

(19) World Intellectual Property Organization
International Bureau



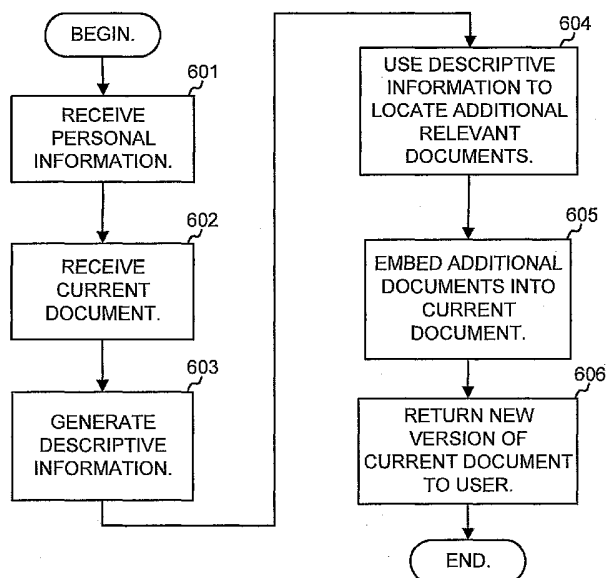
(43) International Publication Date
12 January 2006 (12.01.2006)

PCT

(10) International Publication Number
WO 2006/004657 A1

- (51) International Patent Classification⁷: **G06F 17/30**
- (21) International Application Number:
PCT/US2005/022726
- (22) International Filing Date: 29 June 2005 (29.06.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
10/887,443 30 June 2004 (30.06.2004) US
- (71) Applicant (for all designated States except US):
GOOGLE INC. [US/US]; 1600 Amphitheatre Parkway, Building 41, Mountain View, California 94043 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **BALUJA, Shumeet** [US/US]; 590 Penny Lane, Mountain View, California 94043 (US). **MITTAL, Vibhu** [US/US]; 541 Loch Lomond Ct., Sunnyvale, California 94087 (US). **SAHAMI, Mehran** [US/US]; 3238 Hoover Street, Redwood City, California 94063 (US).
- (74) Agent: **LEDELL, Brian E.**; Harrity & Snyder LLP, 11240 Waples Mill Road, Suite 300, Fairfax, VA 22030 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— with international search 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: ENHANCED DOCUMENT BROWSING WITH AUTOMATICALLY GENERATED LINKS BASED ON USER INFORMATION AND CONTEXT



(57) Abstract: Additional documents are automatically located that are relevant to an original document, such as a document being read by a user, and also potentially relevant to personal information of the user. The additional documents may be located based on descriptive information that includes personal information of the user and content information of the document being read. The additional documents, or links to the additional documents, may be incorporated into the document being read. In some implementations, the additional documents may be presented in-line with the document being read, such as through an in-link link or text snippet. The user can thus be efficiently presented with additional information that is relevant to the original document being read.

ENHANCED DOCUMENT BROWSING WITH
AUTOMATICALLY GENERATED LINKS BASED ON USER INFORMATION AND CONTEXT

BACKGROUND OF THE INVENTION

A. Field of the Invention

5 Systems and methods described herein relate generally to information retrieval and, more particularly, to techniques for browsing information.

B. Description of Related Art

 The World Wide Web ("web") contains a vast amount of information. One very common use of the web is to read documents, such as news articles or other publications.

10 When reading a particular document, such as a news article, it is known to provide links to other documents that are somehow related to the particular document. For example, when a user selects a news document from a news search engine or an online news service, the web site may provide links to other news articles or advertisements that are related to the news document. Typically, such related documents are determined based on the content of the document being read and are shown as
15 additional links displayed outside the content of the document. By providing convenient links to related material, these additional documents can enhance the browsing experience of the reader.

 It would be desirable to provide improved techniques for enhancing document browsing by providing automatically generated links to relevant information to the reader.

SUMMARY OF THE INVENTION

20 According to one aspect, a method of enhancing document browsing includes receiving personal information relating to a user, generating descriptive information based on a content of a first document and the personal information, and identifying additional documents based on the descriptive information. Additionally, a second document may be generated that includes at least a portion of the content of the first document modified to include references to the additional documents.

25 In another aspect, a method includes locating at least one second document that is relevant to a first document and embedding the second document within the first document at a location in the first document at which the second document has relevance.

 In another aspect, a method includes receiving a request for a first document from a user, identifying a named entity in the first document, locating a second document that is relevant to the named
30 entity, and presenting a modified version of the first document to the user in which a link to the second document is displayed in-line in the first document at a location in the first document proximate to the named entity to which the second document is relevant.

BRIEF DESCRIPTION OF THE DRAWINGS

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

 Figs. 1A and 1B are diagrams illustrating exemplary graphical interfaces that may be presented to a user;

 Fig. 2 is an exemplary diagram of a network in which concepts consistent with the principles of
40 the invention may be implemented;

Fig. 3 is an exemplary diagram of a client or server shown in the network of Fig. 2;

Fig. 4 is a block diagram illustrating conceptual elements of the document locator shown in Fig.

2;

Fig. 5 is a diagram illustrating an exemplary implementation of the search component shown in

Fig. 4;

Fig. 6 is a flowchart illustrating exemplary operations performed by the document locator shown in Fig. 2; and

Fig. 7 is a diagram illustrating an exemplary implementation of the document locator in the context of a content serving web site.

DETAILED DESCRIPTION

The following detailed description of the invention refers to the accompanying drawings. The detailed description does not limit the invention.

OVERVIEW

As described herein, additional documents relevant to an original document, such as a document being read by a user, are automatically located. The additional documents can be located based on their content and/or based on personal information of the user. The additional documents can be displayed in-line with the original document. The user can thus be efficiently presented with additional information that is relevant to the original document being read.

Figs. 1A and 1B are diagrams illustrating exemplary graphical interfaces that may be presented to a user. The graphical interfaces may be presented via a web browser 100 that is being used to browse the web. The example document 105 shown in Figs. 1A and 1B relates to a hiker's (Bill Cross) effort to climb Mt. Everest.

A number of additional documents may be relevant to document 105. In Fig. 1A, for example, links to three additional articles 110, 112, and 114 are embedded within document 105. Link 110 may reference a document about Mt. Everest, link 112 may reference a document about the Novolog Peaks and Poles Challenge, and link 114 may reference a document about diabetes. Each of links 110, 112, and 114 reference content that is in some way related to original document 105. In this example, links 110, 112, and 114 are displayed with brief summary text (e.g., "related content: Mt. Everest") that informs the reader of the content of the underlying link. Additionally, the summary text is underlined, indicating that the summary text is associated with a link. Assume that the reader of document 105 in Fig. 1A is located in San Jose, California. An advertisement 115 may additionally be displayed for a hiking equipment retailer in San Jose. Additionally, the documents referenced by links 110, 112, and 114 may be documents that are particularly appropriate for a reader in the San Jose area.

Although not shown in Fig. 1A, other links, such as links even more directly customized to the personal information of the reader, may also be displayed. For example, if the reader has previously entered search queries to a search engine, such as photography related search queries, the other links may be links to documents describing "Everest Photography."

Document 105 in Fig. 1B is identical to that in Fig. 1A. Multiple links 120, 122, and 124 are included in document 105 of Fig. 1B. In this example, links 120, 122, and 124, instead of being shown as linked summary text, are implemented by simply modifying the formatting or display associated with certain words or phrases in document 105. For instance, link 120 is shown to the reader by underlining "Mt. Everest," thus illustrating to the reader that the link references a document that pertains to Mt.

Everest. Another link 126 is inserted in-line in document 105 that includes summary text similar to links 110, 112, and 114. Assume, for this example, that the reader is from Seattle instead of San Jose. Link 126, which may be generated based on this fact, references a document about hiking Mt. Rainer—a mountain near Seattle.

EXEMPLARY NETWORK OVERVIEW

Fig. 2 is an exemplary diagram of a network 200 in which concepts consistent with the principles of the invention may be implemented. Network 200 may include multiple clients 210 connected to a server 220 via a network 240. Network 240 may include a local area network (LAN), a wide area network (WAN), a telephone network, such as the Public Switched Telephone Network (PSTN), an intranet, the Internet, or a combination of networks. Two clients 210 and one server 220 have been illustrated as connected to network 240 for simplicity. In practice, there may be more clients and/or servers. Also, in some instances, a client may perform one or more functions of a server and a server may perform one or more functions of a client.

A client 210 may include a device such as a wireless telephone, a personal computer, a personal digital assistant (PDA), a lap top, or another type of computation or communication device, a thread or process running on one of these devices, and/or an object executable by one of these devices. Server 220 may include a server device that processes, searches, and/or maintains documents and images in a manner consistent with the principles of the invention. Clients 210 and server 220 may connect to network 240 via wired, wireless, or optical connections.

Server 220 may include additional document locator component 225 (also called simply "document locator 225" herein). Document locator 225 may locate and add references to other documents related to an input document, such as the references added to document 105 (Figs. 1A and 1B).

A document, as the term is used herein, is to be broadly interpreted to include any machine-readable and machine-storable work product. A document may be an e-mail, a web log (blog), a file, a combination of files, one or more files with embedded links to other files, a news group posting, etc. In the context of the Internet, a common document is a web page, such as an HTML 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.). Documents discussed herein generally include embedded images. A "link" as the term is used herein is to be broadly interpreted to include any reference to/from a document from/to another document or another part of the same document.

EXEMPLARY CLIENT/SERVER ARCHITECTURE

Fig. 3 is an exemplary diagram of a client 210 or server 220. Client/server 210/220 may include a bus 310, a processor 320, a main memory 330, a read only memory (ROM) 340, a storage device 350, an input device 360, an output device 370, and a communication interface 380. Bus 310 may include conductors that permit communication among the components of client/server 210/220.

Processor 320 may include conventional processors, microprocessors, or processing logic that interpret and execute instructions. Main memory 330 may include a random access memory (RAM) or another type of dynamic storage device that stores information and instructions for execution by processor 320. ROM 340 may include a conventional ROM device or another type of static storage device that stores static information and instructions for use by processor 320. Storage device 350 may include a magnetic and/or optical recording medium and its corresponding drive.

Input device 360 may include one or more conventional mechanisms that permit a user to input information to client/server 210/220, such as a keyboard, a mouse, a pen, voice recognition and/or biometric mechanisms, etc. Output device 370 may include one or more conventional mechanisms that output information to the user, including a display, a printer, a speaker, etc. Communication interface 380 may include any transceiver-like mechanism that enables client/server 210/220 to communicate with other devices and/or systems. For example, communication interface 380 may include mechanisms for communicating with another device or system via a network, such as network 240.

Server 220, consistent with the principles of the invention, may implement additional document locator 225. Additional document locator 225 may be stored in a computer-readable medium, such as memory 330. A computer-readable medium may be defined as one or more physical or logical memory devices and/or carrier waves.

The software instructions defining additional document locator 225 may be read into memory 330 from another computer-readable medium, such as data storage device 350, or from another device via communication interface 380. The software instructions contained in memory 330 may cause processor 320 to perform processes that will be described later. Alternatively, hardwired circuitry or other logic may be used in place of, or in combination with, software instructions to implement processes consistent with the invention. Thus, implementations consistent with the principles of the invention are not limited to any specific combination of hardware circuitry and software.

DOCUMENT LOCATOR 225

Fig. 4 is a block diagram illustrating conceptual elements of document locator 225. Document locator 225 may include a descriptive information generator 405, search component 410, and formatting component 415. Descriptive information generator 405 may generate descriptive information describing the current document and based on personal information of the user. In one implementation, the descriptive information may include a search query. Descriptive information generator 405 may generate the descriptive information based on the personal information of the user and/or the current input document (or information relating to the current document).

The descriptive information output from descriptive information generator 405 may be input to search component 410, which may use the descriptive information to generate additional documents. Links or other references to the additional documents may be inserted with the original document by formatting component 415.

Descriptive information generator 405, search component 410, and formatting component 415 will each be described in more detail below.

DESCRIPTIVE INFORMATION GENERATOR 405

As mentioned, descriptive information generator 405 may generate descriptive information, such as a search query. The descriptive information may be generally based on information relating to the document the user is currently viewing (or has requested for viewing) as well as personal information of the user. The information relating to the current document may include information based on the text of the current document. The text may be processed to obtain, for example: (1) all terms appearing more than some pre-determined number of times, (2) named entities that can be automatically extracted; (3) dates in the document; (4) author and publication names; and/or keyword or category extraction.

Regarding (1) above, terms appearing more than some pre-determined number of times may be considered to be important or particularly descriptive terms in the document and may be considered to

be descriptive information for the document. The number of terms selected to include in the descriptive information may, for example, be limited to a predetermined number of most frequently occurring terms. In a possible variation of this concept, the number of times a term occurs may be considered in conjunction with the general frequency with which the term appears in the language of the document.

5 Thus, terms that tend to occur relatively rarely in a language may be selected before a common term that occurs more times in the document.

A list of predetermined named entities or other nouns may be stored by descriptive information generator 405. For example, location names, celebrity names, names of well known commercial or consumer products, and company names may be pre-generated by either manual (i.e., entered by a
10 human operator) or automatic techniques. As mentioned above, the text of the document can be compared to these named entities and matches included in the descriptive information for the document. Referring to the examples of Figs. 1A and 1B, the list of predetermined named entities may have included terms such as "Mt. Everest" and "Novolog Peaks and Poles Challenge," causing these terms to be included in the descriptive information for document 105.

15 Dates in the document (item (3) above), the document author, and publication names (item (4)) may be included in the descriptive information. Such information can often be automatically determined through pattern matching techniques applied to the document. The date of the document may be used to locate other contemporaneously published documents. Similarly, the publishing entity (e.g., web site) and document author may be used to locate documents from the same or similar publications or
20 documents written by the same author. The document date, author, and publication may be particularly useful in the context of news stories. Regarding (5), a document can be analyzed for its keywords, such as keywords extracted based on term frequency or through named entity extraction.

In addition to generating the descriptive information based on the document, descriptive information generator 405 may generate descriptive information based on information specific to the user
25 ("personal information"). The personal information can include, for example, the geographical location of the user (e.g., previous search queries submitted or links selected), personal information provided by the user when registering an account, personal information based on the browsing history of the user, personal information extracted from documents generated by the user, or other sources of personal information. The geographic location of the user may be estimated based on the user's IP address. The
30 personal information can also include temporal information, such as the current date or season. Temporal information can be useful to correlate events with personal preferences or document content. For example, if a document being browsed is about Edinburgh, and the current month is July or August, then related documents about the Edinburgh Arts Festival may be shown.

In one implementation, the personal information can be based on user profiles constructed
35 from previous search queries submitted to a search engine. Category matching techniques can be used to infer user interests from search terms. For example, even if the user never actually enters the search term "photography," but instead queried the terms "Nikon," "aperture," and "f-stop," these terms may be used to infer that the user is interested in photography.

One technique for generating category mappings from search queries is based on gathering a
40 large number of historical user search queries that are labeled based on user search sessions. The rationale is that people that search for a search term such as "Canon" are likely to also, in the same search session, enter other search queries, such as "photography" or "f-stop," that are related to the

same category. By analyzing many such search query-sessions, category inferences can be made (e.g., if a person searches for "Nikon" it is likely that they are interested in photography).

Descriptive information generator 405 may format the descriptive information as a search query. In one implementation, the search query may be obtained by concatenating the descriptive information (e.g., the personal information of the user and the document-related descriptive information) to obtain the search query. As an example, consider document 105 in Fig. 1A. Based on an analysis of the document and the personal information of the user, descriptive information generator 405 may generate the descriptive information "Mt. Everest," "Novolog Peaks and Poles Challenge," "diabetes," "San Jose," and "photography." These terms may be combined into a single search query "Mt. Everest Novolog Peaks Poles Challenge diabetes San Jose photography." In other implementations, multiple search queries, each including a subset of terms from the document and the user's personal information, may be generated, such as the search queries: "Mt. Everest San Jose," "hiking San Jose," "photography Mt. Everest," etc.

One of ordinary skill in the art will recognize that other techniques for formulating the search queries from the generated descriptive information can be used. For example, additional information can be used in determining whether to include a term in the query, such as the general frequencies of the occurrence of the term in the language. Additionally, certain names, entities or other predefined terms can be given additional weight in determining whether to include them in the query. Some terms, such as geographic names, can be weighted differently than other terms, such as product names. Product names can be automatically limited by appending their associated company name after the product name. Still additionally, the descriptive information can be used with cluster or category matching techniques, such as those described above, to generate other terms that can be used in the search queries.

SEARCH COMPONENT 410

Fig. 5 is a diagram illustrating an exemplary implementation of search component 410 in additional detail. Search component 410 may include a search engine 505 and a ranking component 510.

Search engine 505 may receive the descriptive information from descriptive information generator 405, and in response, locate one or more documents relevant to the descriptive information. Search engine 505 may be a query-based search engine that returns a ranked set of documents related to the input search query. Search engine 505 may be a general search engine, such as one based on all documents from a large collection, such as documents on the web, or a more specialized search engine, such as a news search engine. Techniques for implementing search engines are generally known in the art and will thus not be disclosed further herein.

Ranking component 510 may operate to rank and/or prune the set of documents returned by search engine 505. In one implementation, ranking component 510 may sort the set of returned documents based on a query match score that defines how well each document in the set of returned documents matches the search query. Documents that are a "better" match to the search query, such as documents that include multiple instances of terms in the search query, will tend to have higher relevance scores than documents that are not matched as well. Ranking component 510 may also sort the documents based on other measures of relevance or quality, such as based on a link-based

measurement of document quality. The top N sorted documents (e.g., $N=3$) may be selected by ranking component 510 for presentation to a user.

Other techniques for ranking or pruning the set of relevant documents may be used by ranking component 510. For example, documents may be selected that appear in multiple document sets corresponding to multiple related search queries, documents may be selected as those that are most recent, documents may be selected as those that are the most popular (e.g., based on the number of times the document link was selected). As other examples, documents from commercial sites may be explicitly excluded (or included).

In some implementations, multiple potential search queries can be received corresponding to the descriptive information, and the queries that return the "best" results can be used. The "best" results can be measured in a number of ways, such as by looking at objective ranking values corresponding to documents returned from a search engine in response to the potential search query. In addition, multiple different search engines could be used, such as a news search engine, a product search engine, or a general web-based search engine.

FORMATTING COMPONENT 415

Formatting component 415 may incorporate the additional documents located by search component 410 into the current document (i.e., the document currently being viewed by the user) or into a new document that includes the current document. The additional documents may be incorporated with the current document in a manner that informs the user that the documents are available without being overly disruptive to the user's reading of the current document.

In one implementation, formatting component 415 may insert links (e.g., hyperlinks) into the additional documents in-line with the text of the current document. When possible, the link to each additional document may be inserted in a section in the current document that is particularly relevant to the additional document. This concept is illustrated in Figs. 1A and 1B, in which links to related content, such as a link to a document about Mt. Everest, is inserted near the term "Mt. Everest" in document 105. Although links in Figs. 1A and 1B are shown as including parenthetical summary information and as links that are identified by modifying the display of words in the current input document, other techniques for graphically displaying the links can be used.

Techniques other than in-line hyperlinks may be used to embed the additional documents into the current document. For example, "float-over" text may be used that appears when the user positions the cursor over a particular word, image, or other object in the current document.

OPERATION OF DOCUMENT LOCATOR 225

Fig. 6 is a flowchart illustrating exemplary operations performed by document locator 225. Document locator 225 may begin operation in response to a user requesting a document, such as a request made from a web site or a search engine.

Document locator 225 may receive or locate personal information of the user (act 601). The personal information can include information such as, for example, the geographic location of the user, personal information provided by the user when registering an account (or at another time), personal information based on the browsing history of the user, or personal information extracted from documents generated by the user. Document locator also receives the current input document that the user is requesting (act 602).

Descriptive information relating to the input document may be generated (act 603). The descriptive information may, as previously discussed, be generated by descriptive information generator 405 and may include a search query that contains terms related to the current input document and the personal information of the user. The descriptive information may be used to locate additional relevant documents (act 604). As discussed, this may be performed by search component 410 submitting a search query to a search engine.

One or more of the additional relevant documents may be embedded or otherwise associated with the current input document (act 605). As shown in Figs. 1A and 1B, the additional relevant documents may be embedded in-line with the current input document. The modified version of the current input document, including the links to the additional relevant documents, may then be presented to the user (act 606).

EXEMPLARY IMPLEMENTATION OF DOCUMENT LOCATOR 225

Fig. 7 is a diagram illustrating an exemplary implementation of additional document locator 225 implemented in the context of a content serving web site, such as a web site devoted to articles about a particular hobby (e.g., automobiles). One of ordinary skill in the art will appreciate that document locator 225 could be implemented in a number of additional networked environments, such as in the general context of a news search engine or a more general search engine.

A number of users 705 may connect to content web site 710 over a network 715. The users may request particular documents from content web site 710. Before returning the requested document to the user, web site 710 may transmit the document (or information identifying the document), potentially along with personal information of the requesting user, to document locator 225. Document locator 225 may return its modified version of the requested document, as previously discussed, to web site 710, which may then forward the document to the user. In this manner, documents from web site 710 may be auto-augmented to enhance their desirability before returning them to the user.

Many variations on this example are possible. For instance, instead of document locator 225 returning the enhanced document to web site 710, web site 710 may simply redirect the user's document request to document locator 225, which may then return the enhanced document to the user.

CONCLUSION

Techniques for automatically locating additional documents relevant to an original document and/or to personal information of a user, such as a document being read by a user, were described herein. In one implementation, the additional documents were located based on personal information of the user as well as being based on content relevant to the document being read by the user. The additional documents can be presented in-line with the document being read, such as via links inserted at locations in the document that are particularly relevant to the additional document. The user can thus be efficiently presented with additional information that is relevant to the original document being read.

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 invention is not limiting of the 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.

The foregoing description of preferred embodiments of the 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. For example, although many of the operations described above were
5 described in a particular order, many of the operations are amenable to being performed simultaneously or in different orders to achieve the same or equivalent results.

No element, act, or instruction used in 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 potentially allow for one or more items. Further, the phrase "based on" is intended to
10 mean "based, at least in part, on" unless explicitly stated otherwise

What Is Claimed:

1. A method of enhancing document browsing comprising:
receiving personal information relating to a user;
generating descriptive information based on a content of a first document and the personal
information;
5 identifying an additional document based on the descriptive information; and
generating a second document that includes at least a portion of the content of the first document
and including references to the additional documents.
2. The method of claim 1, wherein at least one of the references to the additional
documents includes a link embedded in-line with the first document.
3. The method of claim 2, wherein the link includes text describing the at least one of the
references.
4. The method of claim 2, wherein the link includes float-over text.
5. The method of claim 1, wherein the descriptive information based on a content of the first
document includes named entities matched from a list of predetermined named entities.
6. The method of claim 5, wherein the list of predetermined named entities includes
location names and consumer products.
7. The method of claim 1, wherein the descriptive information based on a content of the first
document includes terms appearing more than a pre-determined number of times in the first document.
8. The method of claim 1, wherein the personal information includes a geographic location
of the user.
9. The method of claim 1, wherein the personal information includes information provided
by the user when registering an account.
10. The method of claim 1, wherein the personal information includes information extracted
from documents generated by the user.
11. The method of claim 1, wherein the personal information is based on a browsing history
of the user.
12. The method of claim 1, wherein the personal information includes temporal information.
13. The method of claim 1, further including:
formatting the descriptive information as a search query; and
locating the additional documents by submitting the search query to a search engine.

14. A system comprising:
means for receiving personal information relating to a user;
means for generating descriptive information based on a content of a first document and the
personal information;

5 means for locating additional documents based on the descriptive information; and
means for generating a second document that includes the content of the first document
including references to the additional documents.

15. The system of claim 14, further comprising:
means for formatting the descriptive information as a search query; and
means for locating the additional documents by submitting the search query to a search engine.

16. A method comprising:
locating at least one second document that is relevant to a first document; and
embedding the second document within the first document at a location in the first document at
which the second document has relevance.

17. The method of claim 16, wherein the relevance of the second document to the first
document is based on a content of the first document and on personal information of an intended reader
of the first document.

18. The method of claim 17, further comprising:
generating descriptive information based on the content of the first document and the personal
information; and
5 locating the at least one second document based on a comparison of the descriptive information
to a database of documents.

19. The method of claim 18, wherein the descriptive information includes named entities
matched from a list of predetermined named entities.

20. The method of claim 19, wherein the list of predetermined named entities includes
location names and consumer products.

21. The method of claim 18, wherein the descriptive information includes terms appearing
more than a pre-determined number of the times in the first document.

22. The method of claim 17, wherein the personal information includes a geographic location
of the intended reader.

23. The method of claim 17, wherein the personal information includes information provided
by the intended reader when registering an account.

24. The method of claim 17, wherein the personal information includes information extracted from documents generated by the intended reader.

25. The method of claim 17, wherein the personal information is based on a browsing history of the intended reader.

26. The method of claim 16, wherein the second document is embedded in the first document as a hyperlink associated with a named entity in the first document.

27. The method of claim 16, wherein embedding the second document within the first document at a location in the first document at which the second document has relevance further includes:

5 inserting a hyperlink that includes text describing the second document at the location.

28. A method comprising:

receiving a request for a first document from a user;

identifying a named entity in the first document;

locating a second document that is relevant to the named entity; and

5 presenting a modified version of the first document to the user in which a link to the second document is displayed in-line in the first document at a location in the first document proximate to the named entity to which the second document is relevant.

29. The method of claim 28, wherein locating the second document includes:

receiving personal information relating to a user;

generating descriptive information based on a content of the first document and the personal information; and

5 locating the second document based on the descriptive information.

30. The method of claim 28, wherein the link to the second document is displayed by modifying a display font of the named entity to indicate a hyperlink.

31. The method of claim 28, wherein the link to the second document is displayed by adding a hyperlink that includes text describing the second document at a location in the first document near the named entity.

32. A computer-readable medium containing instructions, that when executed by processing logic, cause the processing logic to:

receive a request for a first document from a user;

identify a named entity in the first document;

5 locate a second document that is relevant to the named entity; and

present a modified version of the first document to the user in which a link to the second document is displayed in-line in the first document at a location in the first document proximate to the named entity to which the second document is relevant.

33. A device comprising:

a memory containing programming instructions; and

a processor coupled to the memory and configured to execute the programming instructions to:

receive personal information relating to a user;

5 generate descriptive information based on a content of the first document and the personal information;

locate at least one additional document based on the descriptive information; and

generate a second document that includes the content of the first document modified to include references to the at least one additional document.

34. A method of enhancing a first document comprising:

receiving personal information relating to a user;

generating descriptive information based on a content of the first document and the personal information;

5 formatting the descriptive information as a search query; and

locating additional documents by submitting the search query to a search engine; and

generating a second document that includes the content of the first document modified to include a link embedded in-line with the first document that references at least one of the additional documents.

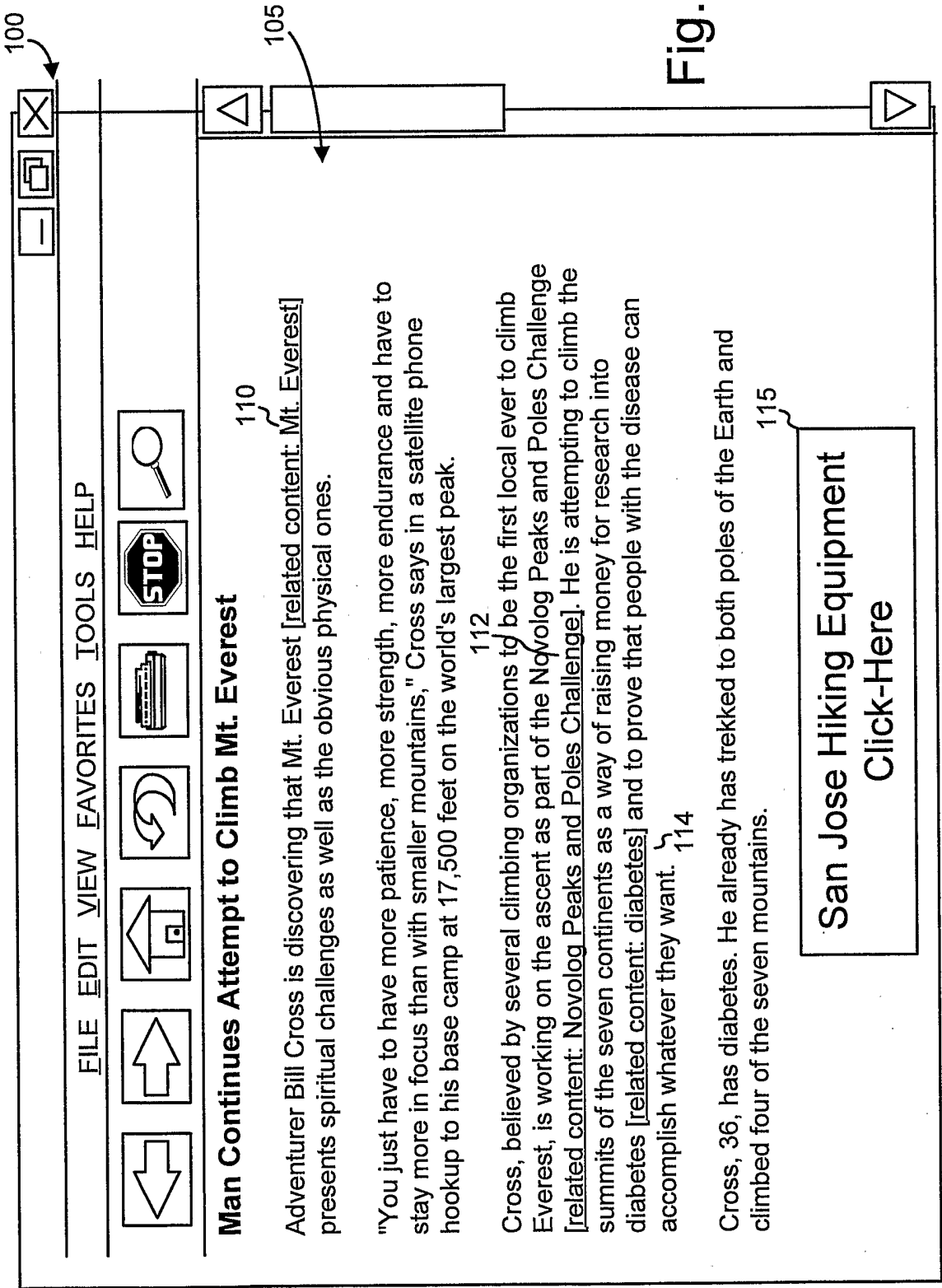


Fig. 1A

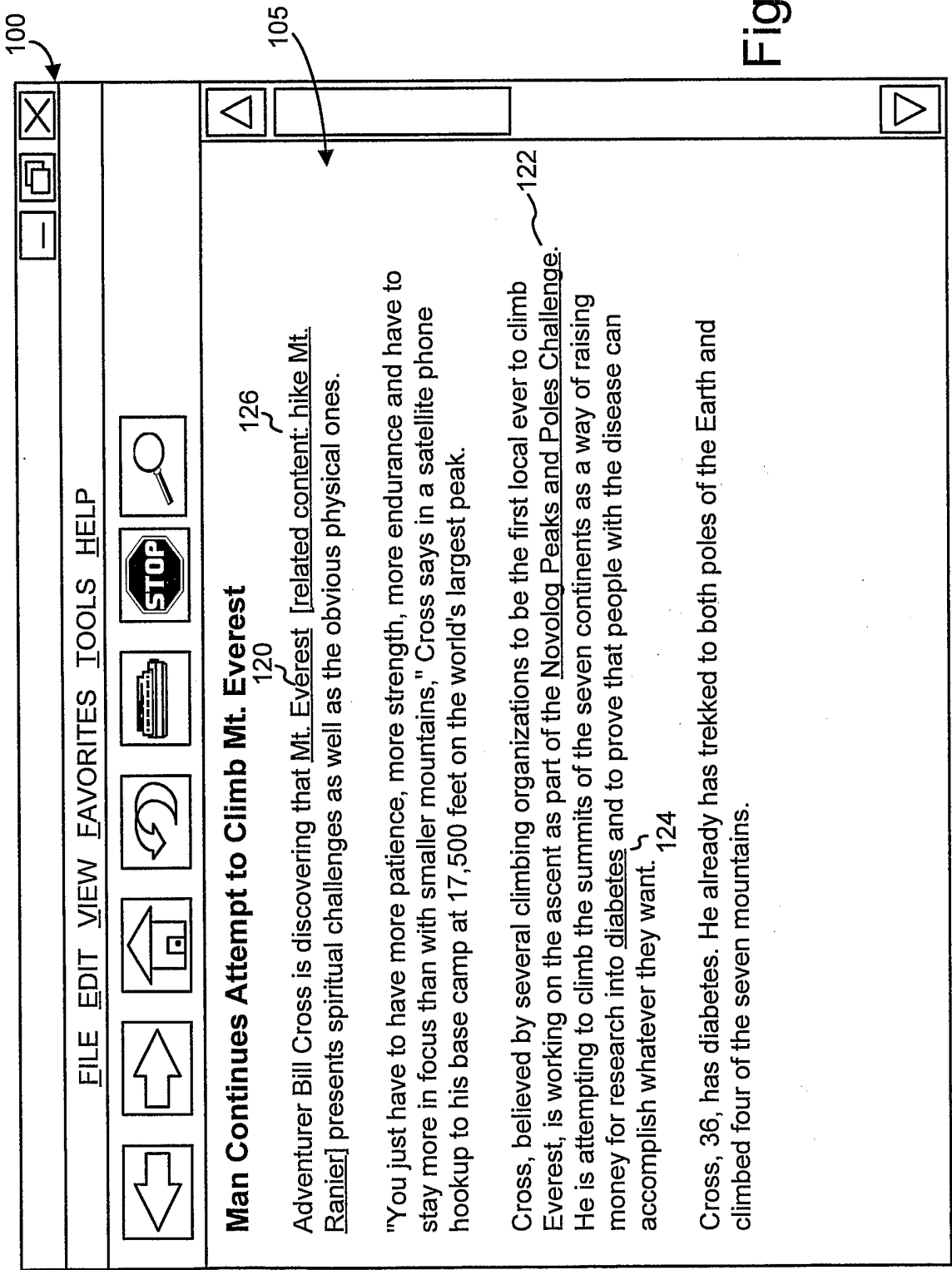


Fig. 1B

3/8

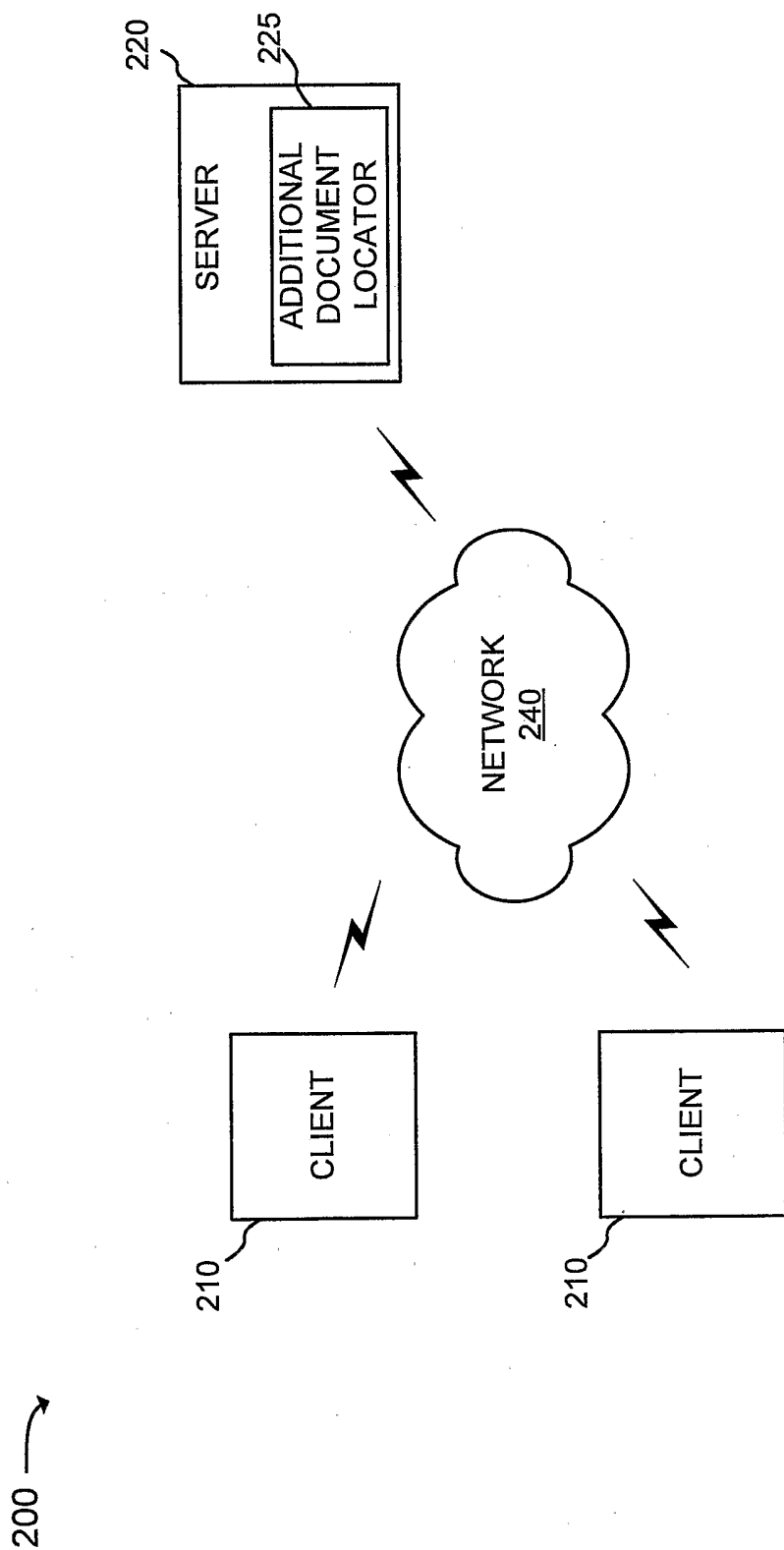


Fig. 2

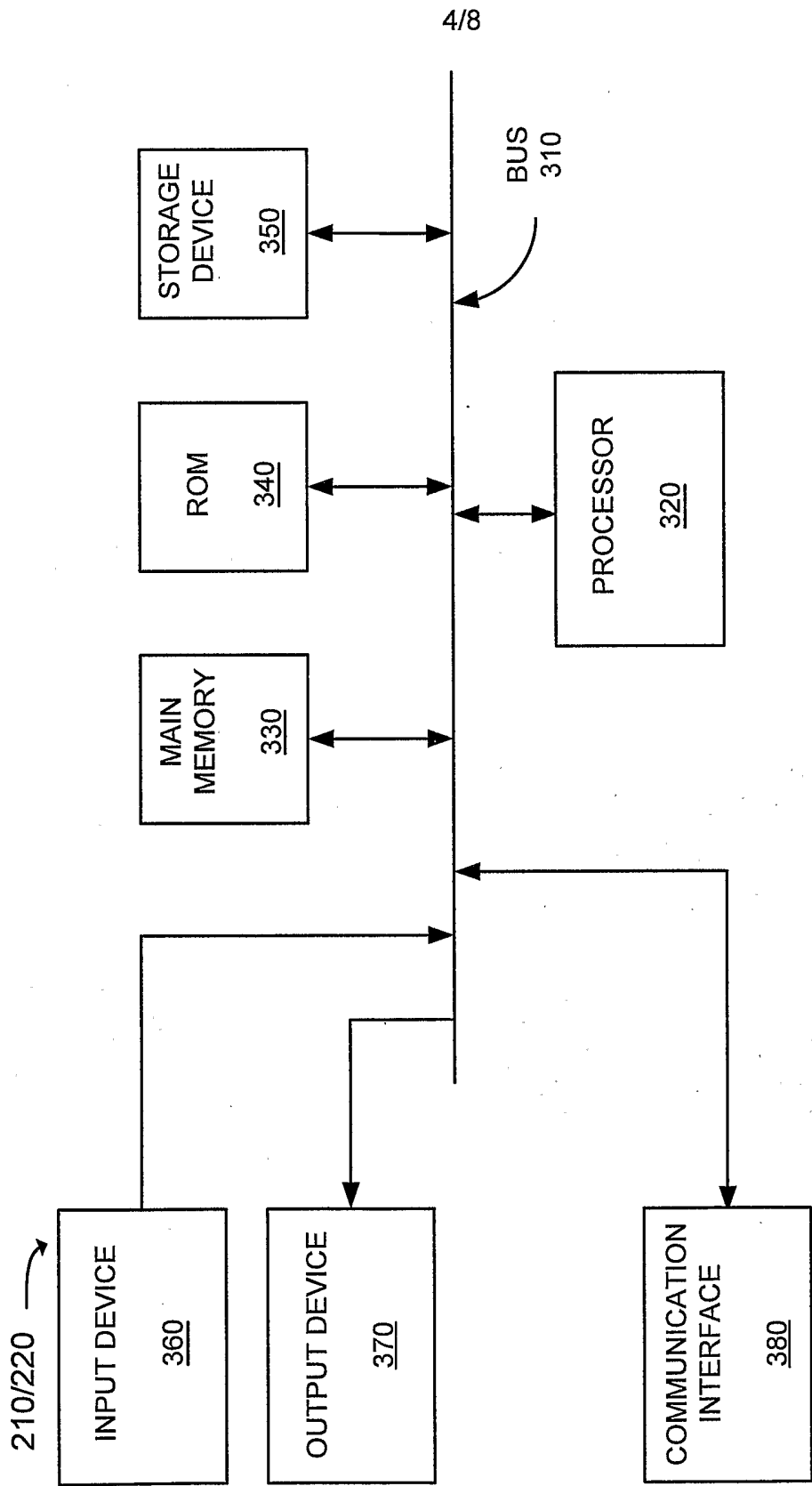


Fig. 3

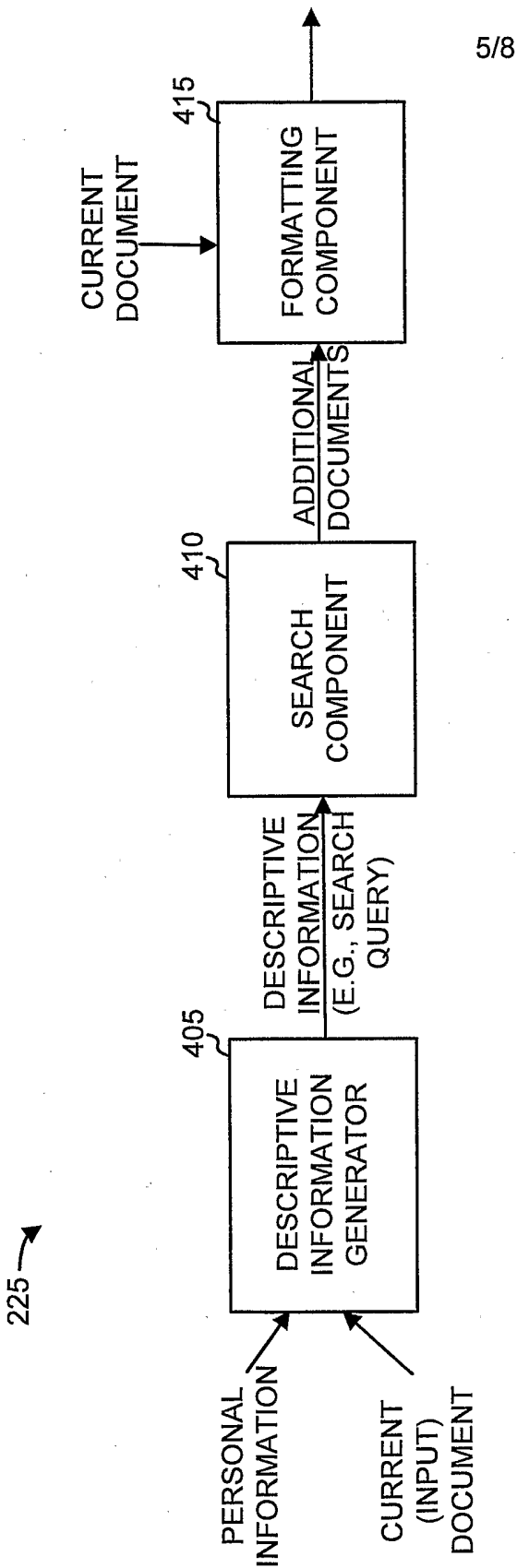


Fig. 4

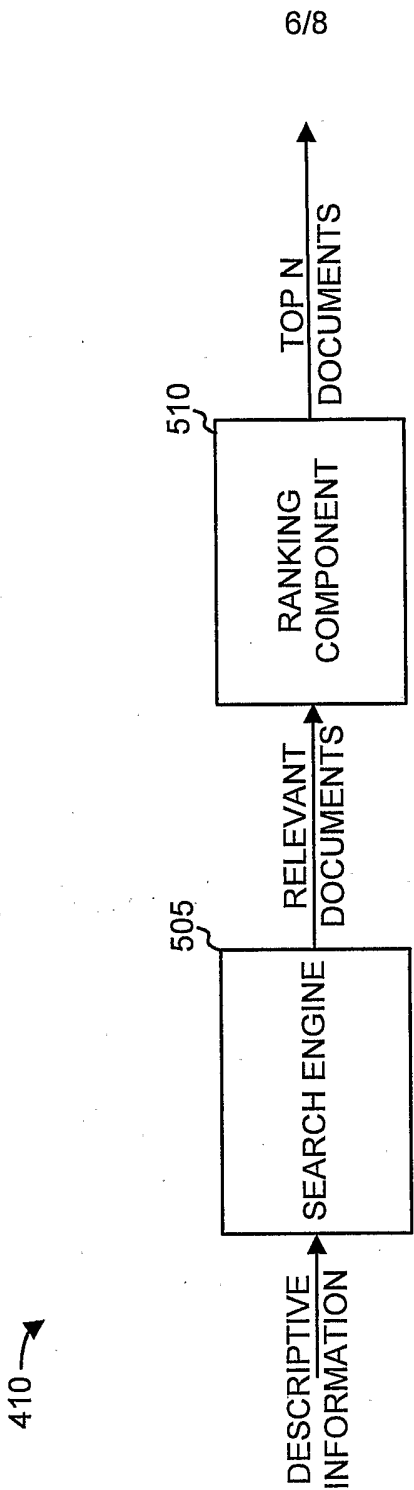


Fig. 5

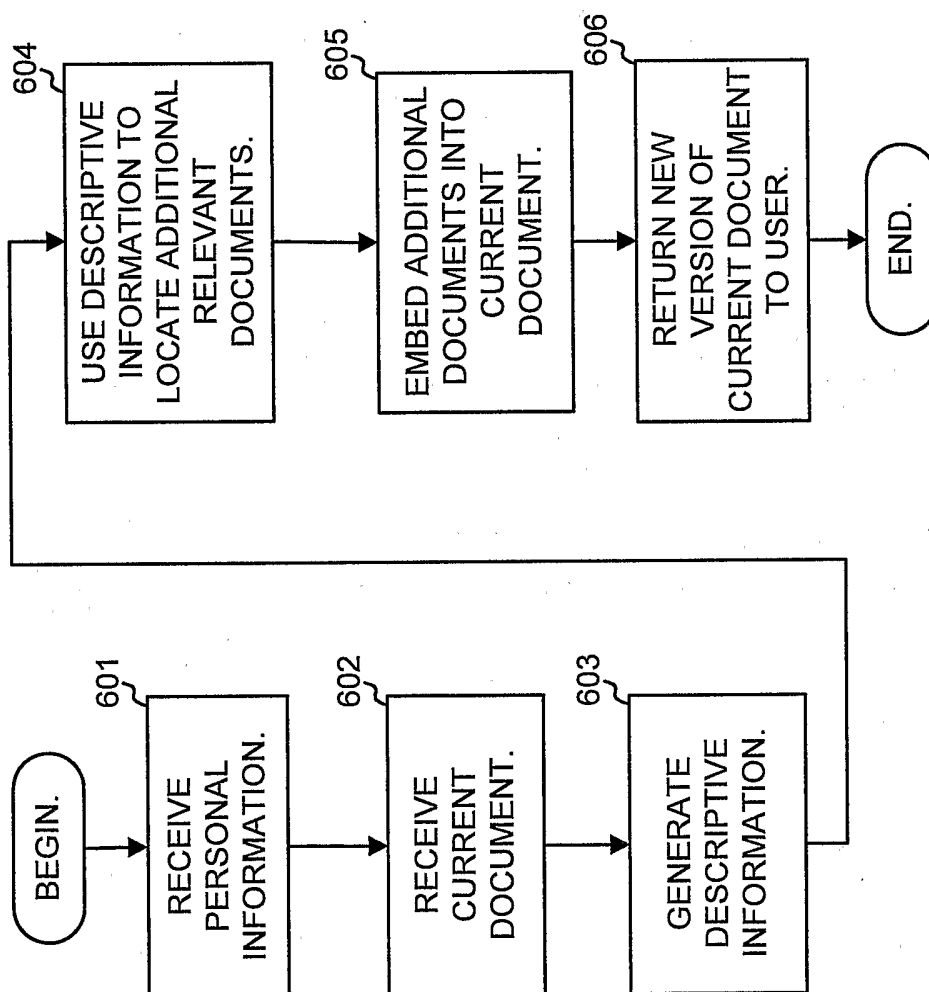


Fig. 6

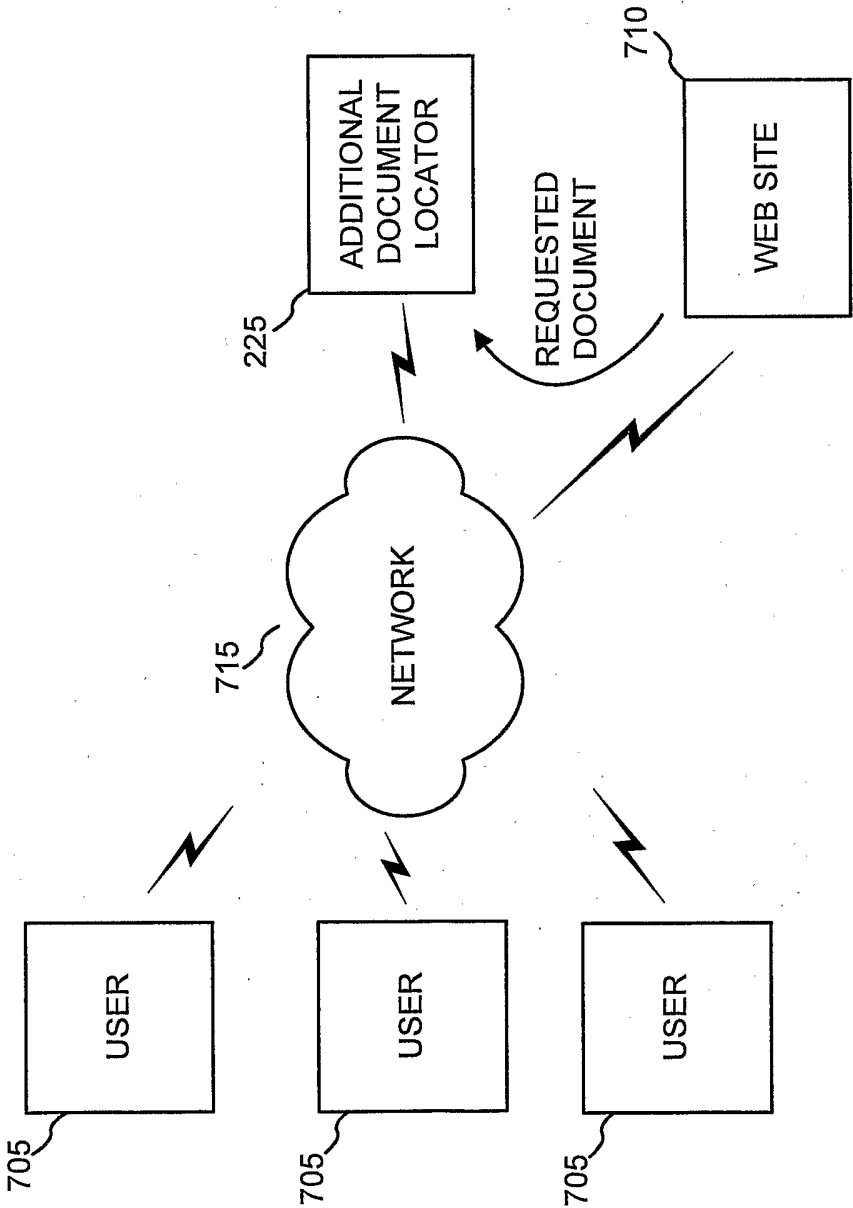


Fig. 7

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US2005/022726

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

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)

IPC 7 G06F

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 practical, search terms used)

EPO-Internal, WPI Data, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>US 6 122 647 A (HOROWITZ ET AL) 19 September 2000 (2000-09-19) abstract column 1, line 7 - line 11 column 2, line 45 - line 51 column 3, line 13 - line 18 column 4, line 26 - line 34 column 7, line 24 - line 27 column 7, line 57 - line 63 column 7, line 66 - column 8, line 9 column 8, line 22 - line 32 column 10, line 8 - line 27; figures 3,8 ----- -/--</p>	1-34



Further documents are listed in the continuation of box C.



Patent family members are listed in 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 document but 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.

G document member of the same patent family

Date of the actual completion of the international search

11 October 2005

Date of mailing of the international search report

18/10/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Haffner, R

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/022726

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00/43918 A (WEST PUBLISHING COMPANY D.B.A. WEST GROUP; ROSENOFF, DOUGLAS, T; MEDIN) 27 July 2000 (2000-07-27) abstract page 3, line 14 - line 27 page 5, line 9 - line 17 page 7, line 19 - line 24 page 10, line 5 - line 7 page 11, line 14 - line 16; figure 3 -----	1-34
X	GB 2 368 167 A (RADKA RADANA * DVORAK) 24 April 2002 (2002-04-24) abstract page 7, line 14 - line 20 page 7, line 27 - page 8, line 2 page 9, line 7 - line 13 page 10, line 24 - line 28 page 12, line 12 - line 34 page 16, line 25 - page 17, line 9; figure 6 -----	1-34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US2005/022726

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6122647	A	19-09-2000	NONE	
WO 0043918	A	27-07-2000	AU 2740200 A CA 2361021 A1 EP 1161733 A2 JP 2002535779 T NZ 513147 A NZ 529689 A	07-08-2000 27-07-2000 12-12-2001 22-10-2002 27-02-2004 29-07-2005
GB 2368167	A	24-04-2002	NONE	