indexing strategies, ranking systems based on training queries, expert systems using rule-based deduction methodologies, and inference networks. These approaches were used to examine knowledge base descriptions of client document collections or databases.

However, the key word searching approaches utilized by previously known search engines suffer from a number of significant disadvantages. Most search systems are viewed as often ineffective in identifying the likely most relevant documents. Accordingly, the users are often presented with overwhelming amounts of information in response to their key words. Thus, using proper key word searching techniques becomes an art in itself—an art that is outside the capabilities of most Internet users.

Most importantly, typical key word and even more advanced searches only provide the user with search results that depend entirely on the search string entered by the user,

without any regard to the user's cultural, educational, social backgrounds or the user's psychological profiles. The results returned by the search engines are tailored only to the search string provided by the user and not to the user's background. None of the previously known search engines tailor results of user's searches based on his or her background and unexpressed interests. For example, a twelve year old child using key word searches on the Internet for some information on computers may be presented with a multitude of documents that are far above the child's reading and educational level. In another example, a physician searching the Internet for information on a particular disease may be presented with dozens of web sites that contain very generic information, while the physician's "unexpressed" interest was to find web sites about the disease that are on his educational and professional level.

It would thus be desirable to provide a system and method for extracting and using linguistic patterns of textual data to assist a user in locating requested data that, in addition to matching the user's specific request, also corresponds to the user's professional, cultural, educational, and social backgrounds as well as to the user's psychological profile and thus addresses the user's "unexpressed" requests.

SUMMARY OF THE INVENTION

This invention relates to use of linguistic patterns of documents to assist a user in locating requested data that, in addition to matching the user's specific request, also corresponds to the user's cultural, educational, professional, and social backgrounds as well as to the user's psychological profile, and thus addresses the user's "unexpressed" requests. The present invention provides a system and method for automatically generating a personalized user profile based on linguistic patterns of documents provided by the user and for utilizing the generated profile to perform adaptive Internet or computer data searches.

The system of the present invention advantageously overcomes the drawbacks of previously known data searching techniques. As was noted earlier, typical key word and even more advanced searches only provide the user with search results that depend entirely on the search string entered by the user, without any regard to the user's cultural, educational, professional, and social backgrounds or the to user's psychological profile.

All texts composed by the user, or adopted by the user as favorite or inimical (such as a favorite book or short story), contain certain recurring linguistic patterns, or combinations of various parts of speech (nouns, verbs, adjectives, etc.) in sentences that reflect the user's cultural, educational, social backgrounds and the user's psychological profile. Research has shown that most people have readily identifiable linguistic patterns in their expression and that people with similar cultural, educational, and social backgrounds will have similar linguistic patterns. Furthermore, research has shown that such factors as psychological profile, life experience, profession, socioeconomic status, educational background, etc. contribute to determining the frequency of occurrences of particular linguistic patterns within the user's written expression.

In accordance with the present invention, particular linguistic patterns and their frequencies of occurrence are extracted from the texts provided by a user of the system of the present invention and stored in a user profile data file. The user profile data file is thus representative of the user's overall linguistic patterns and their respective frequencies. All documents in a remote computer system, such as the Internet, are likewise analyzed and their linguistic patterns and frequencies thereof 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 relative frequencies of matching patterns in the user profile and the document profile. 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. Thus, a world renowned physicist searching for information on quasars would be presented with very sophisticated physics documents that are oriented to wards his level of expertise.

It should be noted that the user's background and psychological characteristics are not evident directly from the linguistic patterns themselves or form their frequencies. Accordingly, the system of the present invention matches the user's linguistic patterns to the linguistic patterns of data requested by the user without extracting any actual information about the user's background and psychological characteristics from the user profile. Thus, the user's privacy is not impinged by the creation and retention of the user profile.

The profiling/search system includes a local computer system, connected to a remote computer network (e.g. the Internet) via a telecommunication link. The local computer system includes a control unit and related circuitry for controlling the operation of the local computer system and for executing application programs, a memory for temporarily storing control program instructions and variables during the execution of application programs by the control unit; a storage memory for long term storage of data and application programs; and input devices for accepting input from the user. The local computer system further includes: output devices for providing output data to the user and a communication device for transmitting to, and receiving data from, the remote computer system via the telecommunication link. The remote computer system includes a communication gateway connected to the telecommunication link, a remote data storage system for long term data storage, and a remote computer system control unit (hereinafter RCS control unit).

In summary, the system of the present invention operates in three separate independent stages, each stage being controlled by a particular control program executed by one of the local computer system and the remote computer system. In a first stage, a user profiling control program is executed to generate or update a user profile computer file representative of the user's linguistic patterns and the frequencies with which these patterns recur in texts submitted by the user and/or automatically acquired by the inventive system. The user is then invited to provide textual data composed by the user such as e-mail messages, memorandums, essays as well as documents composed by others that the user has adopted as "favorites", such as favorite web sites, short stories, etc. These textual documents are temporarily stored in a user data file. The inventive system also monitors the user's data searching and data browsing (e.g. Internet browsing) to automatically add additional textual information to the user data file. Once the user data file attains a sufficient size, or when other criteria for updating the user profile are met, the