



(19) **United States**

(12) **Patent Application Publication**
Tong

(10) **Pub. No.: US 2004/0255237 A1**

(43) **Pub. Date: Dec. 16, 2004**

(54) **DOCUMENT SEARCH ENGINE INCLUDING HIGHLIGHTING OF CONFIDENT RESULTS**

(57) **ABSTRACT**

(76) Inventor: **Simon Tong**, Mountain View, CA (US)

Correspondence Address:
HARRITY & SNYDER, LLP
11240 WAPLES MILL ROAD
SUITE 300
FAIRFAX, VA 22030 (US)

(21) Appl. No.: **10/457,488**

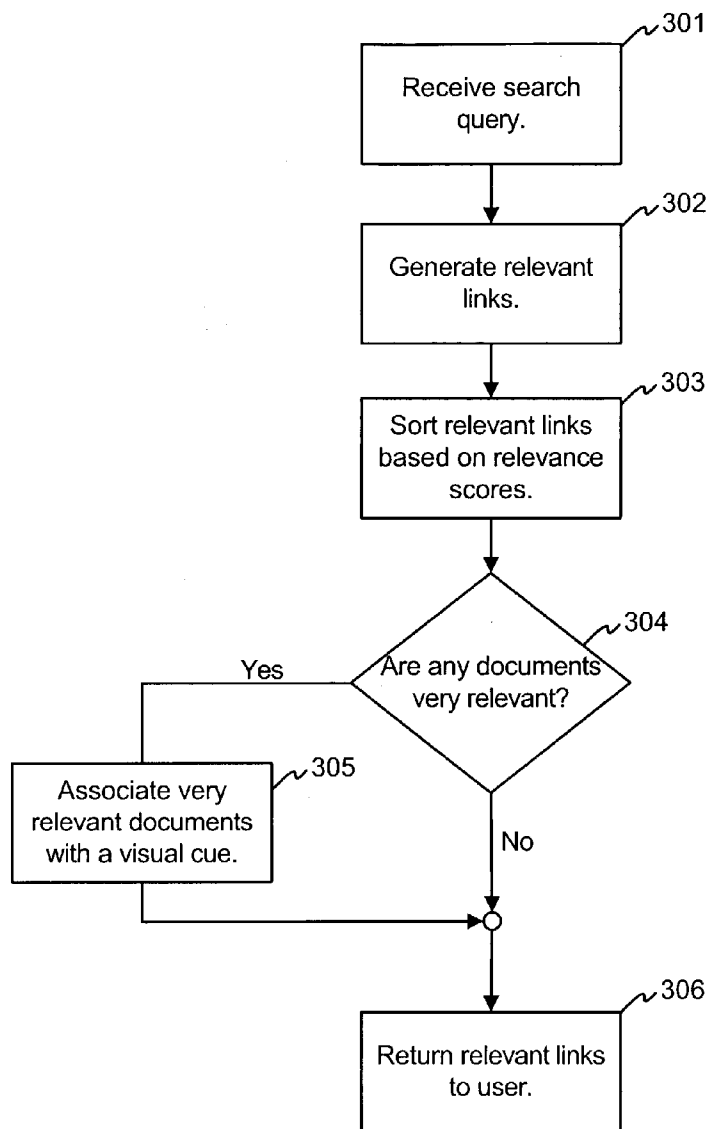
(22) Filed: **Jun. 10, 2003**

Publication Classification

(51) **Int. Cl.⁷ G06F 15/00**

(52) **U.S. Cl. 715/501.1**

A search engine includes a decision component that determines whether documents that are returned in response to a user search query are likely to be very relevant to the search query. Links that refer to documents that the search engine determines to likely be very relevant may be displayed with visual cues that assist the user in browsing the links. The decision component may base its decision on a number of parameters, including: (1) the position of the document in a ranked list of search results, (2) the click through rate of the document, (3) relevance scores for the document and other documents that are returned as hits in response to the search query, and (4) whether the document is classified as a pornographic document (the search engine may refrain from showing visual cues for potentially pornographic documents).



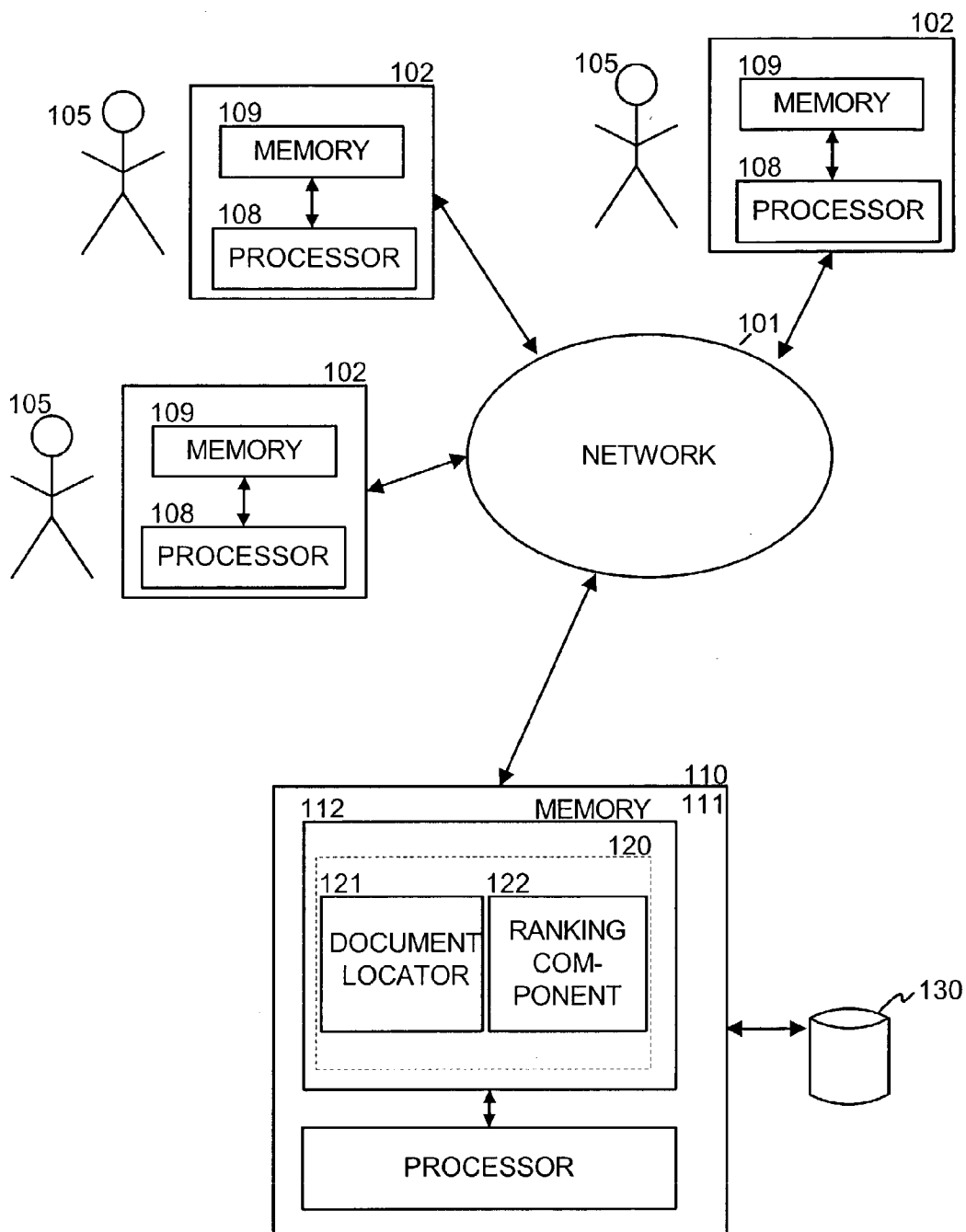



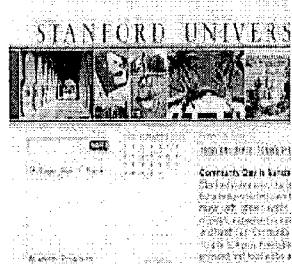
Fig. 1

[Advanced Search](#) [Preferences](#) [Language Tools](#) [Search Tips](#) 200


 Web · Images · Groups · Directory · News ·
 Searched the web for stanford. Results 1 - 10 of about 6,280,000. Search took 0.22 seconds.

Categories: [Regional > North America > ...](#) [Regional > North America > ... > South Bay > Education](#) ~ 215

News: [Stanford Super Regional notes](#) - ESPN - 1 hour ago
[Students end Stanford fast](#) - San Jose Mercury News - Jun 5, 2003
[Draft taps Cal. Stanford](#) - San Francisco Chronicle - Jun 4, 2003
 Try Google News: [Search news for stanford](#) or [browse the latest headlines](#) } 216


[Stanford University](#) ~ 210
 The Stanford University's homepage allows you to find web resources and websites. Pictures of Stanford University. Stanford Web People. ... Admission. About Stanford. } 220
 Description: [Stanford University](#)
 Category: [Regional > North America > ...](#) ~ 221
[www.stanford.edu/](#) - 38k - [Cached](#) - [Similar pages](#)

[The Martin Luther King, Jr. Papers Project](#) ~ 211
 Martin Luther King, Jr. Speaks Out on War and Peace. Nonviolence is not sterile passivity, but a powerful moral force which makes for social transformation. ...
 Description: [Secondary documents written about Martin Luther King, Jr. , as well as primary documents written during...](#)
 Category: [Society > History > ... > Personas > King, Martin Luther, Jr.](#)
[www.stanford.edu/group/King/](#) - 15k - [Cached](#) - [Similar pages](#)
[\[More results from www.stanford.edu \]](#)

[Stanford Encyclopedia of Philosophy](#) ~ 212
 Stanford Encyclopedia of Philosophy. Edited by Edward N. Zalta Editorial Information. [Link ... Principal Site: USA Stanford University. Mirror ...
 Description: [Online philosophy reference work, articles are authored and updated by experts in the field. Edited...](#)
 Category: [Society > Philosophy > ... > Stanford Encyclopedia of Philosophy](#)
[plato.stanford.edu/](#) - 3k - [Cached](#) - [Similar pages](#)

[Stanford Encyclopedia of Philosophy Table of Contents \(Abridged\)](#) ~ 213
 Stanford Encyclopedia of Philosophy Abridged Table of Contents. Welcome to the Stanford Encyclopedia of Philosophy, which was designed ...
 Description: [Dynamic, evolving philosophy reference. Articles include bibliographies for further reading.](#)
 Category: [Reference > Encyclopedias > Subject Encyclopedias > Philosophy](#)
[plato.stanford.edu/contents.html](#) - 101k - [Cached](#) - [Similar pages](#)
[\[More results from plato.stanford.edu \]](#)

[SUL: Copyright & Fair Use](#) ~ 214
 Welcome to fairuse.stanford.edu. ... Fair Use Copyright Reminder from Stanford Provost Condoleezza Rice, October 30, 1998 (superseded). ...
 Description: [A very good starting place built by the Stanford University.](#)
 Category: [Arts > Music > Resources > Copyright](#)
[fairuse.stanford.edu/](#) - 6k - [Cached](#) - [Similar pages](#)

Fig. 2

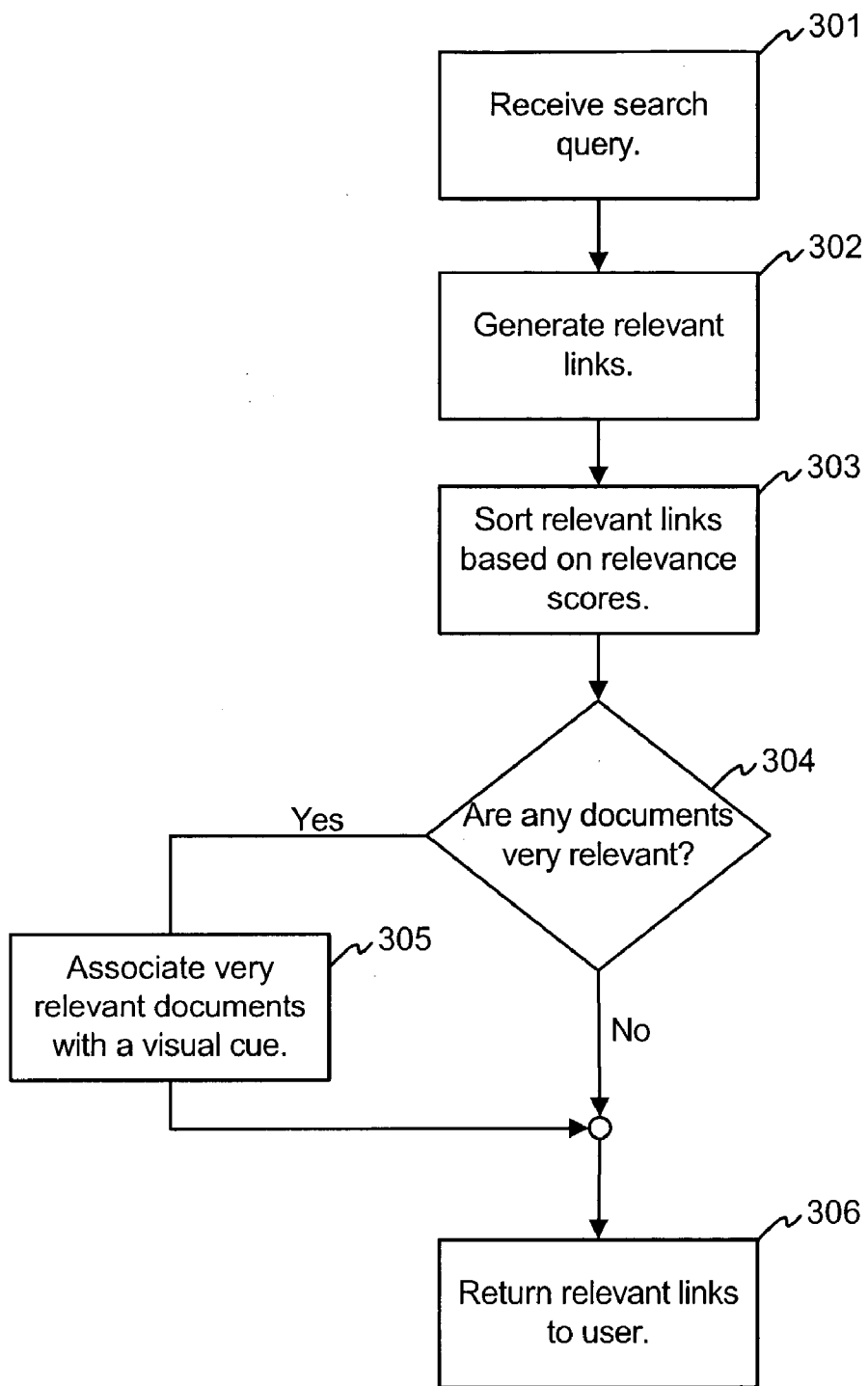


Fig. 3

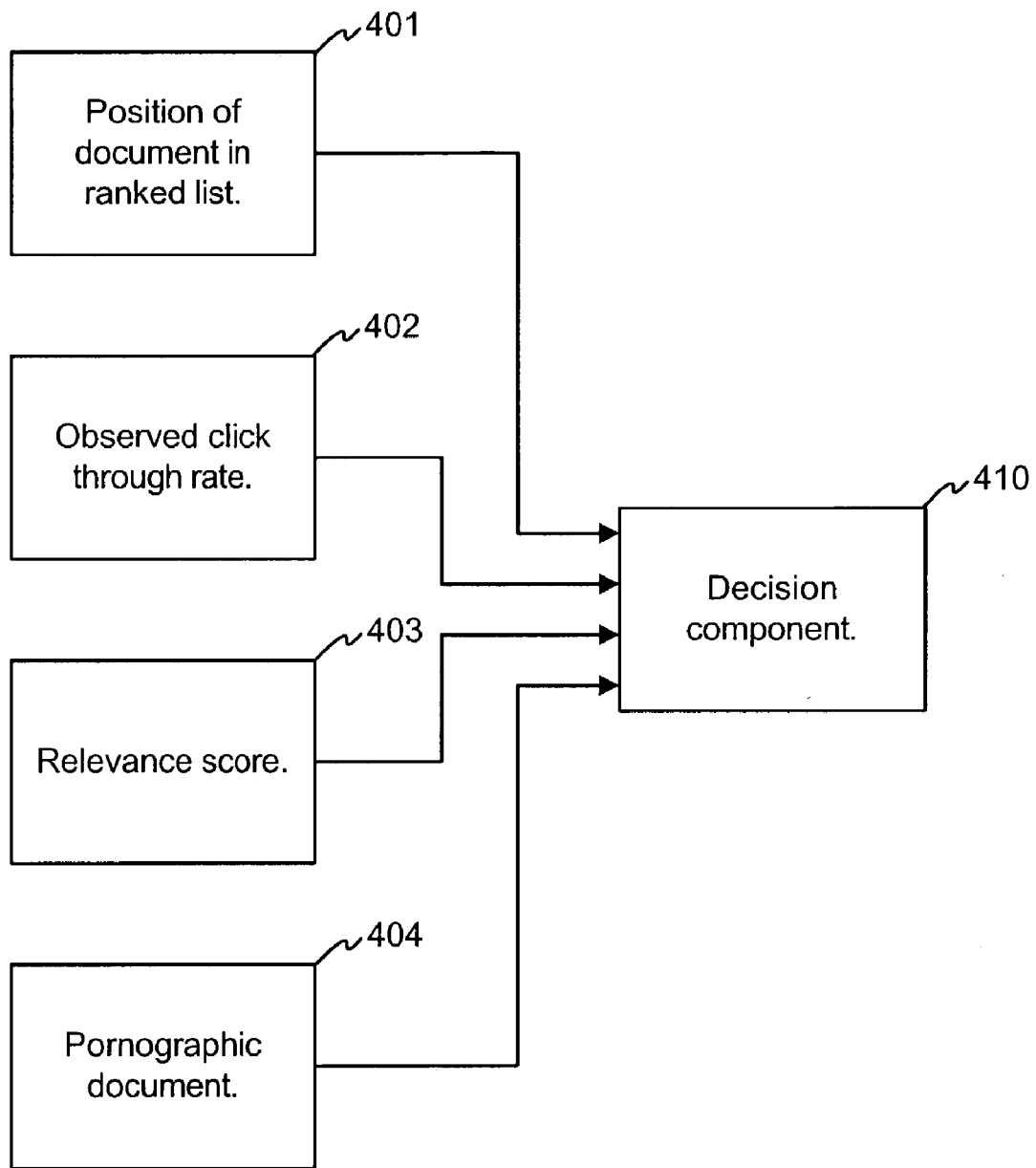


Fig. 4

DOCUMENT SEARCH ENGINE INCLUDING HIGHLIGHTING OF CONFIDENT RESULTS

BACKGROUND OF THE INVENTION

[0001] A. Field of the Invention

[0002] The present invention relates generally to document search engines and, more particularly, to highlighting select documents returned from a search engine.

[0003] B. Description of Related Art

[0004] The World Wide Web (“web”) contains a vast amount of information. Locating a desired portion of the information, however, can be challenging. This problem is compounded because the amount of information on the web and the number of new users inexperienced at web searching are growing rapidly.

[0005] Search engines attempt to return hyperlinks to web pages in which a user is interested. Generally, search engines base their determination of the user’s interest on search terms (called a search query) entered by the user. The goal of the search engine is to provide links to high quality, relevant results to the user based on the search query. Typically, the search engine accomplishes this by matching the terms in the search query to a corpus of pre-stored web pages. Web pages that contain the user’s search terms are “hits” and are returned to the user.

[0006] In an attempt to increase the relevancy and quality of the web pages returned to the user, a search engine may attempt to sort the list of hits so that the most relevant and/or highest quality pages are at the top of the list of hits returned to the user. For example, the search engine may assign a rank or score to each hit, where the score is designed to correspond to the relevance or importance of the web page. Existing techniques of determining relevance are based on the contents of the web page. More advanced techniques determine the importance of a web page based on more than the content of the web page. For example, one known method, described in the article entitled “The Anatomy of a Large-Scale Hypertextual Search Engine,” by Sergey Brin and Lawrence Page, assigns a degree of importance to a web page based on the link structure of the web page. In other words, the Brin and Page algorithm attempts to quantify the importance of a web page based on more than just the content of the web page.

[0007] The goal of a search engine is to return the most desirable set of links for any particular search query. However, in addition to just returning the most desirable set of links, it may also be beneficial to present the set of links to the user in a manner that clearly and quickly informs the user of the content associated with each of the links.

[0008] One method of apprising the user of the content associated with a particular link is to also display a “snippet” of text with the link. Ideally, the snippet of text should summarize the content of the link. In practice, the snippets are typically drawn from text of the document referenced by the link. Although text snippets can be useful in summarizing the link, the quality of the snippet in accurately summarizing the link can vary based on factors such as the automated nature of snippet generation and the quality of the underlying document. Also, for snippets to be useful, the user must take time to read the snippet.

[0009] There is a need in the art for improved techniques for apprising the user of the contents of links returned in response to a search query.

SUMMARY OF THE INVENTION

[0010] A search engine returns a list of links to documents that are relevant to the search query. Links that refer to documents that the search engine determines to be very relevant to a user query may be displayed with visual cues that assist the user in browsing the links.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate the invention and, together with the description, explain the invention. In the drawings,

[0012] FIG. 1 is a diagram illustrating an exemplary system in which concepts consistent with the present invention may be implemented;

[0013] FIG. 2 is a diagram illustrating a document that includes links to web documents displayed in response to a search query;

[0014] FIG. 3 is a flow chart illustrating operation of a search engine consistent with an aspect of the invention; and

[0015] FIG. 4 is a block diagram conceptually illustrating the determination of whether a document is very relevant to a user search query.

DETAILED DESCRIPTION

[0016] The following detailed description of the invention refers to the accompanying drawings. The detailed description does not limit the invention. Instead, the scope of the invention is defined by the appended claims and equivalents.

[0017] As described herein, a search engine may determine that it is confident that a particular link is the link that the user intended to find. The search engine may then highlight this link with a visual cue. The visual cue allows the user to quickly locate and determine whether the highlighted link is the link that the user intended to find.

System Overview

[0018] FIG. 1 is a diagram illustrating an exemplary system in which concepts consistent with the present invention may be implemented. The system includes multiple client devices 102, a server device 110, and a network 101, which may be, for example, the internet. Client devices 102 each include a computer-readable medium 109, such as random access memory, coupled to a processor 108. Processor 108 executes program instructions stored in memory 109. Client devices 102 may also include a number of additional external or internal devices, such as, without limitation, a mouse, a CD-ROM, a keyboard, and a display.

[0019] Through client devices 102, users 105 can communicate over network 101 with each other and with other systems and devices coupled to network 101, such as server device 110.

[0020] Similar to client devices 102, server device 110 may include a processor 111 coupled to a computer-readable

memory **112**. Server device **110** may additionally include a secondary storage element, such as database **130**.

[0021] Client processors **108** and server processor **111** can be any of a number of well known computer processors. In general, client device **102** may be any type of computing platform connected to a network and that interacts with application programs, such as a digital assistant or a “smart” cellular telephone or pager. Server **110**, although depicted as a single computer system, may be implemented as a network of computer processors.

[0022] Memory **112** contains a search engine program **120**. Search engine program **120** locates relevant information in response to search queries from users **105**. In particular, users **105** send search queries to server device **110**, which responds by returning a list of relevant information to user **105**. Typically, users **105** ask server device **110** to locate web pages relating to a particular topic and stored at other devices or systems connected to network **101**. Search engine **120** includes document locator **121** and a ranking component **122**. In general, document locator **121** finds a set of documents whose contents match a user search query. Ranking component **122** may rank the located set of documents based on relevance and may generate a relevance score for each document that indicates a level of relevance. Search engine **120** may then return a list of links pointing to the set of documents determined by document locator **121**. The list of links may be sorted based on the relevance scores determined by ranking component **122**.

[0023] Document locator **121** may initially locate documents from a document corpus stored in database **130** by comparing the terms in the user’s search query to the documents in the corpus. In general, processes for indexing web documents and searching the indexed corpus of web documents to return a set of documents containing the searched terms are well known in the art. Accordingly, this functionality of document locator **121** will not be described further herein.

[0024] Ranking component **122** assists search engine **120** in returning relevant documents to the user by ranking the set of documents identified by document locator **121**. This ranking may take the form of assigning a numerical value, called a relevance score, corresponding to the calculated relevance of each document identified by document locator **121**. There are a number of suitable ranking algorithms known in the art, one of which is described in the article by Brin and Page, as mentioned in the Background of the Invention section of this disclosure. Alternatively, the functions of ranking component **122** and document locator **121** may be combined so that document locator **121** produces a set of relevant documents each having rank values.

[0025] As used herein, a document is to be broadly interpreted to include any machine-readable and machine-storable work product. A document may be an email, a file, a combination of files, one or more files with embedded links to other files, etc. In the context of the Internet, a common document is a Web page. Web pages often include content and may include embedded information (such as meta information, hyperlinks, etc.) and/or embedded instructions (such as Javascript, etc.).

Operation of Search Engine **120**

[0026] Consistent with an aspect of the invention, links to sets of web documents returned by search engine **120** may

include, in addition to text snippets that describe the web documents, a visual cue that informs the user that the web document is likely to be relevant to the user’s search query. The visual cue may include graphic information through which the user can quickly and easily verify whether the link is the link that the user intended to locate. In one aspect of the invention, the link corresponding to a document that is determined to be “highly relevant” (i.e., a high confidence that the document is the document that the user would be most interested in viewing) to the user search query is displayed with the visual cue.

[0027] FIG. 2 is a diagram illustrating a document **200** that includes links to web documents that may be displayed to a user at a client device **102** in response to a search query. The document may be generated by search program **120** in response to the user search query and displayed by a web browser at one of client devices **102**.

[0028] In the exemplary document shown in FIG. 2, a search query **201** entered by the user was “stanford.” Document **200** includes a number of links **210, 211, 212, 213,** and **214** to web documents, such as web pages, that search engine **120** determined to be potentially relevant to search query **201**. In addition to the returned links **210-214**, document **200** may include other information, such as the general category **215** and current events/news links **216** that correspond to search query **201**.

[0029] The user may select any of links **210-214** to thereby direct the web browser to return the web document pointed-to by the links. Each of links **210-214** may also include, for example, a snippet of text **220** that describes the linked web document and a description of a category **221** in which the web document is classified.

[0030] Consistent with an aspect of the invention, one or more of the links **210-214** may contain a visual cue **230** corresponding to the link. In the example shown in FIG. 2, visual cue **230** is a miniaturized (“thumbnail”) rendering of the web page corresponding to link **210**. In this example, search query **201** was “stanford.” Search engine **120** determined that the most highly ranked link for “stanford” is the link to the web site of Stanford University (stanford.edu). Accordingly, search engine **120** included visual cue **230** in document **200**.

[0031] In some implementations, search engine **120** may only display visual cues **230** when a link is determined to be “very relevant” to the search query, where “very relevant” is defined by an objective standard for determining relevancy. For example, assume that the large majority of “stanford” search queries eventually result in the user selecting the stanford.edu link. In this situation, search engine **120** may realize that stanford.edu is a very relevant link and therefore include visual cue **230** in document **200**.

[0032] By including visual cue **230** with very relevant links, users may learn to associate the visual cue with links that search engine **120** is confident matches the user’s intentions. As the users begin to trust visual cue **230**, the visual cue allows the user to home in on the relevant search results faster as they will not necessarily need to read the corresponding snippet **220**. This may thus result in a decreased “time to satisfaction” for the user and a concomitant increase in search engine satisfaction.

[0033] FIG. 3 is a flow chart illustrating operation of search engine program **120** consistent with an aspect of the

invention. Search engine program 120 may begin by receiving a search query from one of users 105 (act 301). Based on the search query, document locator 121 may generate a set of links to documents that are relevant to the search query (act 302). The set of links may be sorted based on a relevance metric returned for each of the documents from ranking component 122 (act 303).

[0034] Search engine program 120 may determine whether any of the links returned by document locator 121 are associated with “very relevant” documents (act 304). Documents that are determined to be very relevant may be associated with a visual cue, such as visual cue 230 (act 305). Server device 110 may then return the set of links to the user that entered the search query (act 306). The returned set of links may be transmitted as a web document formatted in a manner similar to document 200, and may particularly include visual cues 230 associated with the very relevant documents.

[0035] FIG. 4 is a block diagram conceptually illustrating the determination of whether a document is very relevant by search engine program 120.

[0036] Whether a particular document, D, is very relevant may depend on a number of parameters. In one implementation, these parameters may be calculated by search engine program 120 and are shown in FIG. 4 as components 401-404. These components include the position of the document D in the ranked list of documents (component 401), the observed historical user click through rate compared to the other documents in the ranked list of documents (component 402), the relevance score for the document relative to the relevance scores for the other documents in the ranked list of documents (component 403), and whether the document is detected as being a pornographic document (component 404).

[0037] Component 401 may be determined directly based on the sorted set of links returned from ranking component 122. In general, a document at the top of the sorted list is more likely to be a very relevant document than a document further down on the list.

[0038] Component 402 may be based on a historical click through rate of other users in response to the links in the returned set of links. For example, search engine program 120 may calculate a ratio between document D and each other document in the returned set of links. The ratio may be calculated, for each other document E, as:

$$\frac{\hat{D}}{\hat{E}}, \tag{1}$$

[0039] where \hat{D} represents the total number of times users have selected document D and \hat{E} represents the total number of times users have selected document E. Thus, for each document E, the ratio gives the popularity of document D relative to document E. In an alternate implementation of formula (1) for measuring historical click through rates, \hat{D} may be defined to represent the total number of times users have selected document D for a particular query Q and \hat{E} may be defined as the number of times users clicked on any document for query Q. Component 403 may relate the

relevance score for document D, as returned from ranking component 122, to the relevance score of each document E. For example, component 403 may store the differences between the relevance scores of document D and the relevance scores of the other documents.

[0040] Component 404 may store determinations of whether the document D is a pornographic document. This determination can be made based on, for example, key word comparison, domain name matching, manually entered information, etc.

[0041] Based on one or more of the parameters stored by components 401-404, decision component 410 determines whether to classify the document D as very relevant. This classification may be based on a number of functions that use the parameters stored by components 401-404. In one implementation, decision component 410 determines document D is very relevant when: (1) document D’s position is first in the ranked list of results (determined based on component 401), (2) document D has a high click through rate, as may be determined by a pre-selected click through threshold (determined based on component 402), and (3) the relevance score for document D is significantly greater than the next highest relevance score in the returned set of documents (determined based on component 403). In some implementations, decision component 410 may additionally require that (4) document D is not classified as a pornographic document (determined based on component 404). The specific threshold values for the high click through rate (2) and the differences between relevance scores (3) may be determined by one of ordinary skill in the art via trial and error techniques.

[0042] Other functions, based on one or more of components 401-404, may be used by decision component 410 to determine whether a document is very relevant. For example, there may be a tradeoff between how high the click through rate for document D should be and how high the relevance score for document D should be. A high relevance score for document D is generally a good indication that document D is a good result. Accordingly, when document D has a high relevance score and is in the first position in the ranked list, decision component 410 may accept a moderate click through rate for document D and still classify the document as very relevant. On the other hand, if the relevance score for document D is low, decision component 410 may require a much higher click through rate and that the document is in the first position before classifying the document as very relevant. A similar tradeoff can be made with the click through rate and the difference in relevance scores between document D and the second place document in the ranked list. In general, one of ordinary skill in the art will recognize that other functions based on the parameters associated with components 401-404, or based on other parameters, can be used to determine whether a document is highly relevant.

[0043] Further, one of ordinary skill in the art will recognize that a number of modifications are possible to the above-discussed techniques for displaying visual cues for select documents. For example, instead of rendering the thumbnails in real-time, the thumbnails may be pre-rendered for documents that have a high click through rate. Thus, documents that are likely to be displayed with a visual cue can be quickly returned to the user without having to first

generate the thumbnail. Further, visual cues other than thumbnails may be presented to the users. For example, other forms of highlighting, such as a logo, contrasting textual fonts (e.g., text contrasted by size, color, or weight) that are designed to stand out, contrasting backgrounds, or textual labels may be used in place of thumbnails.

[0044] Additionally, parameters other than the above-described parameters may be used by decision component 410. For example, a parameter related to the click through rate may be defined as the historical ratio of the number of times that users have selected document D after submitting a query to the number of times that the query was submitted.

[0045] Yet another parameter that could be used by decision component 410 is based on the text of the hyperlinks that connect web documents. In particular, this parameter may be based on the number of links pointing to document D in which the text that describes the link matches the search query. This number may be compared to the number of links pointing to the other documents in the returned set of documents in which the text of the link matches the search query.

Summary

[0046] As described herein, a search engine includes a decision component that determines whether documents are highly relevant. A highly relevant document may indicate that the search engine is confident that the user who entered the search query will be interested in the document. Such documents may then be associated with a visual cue that draws the user's attention and assists the user in analyzing links returned in response to the search query.

[0047] It will be apparent to one of ordinary skill in the art that aspects of the invention, as described above, may be implemented in many different forms of software, firmware, and hardware in the implementations illustrated in the figures. The actual software code or specialized control hardware used to implement aspects consistent with the present invention is not limiting of the present invention. Thus, the operation and behavior of the aspects were described without reference to the specific software code—it

being understood that a person of ordinary skill in the art would be able to design software and control hardware to implement the aspects based on the description herein.

[0048] The foregoing description of preferred embodiments of the present invention provides illustration and description, but is not intended to be exhaustive or to limit the invention to the precise form disclosed. Modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention.

[0049] No element, act, or instruction used in the description of the present application should be construed as critical or essential to the invention unless explicitly described as such. Also, as used herein, the article "a" is intended to include one or more items. Where only one item is intended, the term "one" or similar language is used.

[0050] The scope of the invention is defined by the claims and their equivalents.

What is claimed:

1. A method for performing a document search comprising:
 - generating links to a set documents relevant to a search query submitted by a user;
 - determining whether one or more of the documents in the set of documents is likely to correspond to a document that the user considers to be very relevant to the search query;
 - associating a visual cue with the links to the documents in the set of documents that are determined to be very relevant; and
 - returning the generated links and the associated visual cues to the user.
2. The method of claim 1, wherein the documents are web documents.
3. The method of claim 2, wherein the visual cue is a thumbnail representation of the document associated with the visual cue.

* * * * *