

FREE patent keyword monitoring and additional FREE benefits. ▶ [REG now for FREE](#) ◀

FreshPatents.com
Track New Patents and Technologies



monitor



organizer



location



industry



inventors



ag

Title/Abstract/Num. ▼

FreshPatents Search

Ads by Google

[חיפוש פטנטים](#)

לפני פניה לעורך פטנטים
www.New-Tone.co.il

[Pioneer Recruitment Ltd](#)

Exclusive primary care provider
for GP's work available across
UK
www.pioneergrp.co.uk

[Inventors](#)

Are You an Inventor? Find
Invention Resources & Help
Online Here.
www.PureBusiness.com

[YossiP Projects](#)

יעץ מומחה בתחום המדען הראשי
www.yossip.com

[Medical Directory](#)

news and resources related to
healthcare and the medical field
www.MedMark.org

[Advertise on this site](#)



A **FREE** service from FreshPatents...**PATENT KEY**
3 steps to **Unlock the Secrets of new technologies** and its
workers or other inventors!

1. [Sign up](#) (takes 30 seconds).
2. [Fill in the keywords](#) to be monitored.
3. Each week you receive an email with patent application
keywords. [Start now!](#)

[05/18/06](#) | [#20060106760](#) | [Browse Patent Application](#)
[Browse Industry: USPTO Class 707](#)

Method and apparatus of inter-document data retrieval



[Save to Organizer](#)



[View Organizer](#)



[Monitor Keywords](#)

[Brief Patent Description](#) - [Full Patent Description](#) - [Patent Claims](#)

BACKGROUND OF THE INVENTION

[0001] During the last decades the importance of information technology has increased. The amount of information has dramatically increased. A search engine can access more than 10^{sup.9} web pages. The development of telecommunications and computers technology has allowed millions of people to search for information using a variety of devices. Modern data mining and data management technology has allowed access to relevant information.

[0002] In view of the vast amount of information various data mining techniques and data retrieval methods were are aimed to locate relevant documents out of a large do Common search methods include keyword-based methods and the like.

[0003] The following U.S. patents, all being incorporated by reference, provide a brief view of some state of the art search devices as well of some state of the art data retrieval methods. U.S. Pat. No. 6,523,026 of Gillis, U.S. Pat. No. 6,681,217 of Pat. No. 6,721,728 of McGreevy, U.S. Pat. No. 6,718,322 of U.S. Pat. No. 6,681,217 of Lewak, U.S. Pat. No. 6,151,600 of U.S. Pat. No. 6,026,388 of Liddy et al. U.S. Pat. No. 6,006,221 of Liddy et al., U.S. Pat. No. 5,412,807 of Moreland, U.S. Pat. No. of Meek, and U.S. Pat. No. 5,915,251 of Burrows et al.

[0004] One keyword-based search method is known as proximity search. It allows a client to define a search query that includes keywords and a distance between said keywords. Any document includes all the keywords positioned within said distance as a search result.

[0005] The distance between the keywords can be fixed or defined as a default value that can be altered by the client. A command is typically used at keyword proximity search accompanied by the distance between said keywords.

[0006] There is a growing need to provide an efficient data retrieval method.

SUMMARY OF THE INVENTION

[0007] The invention provides method for data retrieval includes: (i) receiving a query that comprises a set of keywords; (ii) for each desired document responsive keyword association level; (iii) for at least a group of documents, whereas each document includes at least one keywords of the set but does not include all of keywords, but the group of documents as a whole includes all keywords; and whereas an association level between the

groups corresponds to the desired association level.

[0008] The invention provides method for data retrieval includes: (i) receiving a search query that comprises a set of keywords; and (ii) identifying at least one document responsive to the search query at a desired document responsive keyword association level such that at least one document that comprises all the set of keywords; and (iii) searching for at least one group of documents that comprises all the set of keywords, whereas each document of the group does not comprise all the set of keywords; and whereas a document responsive keyword association level between the keywords within each group corresponds to the desired association level.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] In order to understand the invention and to see how it is carried out in practice, a preferred embodiment will now be described by way of non-limiting example only, with reference to the drawings, in which:

[0010] FIG. 1 illustrates a data retrieval system according to an embodiment of the invention;

[0011] FIG. 2 illustrates two documents that include two keywords and two pieces of metadata, according to an embodiment of the invention;

[0012] FIGS. 3 and 4 are flow charts of methods for data retrieval according to various embodiments of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0013] For simplicity of explanation the following detailed description refers to search queries that include two keywords. It is understood that the invention can be applied mutatis mutandis to search queries that include more than two keywords.

[0014] The usage of association level parameters allows for a multi-dimensional relation between keywords. Thus, instead of a simple NEAR operation that can associate between keywords to the same document, the method can link between keywords to indirect correlation between documents, document data, and the like.

[0015] The term "document" as used throughout the specification is defined as a set, group or any other arrangement of multiple signals including text, graphics, video, audio or any digital based signal that may contain valuable information or a combination of two or more of the above. Signals can be in various formats, including, for example, PDF, JPEG, MPEG format, and the like. It is noted that at least one of the databases is searched and that said databases can be stored in one or more locations either close to each other or remotely positioned. Examples of documents can include a book, an article, a song, one of more images, and the like.

[0016] The term "metadata" as used throughout the specification is defined as information representative to a characteristic, parameter or attribute of a certain piece of information.

[0017] FIG. 1 illustrates a data retrieval system 10 according to one embodiment of the invention. For convenience of explanation, it is assumed that multiple client devices are connected via a central database.

[0018] It is noted that the data retrieval method can be implemented in systems that differ from system 10. For example, the multiple databases that are accessed during the data retrieval process can be distributed databases. Yet for another example, the client can perform the search on a local server or computer, and the like. Yet for a further example, multiple client devices can be connected to the central database or to a distributed databases via different types of networks.

[0019] System 10 includes a central database 12 that stores documents 10(1)-10(k), as well as metadata 26 associated with the documents. This central database 12 can include multiple components as well as load balancing components, and the like. The central data base can include cache memories, firewall servers, and the like.

[0020] The central database 12 is connected to a search engine 14. The search engine 14 can include multiple hardware, software and middle ware components. The search engine 14 is connected to multiple client devices 16 via a network 18. It is noted that the client devices can include personal data accessories, cellular phones, set-top-boxes, and the like.

Usually, a client device 16 includes dedicated software client to send search queries and receive results.

[0021] Network 18 can include multiple networks, including networks, local area networks and the like. Network 18 utilizes various communication techniques such as but not limited to wireless communication, terrestrial communication, satellite communication and the like. Network 18 can utilize optical communication and frequency communication techniques, and even a combination of both can be found at hybrid fiber coax networks connected to cable modems and to set top boxes.

[0022] Search engine 14 is capable of receiving a search query that includes a set of keywords and a desired document response association level and is further adapted to search for at least one document within a document database. Each group of documents is defined by the following conditions: (i) the whole set of keywords is included in the group; (ii) each document of the group does not include all of the keywords, and (iii) an document responsive keyword association level between the keywords within the group documents of the group corresponds to the desired association level. For example, if the association level is either equal to said level or higher than that level, then the document database does not include documents that do not satisfy the query then the group of documents is an empty group.

[0023] Conveniently, the invention search engine 14 is capable of receiving other search queries. According to various embodiments, the search engine 14 is capable of performing a prior art keyword search.

[0024] FIG. 3 is a flow chart of method 100 for data retrieval, which is an embodiment of the invention.

[0025] Method 100 starts by stage 110 of receiving multiple search queries and receiving multiple keywords. Referring to the example, a central data base stores a large amount of documents, but only two documents 20(1) and 20(2) are illustrated, for simplicity. In the first document 20(1), the first keyword keyword_1 24(1) appears twice in the first document 20(1) and the second keyword keyword_2 24(2) appears once within the first document 20(1). In the second document 20(2), the first keyword keyword_1 24(1) appears once within the second document 20(2) and the second keyword keyword_2 24(2) appears twice within the second document 20(2).

[0026] Stage 110 is followed by stage 120 of defining, for each keyword a document keyword association parameter. For each keyword a document keyword association parameter can reflect the correlation between a certain document and a certain keyword.

[0027] The definition can require human intervention but can be done automatically, without human intervention. For example, the parameter can be responsive to the number of times the keyword appears within the document, to the location of the keyword within the document (for example whether the keyword appears in the title, the abstract of the document and the like), to the frequency of that keyword and other keywords or words within the document, and the like.

[0028] Referring to the example set forth in FIG. 2, stage 110 includes defining a first document keyword association parameter (denoted D1k1_A_P) 24(11) that has a certain value. It is assumed that D1k1_A_P=50, reflecting a medium correlation between the document and the keyword. Stage 120 also includes defining a second document keyword association parameter (denoted D2k2_A_P) 25(12) that has a certain value. It is assumed that each of the association parameters can range between zero (lowest correlation level) and one hundred (highest correlation level). It is further assumed that D2k2_A_P=20, reflecting a low correlation between the second document and the second keyword.

[0029] Stage 120 is followed by stage 130 of defining, for each pair of keywords an inter-keyword association parameter. Referring to the example set forth in FIG. 2, stage 120 includes defining a second keyword inter-keyword association parameter (denoted k1k2_A_P) 26(12) that has a certain value. We assume that k1k2_A_P=20, reflecting a relatively low correlation between the first and second keywords.

[0030] Stage 130 is followed by stage 140 of receiving a set of keywords and at least one desired document keyword association level (denoted desired_level). The search engine receives a received document responsive keyword association level threshold, a high association level threshold, and a range of association levels. It is assumed that the search engine receives keyword_1 and keyword_2 and that desired_level is a low association level threshold and that it equals eighty.

[0031] Stage 140 is followed by stage 150 of searching a group of documents that as a whole includes the set of keywords, each document of the group includes at least one keyword, the group includes the whole set of keywords; and whereas a document responsive keyword association level of keywords within the group is greater than a desired association level.

[0032] The document responsive keyword association level (keyword_level) of keywords is responsive to a corresponding keyword association parameter and to corresponding document association parameter. The relationship between said association parameters and the document responsive keyword association level is linear, non-linear and the like.

[0033] For simplicity of explanation we assume that the document responsive keyword level is the sum of the association parameters.
$$\text{keyword_level} = k_1k_2_A_P + D_1k_1_A_P + D_2k_2_A_P$$

[0034] In our example $\text{keyword_level} = 50 + 10 + 30 = 90$. $\text{Keyword_level} > \text{desired_level}$, thus the search result will include document_1 and document_2. If, for example, the document responsive keyword association level threshold is 80, then the search result was not relevant.

[0035] According to an embodiment of the invention the association parameters can be updated in response to the client's actions, even in response to client inputs or actions. For example, if a client receives a search result that mentioned various documents, the client can learn from the mere retrieval of these documents that mentioned various documents can be learnt from the mere retrieval of these documents by the client, or even by an additional client storage of said document, the initiation of other search criteria, or even by the client's mention of the document and the like. It is noted that the relationship between various association parameters are related to each other, and the manner in which the search is conducted can be altered.

[0036] Yet for another example the client can provide input indicating the correlation between the retrieved documents. The input can be ("relevant", "not relevant", "slightly relevant") or can be a grade reflecting the correlation between documents and even between keywords, as viewed by the client.

[0037] Stage 150 is followed by stage 160 of providing. Conveniently, the search result includes information regarding documents within a document group. Said information comprises names of the documents, their location, a link for fast retrieval of documents, and the like. The information can further provide about the relevant association parameters, and can also include documents and the like.

[0038] FIG. 4 illustrates a data retrieval method 200, according to an embodiment of the invention.

[0039] Method 200 differs from method 100 by including stage 145 of searching at least one document that includes keywords. The search result provided during stage 160 comprises documents that were found during stages 150 and 154.

[0040] According to other embodiments of the invention, the documents are processed before being used to retrieve documents. The various operations such as parsing or stemming.

[0041] The present invention can be practiced by employing various tools, methodology and components. Accordingly, the details of the tools, component and methodology are not set forth here. As in the previous descriptions, numerous specific details (such as a compression standard) are set forth in order to provide a clear understanding of the present invention. However, it should be understood that the present invention might be practiced without resorting to the specific details specifically set forth.

[0042] Only exemplary embodiments of the present invention are shown and described in the disclosure. It is to be understood that the present invention can be used in various other combinations and environments and that changes or modifications within the scope of the invention are expressed herein.

[Brief Patent Description](#) - [Full Patent Description](#) - [Patent Claims](#)

Click on the above for other options relating to this Method

of inter-document data retrieval patent application.

Related - 20060149710 - ([abstract](#)) - Associating feature such as categories of web page documents, and/or weight

Related - 707003000 - ([abstract](#)) - Automated management images for efficient resource node building within a grid

Related - 707003000 - ([abstract](#)) - Computer readable medium and apparatus for preserving filtering conditions to query sources at various locales when regenerating a report

Related - 707003000 - ([abstract](#)) - Distributed search system

Related - 707003000 - ([abstract](#)) - Image forming apparatus

Related - 707003000 - ([abstract](#)) - Infringer finder

Related - 707003000 - ([abstract](#)) - Method and apparatus for recommendations in a group resource environment

Related - 707003000 - ([abstract](#)) - Method of locating web pages utilizing visual images

Related - 707003000 - ([abstract](#)) - Methods and apparatus of unconsciously captured documents

Related - 707003000 - ([abstract](#)) - Search engine for a video

Related - 707003000 - ([abstract](#)) - Search method and system using the same

Related - 707003000 - ([abstract](#)) - Searching based on ontology

Related - 707003000 - ([abstract](#)) - System and method for

Related - 707003000 - ([abstract](#)) - System and method for results with configurable scoring formula

Related - 707003000 - ([abstract](#)) - System and method for information from citation-rich documents

Related - 707003000 - ([abstract](#)) - System, method, and product for improving accuracy of cache-based searches

###

 How **KEYWORD MONITOR** works... a **FREE** service
FreshPatents

1. [Sign up](#) (takes 30 seconds). 2. [Fill in the keywords](#) to
3. Each week you receive an email with patent application
keywords.

[Start now!](#) - Receive info on patent apps like Method and
inter-document data retrieval or other areas of interest.

###

Previous Patent Application:

[Method and apparatus for predicting selectivity of datab conditions using hypothetical query predicates having sl constants](#)

Next Patent Application:

[Method and system for autocompletion for languages ha and phonetic characters](#)

Industry Class:

[Data processing: database and file management or data :](#)

###

[FreshPatents.com Support](#) | [Advertise on FreshPatents.c](#)

Design/code © 2004-2005 Freshpatents.com. [Website T Conditions](#)

Patent data source: United States Patent and Trademark Information published here is an abstract for research/ec purposes.

Complete official applications are on file at the USPTO additional data/images.

FreshPatents.com is not affiliated with or endorsed by th Thank you for viewing the *Method and apparatus of int retrieval* patent info.

[IP-related news and info](#)

Results in 0.78572 seconds

Other interesting Feshpatents.com categories:

Electronics: [Semiconductor](#), [Audio](#), [Illumination](#), [C Crypto](#),