(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 March 2001 (08.03.2001)

PCT

(10) International Publication Number
WO 01/16780 A1

(51) International Patent Classification[7]: G06F 15/16

(21) International Application Number: PCT/US00/24036

(22) International Filing Date: 31 August 2000 (31.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/386,942 31 August 1999 (31.08.1999) US

(71) Applicant: MCI WORLDCOM, INC. [US/US]; 515 East Amite Street, Jackson, MS 39201 (US).

(72) Inventors: REID, Dylan; 100 Verlaine Place, Atlanta, GA 30327 (US). GOLDMAN, Les; 4 Farm House Lane, Huntington, CT 06484 (US). KENNEDY, Nick; 66 Warncke Road, Wilton, CT 06897 (US). LAMBERT-GORWYN, Adam; 6833 S. Ivy Way, #6-307, Englewood, CT 80112 (US).

(74) Agents: GROLZ, Edward, W. et al.; Scully, Scott, Murphy & Presser, 400 Garden City Plaza, Garden City, NY 11530 (US).

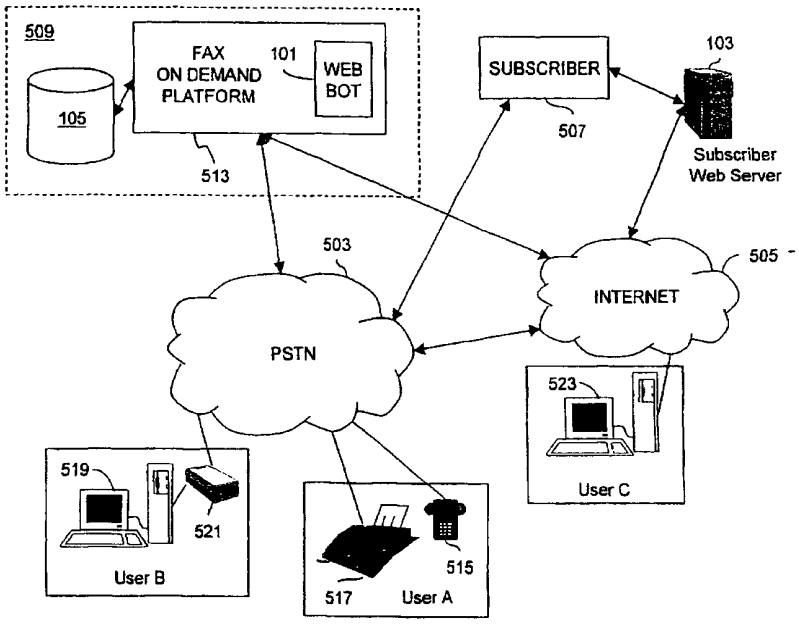
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, 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: WEB INDEXING FOR INFORMATION ON-DEMAND DELIVERY SYSTEMS



(57) Abstract: An approach for automating a process by which source documents residing on a packet switched network, such as the Internet (505) or an intranet, are made available for information on-demand services (509). The information on-demand service (509) can deliver the content of these documents via facsimile (517) and/or e-mail (519). This automatic update approach ensures greater accuracy of maintaining the source documents for document delivery to a user.

WO 01/16780 A1

TITLE OF THE INVENTION

WEB INDEXING FOR INFORMATION ON-DEMAND DELIVERY SYSTEMS

BACKGROUND OF THE INVENTION

Field of the Invention:

The present invention relates to information on demand delivery systems, and more particularly relates to on demand facsimile delivery of information.

Discussion of the Background

Facsimiles (i.e., faxes) continue to be a popular medium for conveniently transmitting textual and graphical information. Business communication has evolved to rely heavily on the use of faxes for day to day transactions. For that matter, personal communication has enjoyed the use of faxes because of the prevalence of fax machines and relatively low transaction cost. Despite the advances in other forms of messaging, fax technology remains popular because of the wide acceptance of fax standards by the manufacturers. A number of factors contribute to this popularity. For example, sending a fax is ideal when the recipient does not have an e-mail address or convenient access to a computer. In addition, the timestamp and authentication capability of faxes provide a mechanism whereby business transactions can be conducted with certain assurances of integrity. Users and manufacturers of fax machines alike have come to

realize that this indispensable tool is not incompatible with emerging technologies. On the contrary, the prevalence of fax usage stems, in large part, from the flexibility of fax technology to adapt and to integrate with other forms of communication, such as e-mail, text paging, and client-server technology.

Furthermore, fax technology has been embraced by the global network of computers, known as the Internet. Usage of the Internet has been greatly stimulated by the World Wide Web (the Web), in large part, because of the Web's graphical user interface (GUI). This easy-to-use and reliable GUI has widened the audience of the Internet, which in the past has been restricted to academia and government entities, to the general population. Because the Web provides a forum that is conducive to collaboration, it has become an invaluable data repository that continues to grow exponentially as users and content providers continually contribute to it.

As a consequence, the capability to deliver documents, on an on demand basis, within this vast data repository via fax is quite desirable. Fax retrieval capability is realized in a number of ways. First, user stations (e.g., personal computers or workstations) connected to the Internet are equipped with fax software, which allows the computer to send and receive faxes over the Internet. These user stations typically access the Internet using dial-up connections through the public switch telephone network (PSTN). In such an arrangement, the user essentially uses his computer or workstation as a fax machine. In another scenario, a user can retrieve information from the Internet by having the information delivered to an accessible fax machine. As will be discussed later, the fax message can be alternatively delivered to the user, who is on a computer, as an e-mail.

Figure 5 illustrates a conventional approach to providing fax on demand service that interfaces with the Web (i.e., Internet) to permit pre-designated users to access various

documents on a subscriber's web site. As shown in the Figure 5, a fax on demand platform 501 has connectivity to the public switch telephone network (PSTN) 503 as well as the Internet 505. Typically, a subscriber 507 has an account with a telecommunication service provider 509 to provide fax on demand services on behalf of the subscriber 507. The subscriber 507 supplies the service provider 509 with a library of source documents that the subscriber 507 desires to make available to its users (or customers). The source documents may detail information about the subscriber 507, along with the services and products of the subscriber 507. The subscriber's users (e.g., users A, B, and C) have the capability to retrieve any or all of the source documents at any time (24 hours a day, 7 days a week) without human intervention by the subscriber 507 or the service provider 509.

In the system of Figure 5, the service provider 509 has a service center 511 that interacts with the fax on demand platform 501. The service center 511 has a staff of customer representatives who communicate with the subscriber 507 over the PSTN 503 to address a variety of requests that cannot be handled directly by the fax on demand platform 501. One such request may involve the subscriber 507 desiring to have the content of its web site resident on web server 508 be available to its users via fax document transmission. The content may be specified through the uniform resource locators (URLs). These URLs enable access to documents not physically on the subscriber's web site as well as files within the web server 508. As more fully discussed below with respect to Figure 6, this arrangement provides the subscriber 507 with a mechanism to supply the content of its web site to its users as a fax message.

The fax on demand platform 501 has the capability to interact with Users A, B, and C via the PSTN 503 as well as the Internet 505. For example, User A can use his telephone 515 to dial into the fax on demand platform 501 through the PSTN 503 and request information to be faxed

to a fax machine 517. User B may also obtain information from the fax on demand platform 501 by using his computer 519. The computer 519 is equipped with faxing software so that User B may receive and send faxes, basically emulating a fax machine. User B, in the alternative, may elect to receive the fax as a file attached to an e-mail or as embedded text into the e-mail itself. User B simply dials into the fax on demand platform 501 using communication software loaded into computer 519 and modem 521. The modem 521 communicates over the PSTN 503 with the fax on demand platform 501 to convey the request by User B for a particular document. In response to the request, the fax on demand platform 501 delivers the corresponding fax of the requested information to computer 519 of User B in the manner that was specified in the request. In yet another configuration, User C has the capability to request information by utilizing a web browser over the Internet 505 using computer 523 to communicate with the fax on demand platform 501. With appropriate software, computer 523 may receive the requested faxes from the fax on demand platform 513 over the Internet 505 without going through the PSTN 503 if User C utilizes a local area network (LAN) (not shown) to directly access an Internet node (not shown). A drawback in the arrangement of Figure 5 is that the update process of supplying the service provider 509 with the URLs contained within the web site is accomplished manually. This manual process is inefficient in terms of time and labor costs, and the probability of errors is increased.

As indicated above, the web site of subscriber 507 may refer to various source documents by indicating the associated URLs. In general terms, a URL specifies an address of an "object" (e.g., file, a web page, newsgroups, etc.) in the Internet 505 by explicitly indicating the method of accessing the resource. A representative format of a URL is as follows: http://www.wcom.com/homepage/document. This example indicates that the file "document" is

accessed using the Hypertext Transfer Protocol (HTTP). HTTP is an application level protocol (Layer 7 of the OSI model) that is employed for information transfer over the Web. RFC (Request for Comment) 2616 specifies this protocol and is incorporated herein in its entirety. The two slashes after the colon signifies that what follows refers to a machine with that designated domain name (wcom.com), while the remaining text refers to the file ("document") and its file structure. Accordingly, it is possible that a web site can serve as a "virtual" library by enumerating multiple URLs. A more detailed definition of URL can be found in RFC 1737, which is incorporated herein in its entirety.

Figure 6 is a flowchart illustrating a conventional method that is used by subscriber 507 to notify the service center 511 that the subscriber 507 desires to update its source document list (i.e., document catalog) to reflect the changes made to the content of its web site. As in step 601, subscriber 507 manually requests, typically through a telephone call to the service center 511 or the e-mailing of a database (spreadsheet) with account changes, an update of the URLs. As previously mentioned, these URLs correspond to source documents or files within the web server 103 and to remote web sites (not shown).

Depending on the type of subscriber 507, the source documents for this particular subscriber may require updating quite frequently. For example, if the subscriber 507 is a retailer that sells a product whose prices fluctuate often, the subscriber 507 would need to update this information throughout all the source documents that contain these prices. If the URLs corresponding to these documents change due to an update of the file structure, for instance, then new URLs are required. The subscriber 507 must accordingly notify the service center 511. Once the subscriber 507 informs the service center 511 of the updated URLs within the web site, the service center 511 manually enters, per step 603, the updated URLs into the fax on demand

platform 501. Typically, the data entry is performed through a computer that is networked over a local area network (LAN) (not shown) to the fax on demand platform 501. In step 605, the fax on demand platform 513 then retrieves the updated documents according to the updated URL list, such that they now appear among the available source documents (i.e., document catalog). Assuming the subscriber 507 has users A, B, and C, any one of them may select, per step 607, the documents listed in the document catalog of the subscriber 507 for retrieval via fax or e-mail. The primary disadvantage with the above approach is that the subscriber 507 and the service center 511 engage in manual operations for synchronizing the web site's content with the fax on demand platform 501. The manual processes of steps 601 and 603 are susceptible to human errors and delay; in addition, these processes are labor intensive.

Based on the foregoing, there is a clear need for improved approaches for updating the document catalog and associated source documents of a subscriber's web site in a fax on demand system.

There is also a need for improved integration between the fax on demand platform and the World Wide Web. There is a further need to minimize errors associated with updating of the document catalog.

Based on the need to provide better integration between the fax on demand platform and the Internet 505, an approach, which is efficient and is not prone to errors, for automatic update of the document catalog to reflect changes in the subscribers web site is highly desirable.

SUMMARY OF THE INVENTION

According to one aspect of the invention, a method is provided for performing on demand fax delivery of information that resides on a packet switched network. The method

comprises accessing a subscriber web site to produce a document catalog, which includes a plurality of entries. Each of the entries contains document identification information and the corresponding address information. The method also includes maintaining a database, which stores a plurality of entries; each of the database entries has one of the source documents and the corresponding document identification information. The source documents are accessible via a fax on demand platform. The method further includes selectively synchronizing the source documents based upon the document catalog via a web robot mechanism that searches for web sites that are accessible through the packet switched network to update the database. Under this approach, the database of source documents corresponding to URLs of a web site is automatically updated, thereby reducing labor costs and errors.

According to another aspect of the invention, a system for providing on demand fax delivery of information that resides on a packet switched network. A fax on demand platform communicates with the packet switch network via a web robot mechanism. The fax on demand platform is configured for accessing a subscriber web site to produce a document catalog, which includes a plurality of entries. Each of the entries contains document identification information and the corresponding address information. A database has a plurality of entries, in which each of the database entries includes a source document and corresponding document identification information. The source documents are accessible via the fax on demand platform. The fax on demand platform launches the web robot to selectively synchronize the source documents based upon the document catalog. The web robot mechanism searches for web sites that are accessible on the packet switched network to update the database. The above arrangement advantageously effects an integrated system of fax on demand and the World Wide Web.

In yet another aspect of the invention, a computer-readable medium carrying one or more sequences of one or more instructions synchronizes a database of source documents. The one or more sequences of one or more instructions include instructions which, when executed by one or more processors, cause the one or more processors to perform the step of establishing a communication session with a subscriber web site. Another step comprises scanning the subscriber web site to produce a document catalog, which includes a plurality of entries, such that each of the entries contains document identification information and corresponding address information. Another step includes searching for web sites that are accessible through the packet switched network. Another step includes selectively synchronizing the source documents residing in a database based upon the document catalog. Another step involves interfacing with an on demand platform to modify the database. This arrangement advantageously permits efficient updating of the source documents within the fax on demand platform to coincide with the content of a subscriber web site.

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by reference to the following detailed description when considered in connection with the accompanying drawings, wherein:

Figure 1 is a block diagram of a fax on demand system in accord with one embodiment of the present invention;

Figure 2 is a flowchart illustrating the document update and retrieval process of the system in Figure 1;

Figure 3 is a flowchart of the operation of the web bot in providing an updated document catalog and document database in the system of Figure 1;

Figure 4 is a block diagram of an exemplary computer system capable of launching a web bot in the system of Figure 1;

Figure 5 is a block diagram of a conventional fax on demand system;

Figure 6 is a flowchart of a traditional approach involving a subscriber requesting update of a document database in the system of Figure 5.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the following description, for the purpose of explanation, specific details are set forth in order to provide a thorough understanding of the invention. However, it will be apparent that the invention may be practiced without these specific details. In some instances, well-known structures and devices are depicted in block diagram form in order to avoid unnecessarily obscuring the invention.

As will become apparent, an approach for automating a process by which source documents residing on a packet switched network, such as the Internet 505 or intranet, are made available for information on demand services. Source documents are defined herein to encompass the following: a fax document, an email document, an image file, a video file, a sound file, and any combination of the above documents and/or files. When a fax machine is employed, source documents refer to facsimile documents or email documents; however, when a computer is utilized, source documents refers generally to any bit stream (e.g., audio, visual, graphical, or textual data). The information on demand platform is able to automatically map the content of a web site for use as documents for end user queries. The approach, in part, utilizes a

web robot (or web bot) mechanism whose task is to examine periodically or upon request a specified web site and map the content utilizing the web site's URLs. The term mechanism is used herein to refer to a process, which may be implemented in software or hardware. The mapping of the content is then cross-indexed to a set of automatically generated document numbers that can be used within the information on demand system to automatically generate the subscriber's document database. To better understand the present invention, it is instructive to discuss briefly the concept and operation of an information on demand platform, in particular, the fax on demand platform's capability to provide fax back functions.

The fax on demand functionality provides a fast, convenient, and cost effective way to supply up-to-date information, on an as needed basis. Accordingly, a subscriber of a fax on demand service can readily "publish" information to a group of pre-designated users (or customers). This service, for instance, is suitable for large groups of customers, prospective clients, sales organizations, association members, and employees, and can be utilized 24 hours a day, 7 days a week. As a specific example, subscribers in the retail industry may shorten their company's sales cycle by getting information to sales representatives in the field or directly to the prospective clients in minutes as compared to days if the information were to be sent via conventional airmail. Such a capability accelerates the decision making process and reduces collateral costs required for brochure creation, updates, printing, distribution and storage costs.

Further, once implemented, subscribers can promote the availability of their fax on demand capability as a value-added service. Callers can, for example, access the service to request information via a unique toll-free number or internationally via a DID (direct inward dialing) long distance number. Following a series of customized voice prompts by the fax on demand platform, the caller is able to select the desired documents and catalogs and is prompted

to enter a fax machine phone number via the touch-tone keypad. Optionally, a caller may be asked to enter their voice phone number, member number, or name for insertion on the cover page. The document is faxed to the caller within a relatively short period. If the receiving fax machine is busy, the service retries sending the information until a connection is established or according to some predetermined setting.

Referring now to the drawings, wherein like reference numerals designate identical or corresponding parts throughout the several views and more particularly to Figure 1 thereof, there is illustrated a block diagram of a fax on demand system utilizing a web robot program. As shown in Figure 1, a web bot software utility 101 resides within the fax on demand platform 513. The web bot software 101 can also be deployed, as an alternative embodiment, in a server (not shown) that is separate from the fax on demand platform 513. The web bot 101 upon command by the fax on demand platform 513 retrieves information automatically from remote web sites using standard Web protocols. That is, the Web robot traverses the webs hypertext structure by retrieving a document as well as documents that have been referred to by the original documents in a recursive fashion. Such a program has also been referred to as web bots, web wanderers, web crawlers, and spiders; as used in this application the term web robot is synonymous with these terms. It should be noted that the term "recursive" used above does not limit the definition to any specific algorithm. It is apparent to one of ordinary skill in the art that even if a web robot applies some heuristic to the selection and order of documents to traverse and spaces out requests by the users over a long period of time, it is still a robot. It should be noted that although the standard web browsers have the ability to traverse the hypertext structure of a web site, these web browsers are not considered robots mainly because they are operated by a human, and thus, do not automatically retrieve reference documents (other than in-line images).

The system of Figure 1 provides an integrated solution between the fax on demand platform 513 and services of the Internet 505. In an exemplary implementation, the fax platform 513 may comprise a number of servers (not shown) to provide interfaces for the voice switching system (not shown) as well as faxing functionalities. The number of servers and type of software depends upon the system capacity and breadth of functions desired. The fax platform has the flexibility to accommodate inquiries by the users from a variety of sources. For example, the subscriber can use a standard Internet connection to access the platform or simply use a touch-tone phone. Although Figure 1 illustrates the Internet 505, it is understood by one of ordinary skill in the art that any type of packet switched network can be used; for example, an intranet. Basically, an intranet is a private packet switched network that is behind a firewall -- i.e., not accessible by the general public. As shown, a subscriber web server 103 directly connects to the Internet 505. The subscriber web server 103 hosts the web site of subscriber 507, whereby users A, B, and C may access documents and other information, that are resident on the web server 103.

One of the unique features of the Web is that web sites are updated frequently to ensure that the content reflect the most updated information. It is common practice for webmasters, who maintain web sites, to revise the web site on a daily basis. For example, a webmaster of a newspaper needs to track the web site with the daily printed publication. Because of the dynamic nature of the Web, subscriber 507 may regularly update the content of its web site. In one exemplary scenario, if the subscriber 507 is a business that sells products and services, it is conceivable that the subscriber 507 would market such products on its web site. Thus, consumers who use the Internet 505 may want to access the subscriber's web server 103 to gain information about the products by retrieving product specification sheets and the like.

Under the arrangement of Figure 1, the fax on demand platform 501 utilizes web bot 101 to automatically retrieve the documents of the subscriber's web site on the subscriber web server 103 as the content of the web site changes. In this manner, the subscriber 507 need not contact the service provider to notify such provider of a change in its web site, in contrast to the conventional approach. Because the web bot 101 can be launched automatically without human intervention, the web bot 101 reduces delay in synchronizing information resident on the subscriber web site with that of the database 105, while decreasing the likelihood of data entry errors. The details of the web bot's 101 functionality and its interaction with the fax on demand platform 513 is more fully described in Figure 3 below. Before examining the operation of the web bot 101, an overview of the update process (Figure 2) according to an embodiment of the present invention is discussed.

Figure 2 is a flowchart delineating the manner in which a user is able to retrieve information off the subscriber web server 103 without needing to utilize a computer that is connected to the Internet 505. In step 201, subscriber 507 modifies its web site on the web server 103 by either adding a new source document, deleting a document, updating existing ones, or any combination thereof. It is important to note that one way to accomplish these tasks is to manipulate the URLs that refer to the corresponding documents. Under the conventional approach, the subscriber 507 at this time would have to contact the service center 511 to notify that a change has been made to the web site, thereby affecting the list of available source documents on the conventional fax on demand platform 501. In contrast, the fax on demand platform 513 of the present invention, as in step 203, launches a web bot 101 to automatically update the document catalog associated with subscriber 507.

The web bot 101 can visit all the URLs that have been marked as a source document based upon the HTML (hypertext mark-up language) code of the subscriber's web site. In step 205, the web bot 101 maps the content of the subscriber web site by identifying the content as source documents. Thereafter, the web bot 101 cross-indexes the map of the content to a set of document numbers (step 207); as will be discussed later, these document numbers can be generated by the fax on demand platform 513, the web bot 101, or obtained from the HTML code of the web site. As in step 209, the web bot 101 produces a document catalog or report of the web site. The document catalog includes information identifying and locating the mapped source documents. Such information is more fully described with respect to Figure 3. Next, the web bot 101 updates the document database 105 based upon a comparison of the document catalog and the entries of the document database 105, per step 211.

As a result, the fax on demand platform 513 is now synchronized with the updated web site of subscriber 507. Anyone of the users A, B, or C, assuming they are authorized, may select documents from the document catalog to be retrieved via fax (step 213). The fax on demand platform 513 maintains an access control list (ACL) to identify which users the subscriber 507 has specified as having authority to access its source documents in the database 105. Further, the fax on demand platform 513 may use the ACL to specify the level of access rights Users A, B, and C possess. In this manner, different users may be restricted from retrieving certain documents. In this example, it is assumed that users A, B, and C are on the ACL with proper access rights.

Turning now to Figure 3, which illustrates a flowchart detailing the operation of the web bot 101. The web bot 101 can be invoked based upon a predetermined schedule or as required by certain events. In particular, the web bot 101 can be available on demand (e.g., run at the

time a subscriber account is registered) or may be run using a scheduler to periodically map and verify that the subscriber web site is registered to the service. That is, the process of mapping the content of a subscriber web site for use as documents by the end users A, B, and C, as described above, can be based upon a configurable timeframe (e.g., nightly, weekly, etc.) as specified by the subscriber 507 or based upon an event (e.g., upon request by the subscriber 507).

Once the web bot 101 is launched by the fax on demand platform 513, the web bot 101 scans the subscriber web site (step 301), which in this example is resident on web server 103, to produce a document catalog (i.e., a report) that includes information about the source document (step 303). The report includes multiple entries corresponding to the source documents. Each entry comprises document identification information (i.e., the document numbers) and the corresponding address information to locate the source documents. In the preferred embodiment, the address information is the URL.

During its execution, the web bot 101 looks for web pages with specific META tags, per step 305. If the specific META tag identifying a document as content for the informetion on demand platform is not present, the web bot 101 ignores the content within that domain. A META tag is a special HTML code that provides information about a web page. Unlike normal HTML tags, META tags do not affect how the web page is displayed. These META tags merely provide information on who created the web page, how often the web page is updated, what the web page is about, and which key words represent the web page's content. With respect to the present invention, the META tag may be designed to direct the web bot 101 to various documents within the web site that are to be included as part of the on demand content (i.e.,

within database 105). As indicated earlier, the web bot 101 outputs a document catalog. This catalog specifies the documents that are to be inserted into the database 105.

Where META tags do not contain document identification information (i.e., document numbers), the fax on demand platform 513 assigns a new document number corresponding to that particular source document. The order in which the catalog displays documents may be based upon information returned from the META tags. Independent of how document numbers are assigned, the catalog may be displayed in any format or order. For example, these documents could be displayed at the bottom of the catalog in a first in first out (FIFO) web page order in accordance to how the web bot 101 traverses the site. Further, according to an exemplary embodiment, the META tags include an index that instructs the web bot 101 on the location within the database 105 where the source document should be inserted. The META tags also comprises a document number that is to be assigned to that URL.

The web bot 101 next has to compare the entries of the document catalog with that of the current database entries so that the document catalog can be synchronized with the database. The process of synchronization involves deleting old or invalid documents as well as inserting new documents into the database 105. In step 307, the web bot 101 compares the current entries of the database 105, which include the document number and the document file itself (optionally, the URL), with the entries of the document catalog that resulted from the scan. The web bot 101, as in step 309, determines whether the entries of the document catalog matches the entries of the database 105. If these entries are not in the database 105 (i.e., are new entries), the documents associated with these particular entries are inserted into the database 105, per step 311. The insertion of the new documents, as previously indicated, can be based upon the index information in the META tags. On the other hand, if the entries are in the database, the web

robot 101, as in step 313, determines whether the entries are still valid. An entry is no longer valid, for instance, if the URL no longer exists. In other words, the source document associated with the URL can no longer be accessed. Accordingly, if the entries are no longer valid, the documents corresponding to these entries are deleted from the database 105, as in step 315. The deletion process is accomplished by the web bot 101, which notifies the fax on demand platform 513 to delete the identified entries in the document database 105. In step 317, a determination is made whether the entries are more recent versions of the source documents than those that are contained within the database 105. This scenario occurs when an entry of the document catalog generated by the web bot 101 and the entry of the database 105 matches, in which case, the most updated version should be stored in the database 105. If the matched entry has a more updated version, the fax on demand platform 513 updates the database accordingly (step 319).

In addition to the functions described above, the web bot 101 provides manual and auto-update of the document database 105 with an option for the subscriber 507 to append to or replace document files. The manual update process enables the service provider 509 to start the web bot 101 on demand and edit the document catalog prior to updating of the database 105. The automatic update process does not require intervention by the service provider 509; this automatic update capability can be used during the initial subscriber account setup, for example.

Advantageously, subscription to the fax on demand service does not require the purchase of any special equipment or software. In an exemplary embodiment, the subscriber 507 simply needs a fax machine (not shown) to load their documents into the fax messaging system 513, and a touch tone telephone (not shown) to create the voice prompt portion of the service. Also, SMTP (simple mail transfer protocol) e-mail document submission allows subscriber 507 to rapidly and conveniently add or edit a document via e-mail. This feature also provides

immediate confirmation of successful content change directly to the users' e-mail client, permitting the subscriber 507 to provide a cleaner document (i.e., higher resolution and free from fax transmission noise).

The described fax on demand platform 513 provides a number of functionalities that permit automatic update of source documents associated with the subscriber's web site. In addition, other related functions such as subscriber configurable web based on demand document handling exists. Specifically, the platform 513 provides the ability to receive document retrieval instructions in the form of an e-mail template by the subscriber 507. For example, the subscriber 507 may place a "button" on their web page to mask a CGI (common gateway interface) script. This script prompts a user to enter various pieces of required and optional demographic information and a delivery destination address -- either phone number for fax delivery or e-mail address. The CGI scrip takes information and masks it in the form of a templated e-mail message that is launched at the fax on demand platform 513.

The fax platform 513 interprets the information contained on the form and sends the URL reference as the on demand document. The CGI scrip further allows the user to select the current web page or select a list of web page documents/URLs when placing the request. One of ordinary skill in the art is capable of developing a standard CGI script to accomplish the above functions.

In the system of Figure 1, subscriber 507 can monitor the effectiveness of the fax on demand service by utilizing the reporting capability of the fax on demand platform 513. All information requested, for example, via the touch-tone keypad of telephone 515 may be included in management reports. Further, these management reports may be delivered via electronic messaging (e.g., e-mail, text paging, etc.) or may be sent to a designated fax machine 517.

This system further provides web based access to the platform 513 for the purpose of maintaining various aspects of the subscriber's account. The subscriber 507 may use a standard Internet connection to access the platform 513 via secure login procedure. Once the login has been validated and secure connection established, the subscriber 507 may view various account parameters: document profile information attributes, personal identification number (PIN) and URL file information, and reporting options.

Web based viewing of system reports is supported by the fax on demand platform 513. The platform 513 is able to deliver faxes to any valid Internet protocol (IP) address worldwide. The fax documents can be delivered via e-mail as a graphic file attached to the e-mail message; for example, a TIF (tagged image file) formatted file. Further, the body of the e-mail may contain the following information: the sender's fax phone number, the time received by the platform 513, and the number of pages contained in the graphical file.

Although the system of Figure 1 employs a web bot 101 within the fax on demand platform 513, another embodiment provides an arrangement whereby the web bot 101 resides within a computer system (not shown in Figure 1) that is separate from the fax on demand platform 513. Figure 4 is a block diagram of computer that is capable of executing the web robot 101.

FIG. 4 illustrates a computer system 401 upon which an embodiment according to the present invention may be implemented. Such a computer system 401 may be configured as a server to execute code that implements the web robot as earlier discussed. Computer system 401 includes a bus 403 or other communication mechanism for communicating information, and a processor 405 coupled with bus 403 for processing the information. Computer system 401 also

includes a main memory 407, such as a random access memory (RAM) or other dynamic storage device, coupled to bus 403 for storing information and instructions to be executed by processor 405. In addition, main memory 407 may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 405. Notably, the indexed URLs may be stored in main memory 407. Computer system 401 further includes a read only memory (ROM) 409 or other static storage device coupled to bus 403 for storing static information and instructions for processor 405. A storage device 411, such as a magnetic disk or optical disk, is provided and coupled to bus 403 for storing information and instructions.

Computer system 401 may be coupled via bus 403 to a display 413, such as a cathode ray tube (CRT), for displaying information to a computer user. An input device 415, including alphanumeric and other keys, is coupled to bus 403 for communicating information and command selections to processor 405. Another type of user input device is cursor control 417, such as a mouse, a trackball, or cursor direction keys for communicating direction information and command selections to processor 405 and for controlling cursor movement on display 413.

Embodiments are related to the use of computer system 401 to implement a web robot 101 for updating a source document database 105 of a fax on demand platform 513. According to one embodiment, this automatic update approach is provided by computer system 401 in response to processor 405 executing one or more sequences of one or more instructions contained in main memory 407. Such instructions may be read into main memory 407 from another computer-readable medium, such as storage device 411. Execution of the sequences of instructions contained in main memory 407 causes processor 405 to perform the process steps described herein. One or more processors in a multi-processing arrangement may also be

employed to execute the sequences of instructions contained in main memory 407. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions. Thus, embodiments are not limited to any specific combination of hardware circuitry and software.

The term "computer-readable medium" as used herein refers to any medium that participates in providing instructions to processor 405 for execution of the web robot to synchronize the database 105 of the fax on demand platform 513 with the web site of subscriber 507. Such a medium may take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media includes, for example, optical or magnetic disks, such as storage device 411. Volatile media includes dynamic memory, such as main memory 407. Transmission media includes coaxial cables, copper wire and fiber optics, including the wires that comprise bus 403. Transmission media can also take the form of acoustic or light waves, such as those generated during radio wave and infrared data communications.

Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, a PROM, and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave as described hereinafter, or any other medium from which a computer can read.

Various forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to processor 405 for execution. For example, the instructions may initially be carried on a magnetic disk of a remote computer. The remote computer can load the instructions relating to computation of the shared secret key into its

dynamic memory and send the instructions over a telephone line using a modem. A modem local to computer system 401 can receive the data on the telephone line and use an infrared transmitter to convert the data to an infrared signal. An infrared detector coupled to bus 403 can receive the data carried in the infrared signal and place the data on bus 403. Bus 403 carries the data to main memory 407, from which processor 405 retrieves and executes the instructions. The instructions received by main memory 407 may optionally be stored on storage device 411 either before or after execution by processor 405.

Computer system 401 also includes a communication interface 419 coupled to bus 403. Communication interface 419 provides a two-way data communication coupling to a network link 421 that is connected to a local area network 423. For example, communication interface 419 may be a network interface card to attach to any packet switched LAN. As another example, communication interface 419 may be an asymmetrical digital subscriber line (ADSL) card, an integrated services digital network (ISDN) card, a cable modem, or a modem to provide a data communication connection to a corresponding type of telephone line. Wireless links may also be implemented. In any such implementation, communication interface 419 sends and receives electrical, electromagnetic or optical signals that carry digital data streams representing various types of information.

Network link 421 typically provides data communication through one or more networks to other data devices. For example, network link 421 may provide a connection through local area network 423 to a host computer 425 or to data equipment operated by an Internet Service Provider (ISP) 427. ISP 427 in turn provides data communication services through the Internet 505. In addition, LAN 423 is linked to an intranet 429. The intranet 429, LAN 423 and Internet 505 all use electrical, electromagnetic or optical signals that carry digital data streams. The

signals through the various networks and the signals on network link 421 and through communication interface 419, which carry the digital data to and from computer system 401, are exemplary forms of carrier waves transporting the information.

Computer system 401 can send messages and receive data, including program code, through the network(s), network link 421 and communication interface 419. In the Internet example, a server 431 might transmit a requested code for an application program through Internet 505, ISP 427, LAN 423 and communication interface 419. One such downloaded application provides a web robot program that is used to automatically update a document database 105.

The received code may be executed by processor 405 as it is received, and/or stored in storage device 411, or other non-volatile storage for later execution. In this manner, computer system 401 may obtain application code in the form of a carrier wave.

The techniques described herein provide several advantages over prior fax messaging systems over a packet switched network, such as the Internet or intranet. By utilizing a web bot, the present invention presents an efficient approach to updating the source documents that are made available to users via the fax on demand capability. The fax on demand platform automates the process of honoring requests for information by various users via fax or e-mail.

Obviously, numerous modifications and variations of the present invention are possible in light of the above teachings. It is therefore to be understood that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described herein.

WHAT IS CLAIMED AS NEW AND DESIRED TO BE SECURED BY LETTERS PATENT OF THE UNITED STATES IS:

1. A method for providing on demand fax delivery of information residing on a packet switched network, comprising:

accessing a subscriber web site to produce a document catalog, the document catalog comprising a plurality of entries, each of the entries having document identification information and corresponding address information;

maintaining a database of source documents, the database having a plurality of entries, each of the database entries having one of the source documents and corresponding document identification information, the source documents being accessible via a fax on demand platform; and

selectively synchronizing the source documents based upon the document catalog via a web robot mechanism that searches for web sites that are accessible on the packet switch network to update the database.

2. The method according to claim 1, wherein the web sites in the step of selectively synchronizing include META tags that supply information about the source documents to the web robot mechanism.

3. The method according to claim 2, wherein the META tags comprise an index indicating address location within the database to insert one of the source documents and a document number specifying a document number assigned to the one source document.

4. The method according to claim 2, wherein the step of selectively synchronizing comprises launching the web robot mechanism according to a prescribed schedule.

5. The method according to claim 4, wherein the prescribed schedule is based upon events, the events comprising request by a subscriber.

6. The method according to claim 1, wherein the web robot mechanism in the step of selectively synchronizing is executed within the fax on demand platform.

7. The method according to claim 1, wherein the web robot mechanism in the step of selectively synchronizing is executed external to the fax on demand platform.

8. The method according to claim 1, wherein the address information of the accessing step is a uniform resource locator (URL).

9. The method according to claim 1, wherein the step of selectively synchronizing comprises:

(1) comparing the entries of the document catalog and the entries of the database to determine matching entries;

(2) adding new database entries into the database based upon the comparing step;

(3) determining whether the matching entries are valid and whether the determined valid entries, if any, require updating; and

(4) deleting or updating one or more of the database entries based upon the determining step.

10. The method according to claim 1, wherein the packet switched network is the Internet.

11. The method according to claim 1, wherein the packet switched network is an intranet.

12. The method according to claim 1, further comprising providing the source document from the database to a subscriber.

13. The method according to claim 12, wherein the providing step includes utilizing at least one of a facsimile service, an email service, and a telephone service.

14. The method according to claim 12, wherein the source document comprises at least one of a facsimile document, an email document, an image file, a video file, and a sound file.

15. A system for providing on demand fax delivery of information residing on a packet switched network, comprising:

a fax on demand platform communicating with the packet switch network via a web robot mechanism, the fax on demand platform being configured for accessing a subscriber web site to produce a document catalog, the document catalog comprising a plurality of entries, each of the entries having document identification information and corresponding address information; and

a database having a plurality of entries, each of the database entries including a source document and corresponding document identification information, the source documents being accessible via the fax on demand platform,

wherein the fax on demand platform launches the web robot mechanism to selectively synchronize the source documents based upon the document catalog, the web robot mechanism searching for web sites that are accessible on the packet switched network to update the database.

16. The system according to claim 15, wherein the web sites include META tags that supply information about the source documents to the web robot mechanism.

17. The system according to claim 16, wherein the META tags comprise an index indicating address location within the database to insert one of the source documents and a document number specifying a document number assigned to the one source document.

18. The system according to claim 15, wherein the fax on demand platform launches the web robot mechanism according to a prescribed schedule.

19. The system according to claim 18, wherein the prescribed schedule is based upon events, the events comprising request by a subscriber.

20. The system according to claim 15, wherein the address information of the accessing is a uniform resource locator (URL).

21. The system according to claim 1, wherein the web robot mechanism is configured to performed the steps of:

(1) comparing the entries of the document catalog and the entries of the database to determine matching entries;

(2) adding new database entries into the database based upon the comparing step;

(3) determining whether the matching entries are valid and whether the determined valid entries, if any, require updating; and

(4) deleting or updating one or more of the database entries based upon the determining step.

22. The system according to claim 15, wherein the packet switched network is the Internet.

23. The system according to claim 15, wherein the packet switched network is an intranet.

24. A computer-readable medium carrying one or more sequences of one or more instructions for synchronizing a database of source documents, the one or more sequences of one or more instructions including instructions which, when executed by one or more processors, cause the one or more processors to perform the steps of:

(1) establishing a communication session with a subscriber web site;

(2) scanning the subscriber web site to produce a document catalog, the document catalog comprising a plurality of entries, each of the entries having document identification information and corresponding address information;

(3) searching for web sites that are accessible through a packet switched network; and

(4) selectively synchronizing source documents residing in a database based upon the document catalog, the database having a plurality of entries, each of the database entries including a source document and corresponding document identification information; and

(5) interfacing with a fax on demand platform to modify the database.

25. The computer-readable medium according to claim 24, wherein the web sites in include META tags that supply information about the source documents.

26. The computer-readable medium according to claim 25, wherein the META tags comprise an index indicating address location within the database to insert one of the source documents and a document number specifying a document number assigned to the one source document.

27. The computer-readable medium according to claim 24, wherein the steps of (1)-(5) are executed according to a prescribed schedule.

28. The computer-readable medium according to claim 27, wherein the prescribed schedule is based upon events, the events comprising request by a subscriber.

29. The computer-readable medium according to claim 24, wherein the address information is a uniform resource locator (URL).

30. The computer-readable medium according to claim 24, wherein the selectively synchronizing step comprises the steps of:

(1) comparing the entries of the document catalog and the entries of the database to determine matching entries;

(2) adding new database entries into the database based upon the comparing step;

(3) determining whether the matching entries are valid and whether the valid entries require updating; and

(4) deleting or updating one or more of the database entries based upon the determining step.

31. The computer-readable medium according to claim 24, wherein the packet switched network is the Internet.

32. The computer-readable medium according to claim 24, wherein the packet switched network is an intranet.

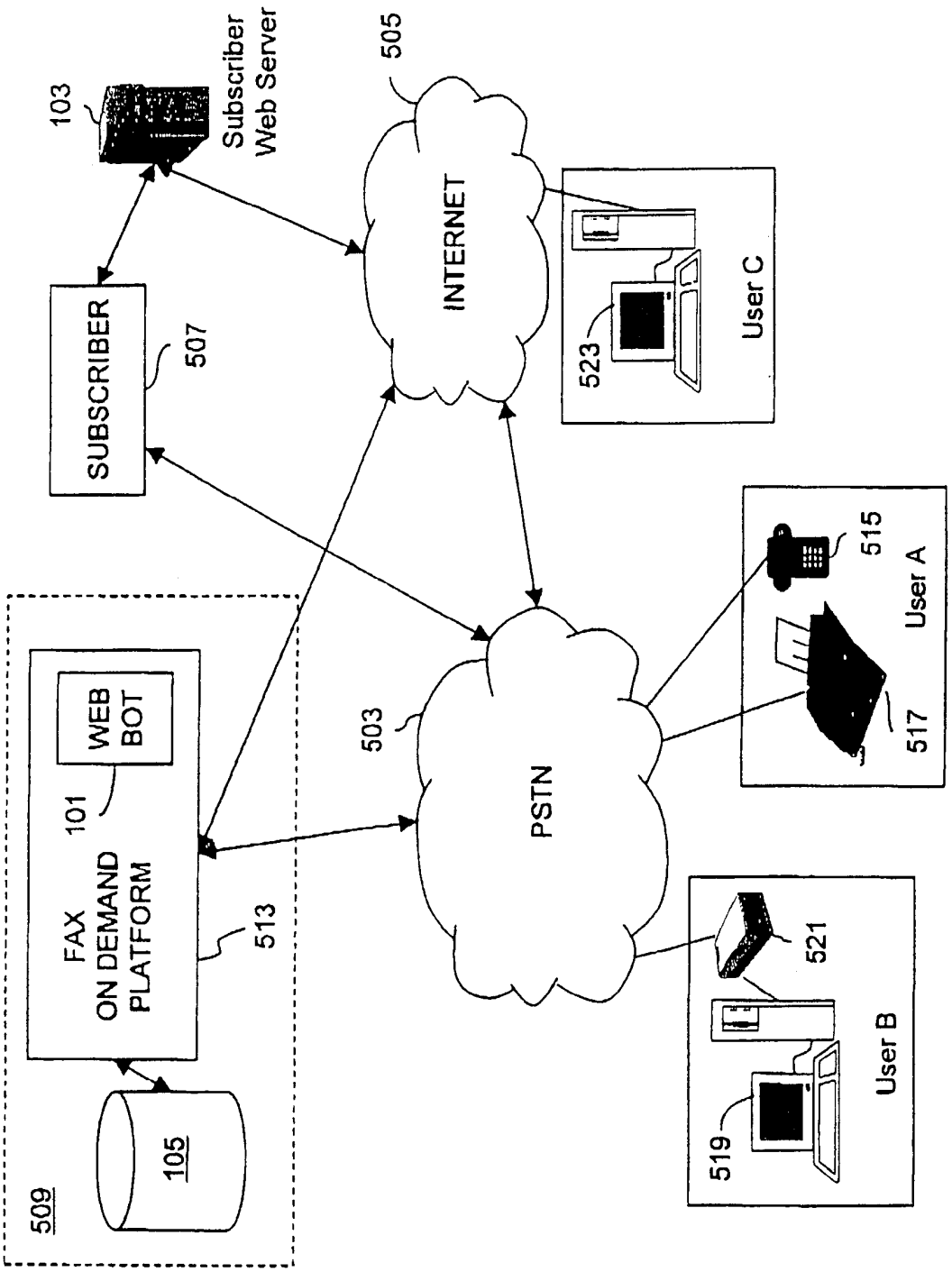

103
Subscriber Web Server
505
INTERNET
SUBSCRIBER
507
523
User C
WEB BOT
101
FAX ON DEMAND PLATFORM
513
503
PSTN
515
User A
517
509
105
521
519
User B

FIG. 1

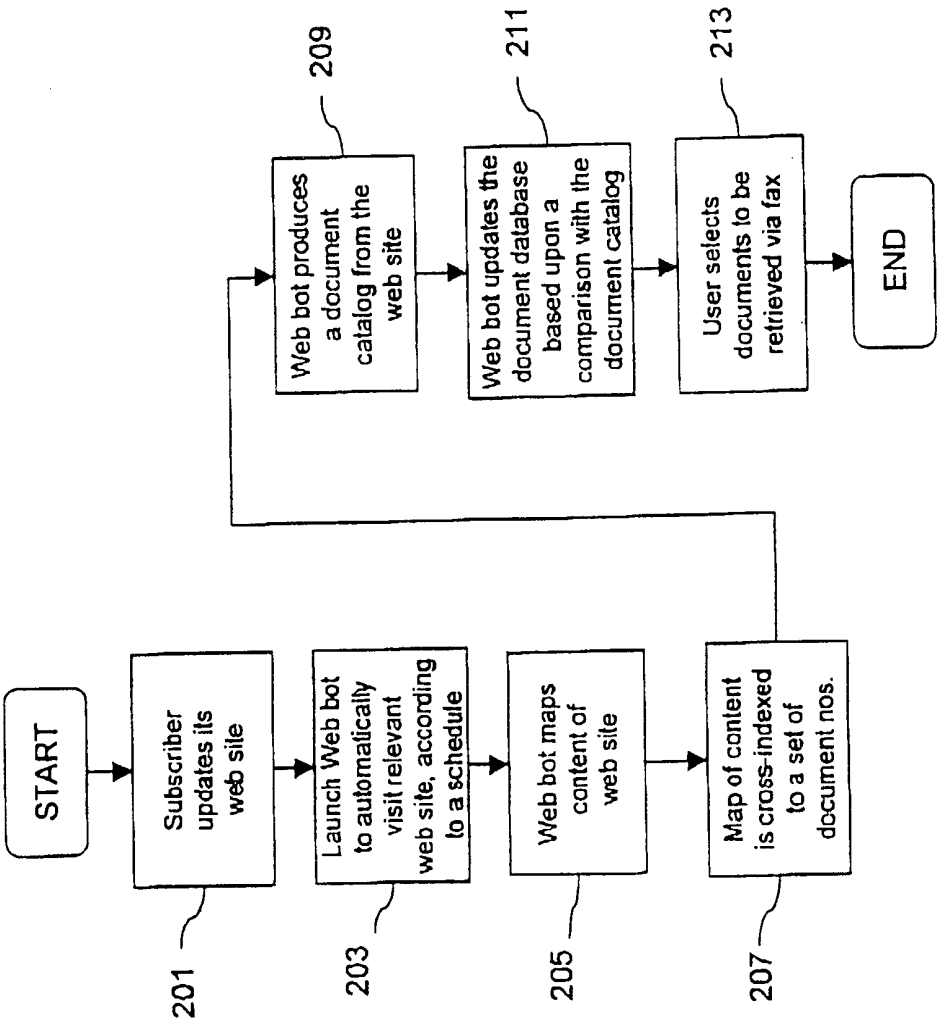

START
Subscriber updates its web site
Launch Web bot to automatically visit relevant web site, according to a schedule
Web bot maps content of web site
Map of content is cross-indexed to a set of document nos.
Web bot produces a document catalog from the web site
Web bot updates the document database based upon a comparison with the document catalog
User selects documents to be retrieved via fax
END
201
203
205
207
209
211
213

FIG. 2

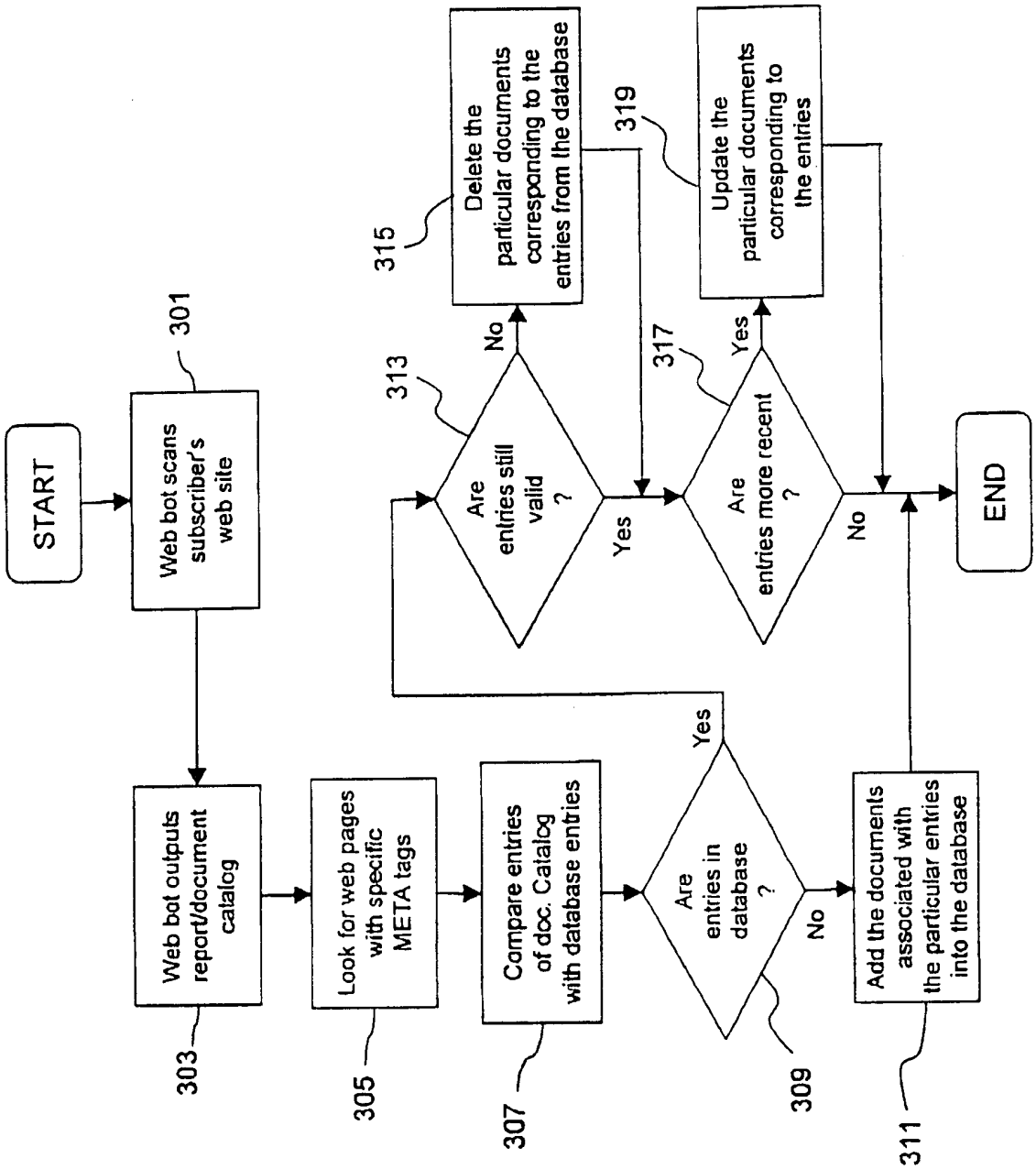

START
Web bot scans subscriber's web site
301
Web bot outputs report/document catalog
303
Look for web pages with specific META tags
305
Compare entries of doc. Catalog with database entries
307
Are entries in database ?
309
Yes
No
Add the documents associated with the particular entries into the database
311
Are entries still valid ?
313
No
Yes
Delete the particular documents corresponding to the entries from the database
315
319
Are entries more recent ?
317
Yes
No
Update the particular documents corresponding to the entries
END

FIG. 3

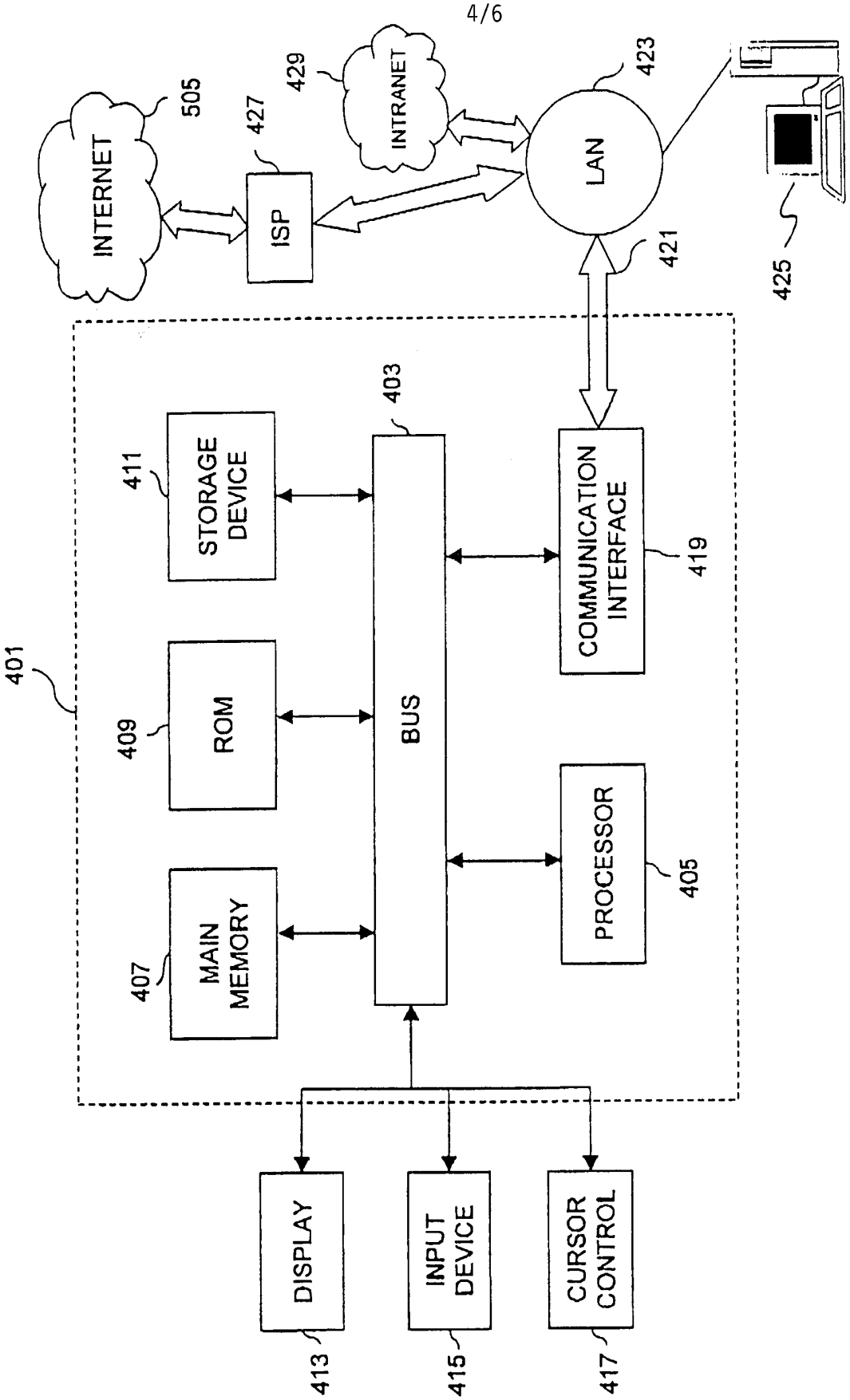

INTERNET
505
ISP
427
INTRANET
429
LAN
423
421
425
401
403
411
STORAGE DEVICE
409
ROM
407
MAIN MEMORY
BUS
COMMUNICATION INTERFACE
419
PROCESSOR
405
DISPLAY
413
INPUT DEVICE
415
CURSOR CONTROL
417

FIG. 4

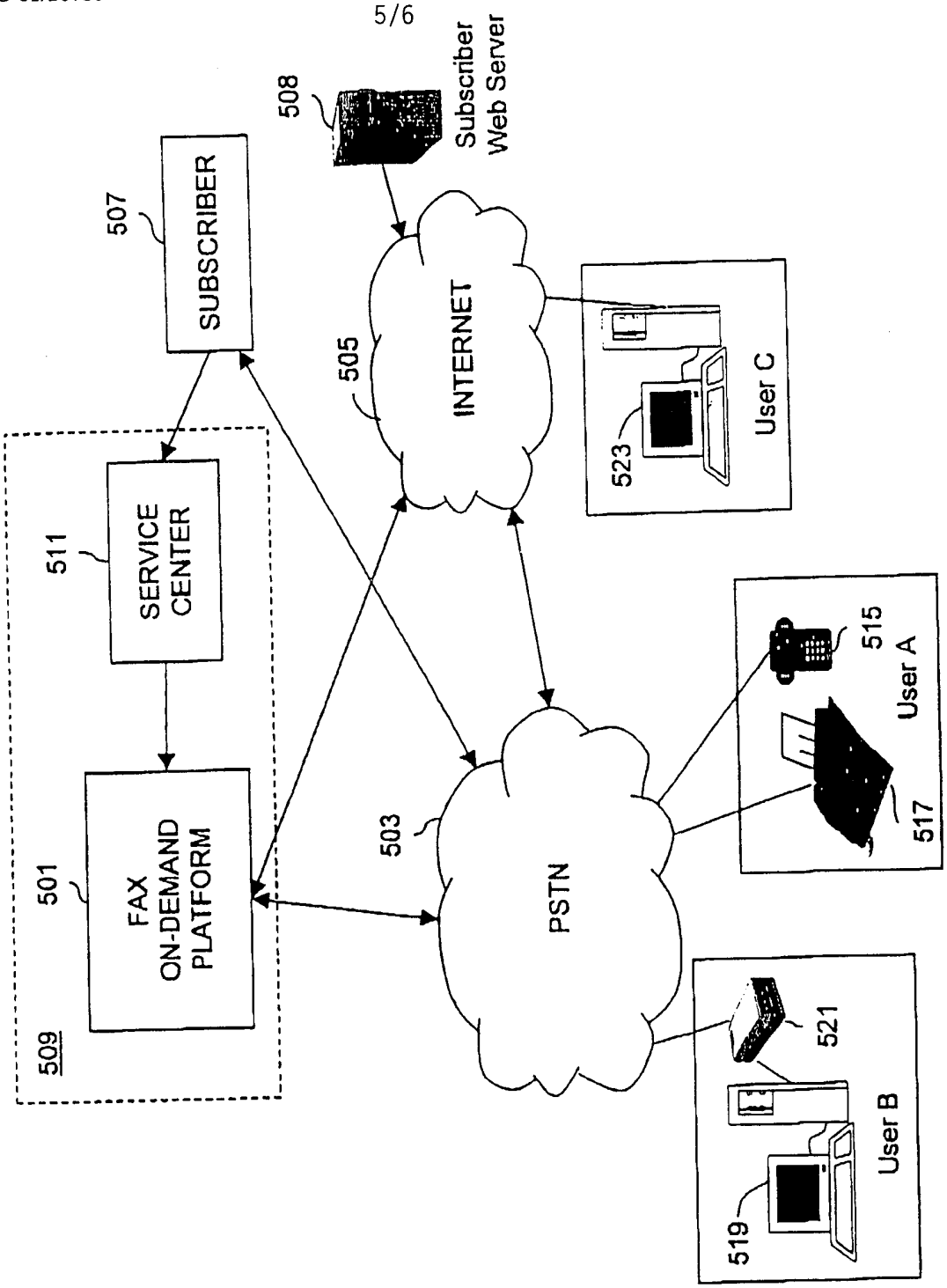


FIG. 5 PRIOR ART

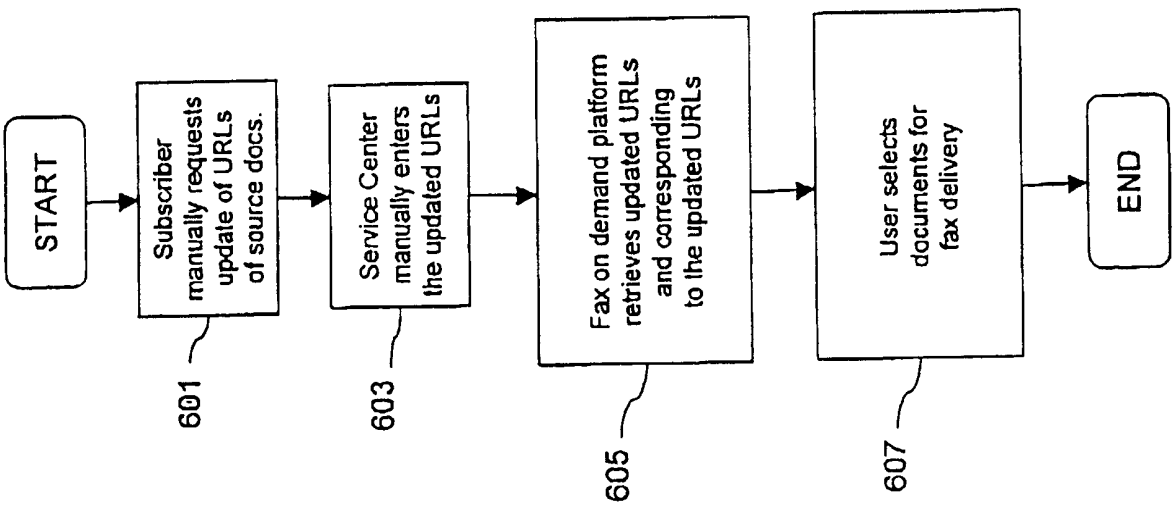


FIG. 6 PRIOR ART

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/24036

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 15/16
US CL :358/1.15; 709/206

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)

U.S. : Please See Extra Sheet.

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 practicable, search terms used)

EAST
search terms: fax, e-mail, on-demand, web site, URL, Internet, metadata, META

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,937,162 A (FUNK et al) 10 AUGUST 1999, see entire document.	1, 5-7, 9, 10, 12-15, 18, 19, 21, 22, 24, 27, 28, 30, 31
X	US 5,862,325 A (REED et al) 19 JANUARY 1999, see entire document.	1-5, 8, 10-14, 15-17, 20, 22-29, 31, 32
A	US 5,793,972 A (SHANE) 11 AUGUST 1998, see entire document.	1-32
A	US 5,761,662 A (DASAN) 02 JUNE 1998, see entire document.	1-32

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"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
"A"	document defining the general state of the art which is not considered to be of particular relevance	"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
"E"	earlier document published on or after the international filing date	"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
"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)	"&"	document member of the same patent family
"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		

Date of the actual completion of the international search 17 OCTOBER 2000	Date of mailing of the international search report 14 NOV 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JOSEPH R. POKRZYWA Telephone No. (703) 305-3800/4700

Form PCT/ISA/210 (second sheet) (July 1998)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/24036

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,740,549 A (REILLY et al) 14 APRIL 1998, see entire document.	1-32
A	US 4,604,686 A (REITER et al) 05 AUGUST 1986, see entire document.	1-32

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/24036

B. FIELDS SEARCHED
Minimum documentation searched
Classification System: U.S.

358/1.15, 402, 403, 407, 442-444, 468; 709/200, 204-207, 213, 216-219, 238-244; 379/90.01, 93.01, 93.05, 93.24, 93.25, 100.01, 100.08, 100.09, 100.11