(12) PATENT APPLICATION **(11)** Application No. **AU 200146223 A1**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Internet archive service providing persistent access to web resources

(51)[7] International Patent Classification(s)
G06F 017/30 **G06F 017/60**

(21) Application No: **200146223** (22) Application Date: **2001.05.23**

(30) Priority Data

(31) Number	(32) Date	(33) Country
09580149	**2000.05.30**	**US**

(43) Publication Date : **2001.12.06**
(43) Publication Journal Date : **2001.12.06**

(71) Applicant(s)
Lucent Technologies Inc.

(72) Inventor(s)
Ping-Wen Ong

(74) Agent/Attorney
SPRUSON and FERGUSON,GPO Box 3898,SYDNEY NSW 2001

INTERNET ARCHIVE SERVICE PROVIDING PERSISTENT ACCESS TO WEB RESOURCES

ABSTRACT

An Internet archive service is disclosed that provides persistent access to Web resources. Time-stamped Uniform Resource Locators ("URLs") that identify Web resources are combined with URL redirect techniques to create an Internet archive service to make the Internet persistent. The present invention provides persistent Web access, without requiring all web servers to directly handle time-stamps in a URL. A user enters a dated URL, for example, using a web browser (100), which is transmitted to a web server 140-150. The Web server receives the URL (containing a time stamp, a relative time-stamp or a variable time-stamp), and retrieves the correct Web page(s) from its own archive (45), or redirects the user to a predefined Internet archive site (160). An archive finder may optionally be provided to identify the best Internet archive service where the desired resource has been archived. The Web server interprets the extracted URL in accordance with the selected time stamp format. The present invention ensures that a time-stamped reference to any Web resource refers to the desired material.

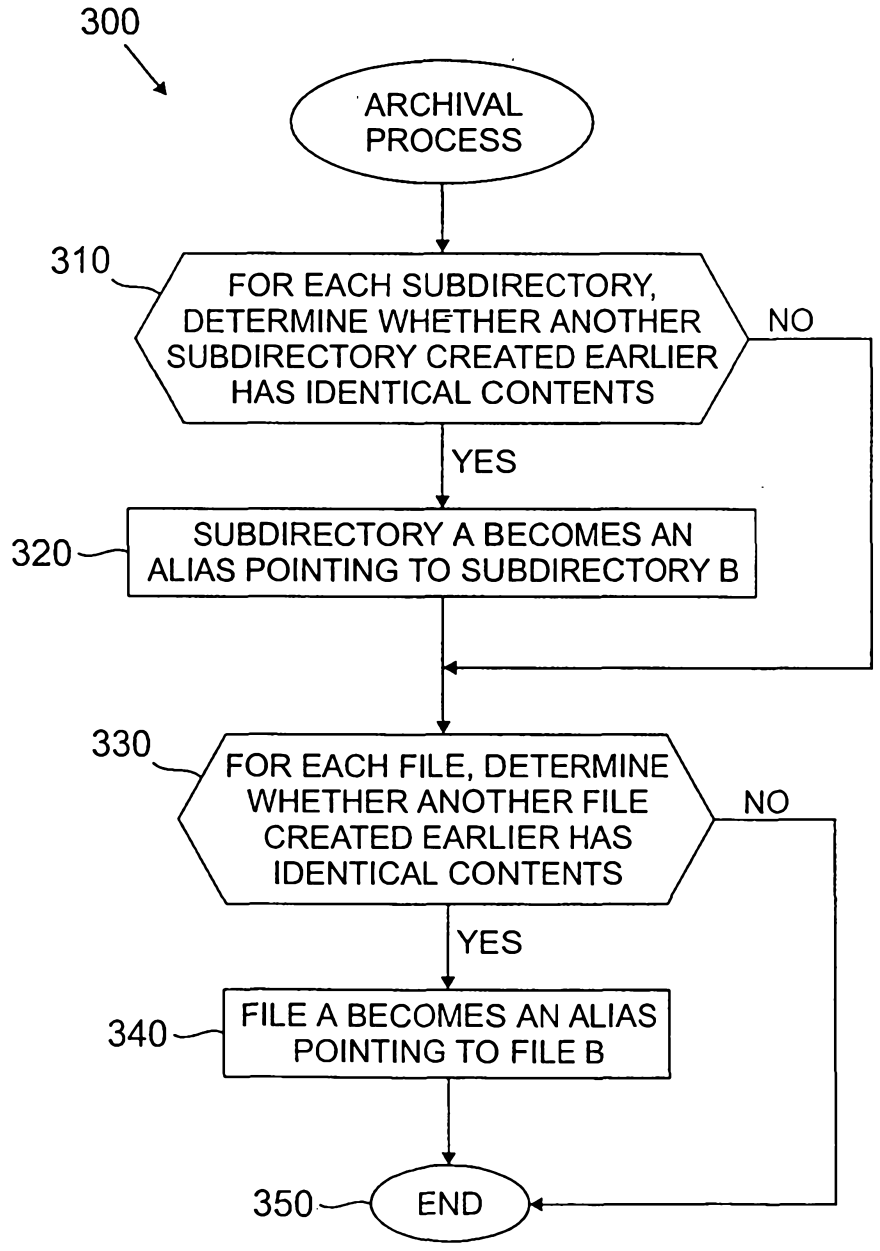
300
ARCHIVAL PROCESS
310
FOR EACH SUBDIRECTORY, DETERMINE WHETHER ANOTHER SUBDIRECTORY CREATED EARLIER HAS IDENTICAL CONTENTS
NO
YES
320
SUBDIRECTORY A BECOMES AN ALIAS POINTING TO SUBDIRECTORY B
330
FOR EACH FILE, DETERMINE WHETHER ANOTHER FILE CREATED EARLIER HAS IDENTICAL CONTENTS
NO
YES
340
FILE A BECOMES AN ALIAS POINTING TO FILE B
350
END

FIG. 3

S&F Ref: 550885

AUSTRALIA

PATENTS ACT 1990

COMPLETE SPECIFICATION

FOR A STANDARD PATENT

ORIGINAL

Name and Address of Applicant :	Lucent Technologies Inc. 600 Mountain Avenue Murray Hill New Jersey 07974-0636 United States of America
Actual Inventor(s):	Ping-Wen Ong
Address for Service:	Spruson & Ferguson St Martins Tower,Level 35 31 Market Street Sydney NSW 2000
Invention Title:	Internet Archive Service Providing Persistent Access to Web Resources

The following statement is a full description of this invention, including the best method of performing it known to me/us:-

Australia
Document Received on:
23 MAY 2001

5845c

INTERNET ARCHIVE SERVICE PROVIDING PERSISTENT ACCESS TO WEB RESOURCES

Cross-Reference to Related Applications

The present invention is related to United States Patent Application Serial No. 09/201,749, filed December 1, 1998, entitled "A Method And Apparatus For Resolving Domain Names of Persistent Web Resources," United States Patent Application Serial No. 09/201,750, filed December 1, 1998, entitled "A Method And Apparatus For Persistent Access to Web Resources Using Relative Time-Stamps," United States Patent Application Serial No. 09/201,751, filed December 1, 1998, entitled "A Method And Apparatus For Persistent Access to Web Resources," United States Patent Application Serial No. 09/201,752, filed December 1, 1998, entitled "A Method And Apparatus For Persistent Storage of Web Resources," and U.S. Patent Application Serial No. 09/342,408, filed June 28, 1999, entitled "A Method And Apparatus For Persistent Access to Web Resources Using Variable Time-Stamps,"each assigned to the assignee of the present invention and incorporated by reference herein.

Field of the Invention

The present invention relates to Internet resource access techniques, and more particularly, to a method and apparatus for ensuring persistent access to Internet resources.

Background of the Invention

The World Wide Web (the "Web") provides a dynamic way to present and distribute a vast amount of information. Anyone who is connected to the Internet and has a browser, such as Netscape Navigator Communicator™, commercially available from Netscape Communications Corporation of Mountain View, CA, can access information on the Web. The Web provides users with

many media options and is becoming ubiquitously available in an expanding variety of personal electronic devices, far beyond its initial limited availability to users via computer terminals. In addition, as display technologies continue to improve, the Web may ultimately replace traditional paper-based media altogether.

Paper-based media generally have an associated time stamp, and permit an easy determination of the information that was available at a given time. For example, a newspaper article can be cited as an authoritative reference, provided that the particular date of the newspaper publication is specified. Due to the dynamic nature of Web content, however, a Web document is generally not a reliable reference source. Currently, Web content cannot reliably be expected to be available in the same form and addressed by the same Uniform Resource Locator ("URL") at a future time. While some Web sites may provide access to some archived Web documents, the historical Web documents may not be accessed by users in a consistent and predictable manner, if at all.

The Online Computer Library Center, Inc. ("OCLC"), a nonprofit computer library service and research organization, provides a software tool, referred to as OCLC PURL ("Persistent Uniform Resource Locator"), for managing Internet addresses and aliases for general Internet resources. A Persistent Uniform Resource Locator provides flexible naming and name resolution services for Internet resources to ensure reliable, long-term access to Internet resources with minimal maintenance. Generally, OCLC PURL assists Internet users in locating Web resources. As previously indicated, the Internet is constantly expanding and changing. Once a Uniform Resource Locator (URL) changes, all previous references to that URL become invalid, thereby preventing users from accessing the Internet resource. The management of these changes often becomes burdensome.

While a URL points directly to the location of an Internet resource, a PURL points to an intermediate resolution service, which translates the PURL

into the actual URL. Once a Web resource has been registered with the OCLC and assigned a PURL, the Web resource may be accessed by means of the PURL. A PURL assigns a persistent name to a resource even if the location of the resource changes. In this manner, PURLs referenced in Web documents and other resources can remain viable over time without having to update the references each time the Web resource is moved. The PURL "forwarding" address maintained by OCLC, however, must be kept up-to-date. In other words, each time the document is moved, OCLC must be notified of the new address for the document.

The Internet Archive, a non-profit research organization, provides a regular snapshot of the contents of the Internet, in order to preserve valuable resources. Thereafter, restricted access is provided to authorized individuals for research using the archive. The archive is not available to the public and does not provide persistent access to Web resources.

Summary of the Invention

Generally, an Internet archive service is disclosed that provides persistent access to Web resources. According to an aspect of the invention, time-stamped Uniform Resource Locators ("URLs") that identify Web resources are combined with URL redirect techniques to create an Internet archive service to make the Internet persistent. The time stamp can be specified in the URL in any suitable format. The present invention allows the Web to be an organized and reliable reference source, much like paper-based media.

The present invention provides persistent Web access, without requiring all web servers to directly handle time-stamps in a URL. Generally, a user enters a dated URL, for example, using a web browser, which is transmitted to a web server. The disclosed Web browser can optionally include a mechanism to facilitate the specification of the desired date and time, or the user can manually append the time stamp to the URL indicated in the "Location" window of the browser. The Web server receives the URL (containing a time stamp, a relative

time-stamp or a variable time-stamp), and retrieves the correct Web page(s) from its own archive, or redirects the user to a predefined Internet archive site. The Web server interprets the extracted URL in accordance with the selected time stamp format. The present invention ensures that a time-stamped reference to any Web resource refers to the desired material. In this manner, anyone doing historical research on the Web can retrieve information that is valid in any period of time.

Brief Description of the Drawings

FIG. 1 illustrates an Internet or World Wide Web ("Web") environment in accordance with the present invention;

FIG. 2 illustrates a directory structure that arranges the contents of a Web site chronologically;

FIG. 3 is a flow chart illustrating an exemplary archival process implemented by a Web site to reduce the redundancy of a persistent archive of FIG. 1;

FIGS. 4A and 4B illustrate the aliasing of the contents of a Web site that is identical to previously archived contents by the archival process of FIG. 3;

FIGS. 5A and 5B are a sample table of a DNS server database in accordance with the present invention;

FIG. 6 is a flow chart describing an exemplary DNS server process performed by a DNS server to determine the IP address corresponding to a time-stamped URL; and

FIGS. 7A and 7B illustrates the various communications between the user's browser 100, the web server 140 and the Internet archive service 160 of FIG. 1, in accordance with two embodiments of the present invention.

Detailed Description

FIG. 1 illustrates a Web browser 100 in accordance with the present invention, that accesses information from one or more persistent Web servers 140, 150 over the Internet or World Wide Web ("Web") environment 130. The present invention provides persistent access to Web resources or electronic documents, including textual, audio, video or animation documents. According to a feature of the present invention, the Uniform Resource Locators ("URLs") that identify Web resources are augmented to include a time stamp. The Web browser 100 and persistent Web servers 140, 150 accommodate the additional time stamp parameter and allow a user to refer to any Web address with a precise target date. For example, the URL "http://cnn.com?time=2+2+1997" specifies the Web contents of the CNN home page on February 2, 1997. In this manner, the Web can be an organized and reliable reference source, much like paper-based media.

As shown in FIG. 1, an Internet archive service 160, discussed further below in a section entitled "Internet Archive Service," provides persistent access to Web resources. According to one feature of the present invention, time-stamped URLs are combined with URL redirect techniques to create an Internet archive service to make the Internet persistent. Generally, a user enters a dated URL, for example, using the browser 100, which is transmitted to a web server, such as the persistent web server 140. As discussed below, the disclosed Web browser 100 can optionally include a mechanism to facilitate the specification of the desired date and time, or the user can manually append the time stamp to the URL indicated in the "Location" window of the browser. The Web server 140 receives the URL (containing a time stamp, a relative time-stamp or a variable time-stamp), and retrieves the correct Web page(s) from its own archive 145, or redirects the user to a predefined Internet archive site, such as the Internet archive service 160. The Internet archive service 160 uses the dated URL to retrieve the desired resource (or a link to the resource).

ILLUSTRATIVE TIME STAMP FORMATS

The time stamp can be included in the Uniform Resource Locator ("URL") in any suitable format, as would be apparent to a person of ordinary skill. For example, to refer to the web page, www.Lucent.com, as it existed on February 2, 1998, the URL can be represented as:

http://www.lucent.com?time=2+2+1998.

In a further variation, additional time granularity can be indicated by including the time-of-day in the URL. For example, the web page, www.Lucent.com, as it existed at 1:23 p.m. on February 2, 1998, the URL can be represented as:

http://www.lucent.com?time=13+23+2+2+1998.

Unless otherwise specified, the time zone is assumed to be the user's default time zone. The illustrative time stamp format described above is a Common Gateway Interface (CGI) search argument. Of course, the month, day and year (or other time units) can be expressed in any order. For a URL without a time stamp, the default value will be the most recent version.

In addition, relative time stamps can be utilized. For example, to refer to the web page, www.Lucent.com, as it existed yesterday, the URL can be represented as:

http://www.lucent.com?time=yesterday.

Furthermore, if an embedded hyperlink contains a relative time stamp, the relative time stamp is based on the current web page. Thus, if a current web page has a URL in the form:

http://www.lucent.com?time=1998_1_10

and this page contains an embedded hyperlink in the form:

http://www.lucent.com?time=yesterday

The browser and server will interpret and translate the URL as

http://www.lucent.com?rtime=1998_1_9.

Other relative time stamps can include time offsets from the time of the current web page, such as plus or minus a specified period of time. For example, "+10D" can indicate plus ten days to the time of the currently viewed web page.

In order to refer to the previous or subsequent archived version of a document (relative to the time stamp of the current document), the URL can be represented using the labels "next_archive," or "previous_archive." In another variation, the first or most recent archived version of a document can be represented using the labels "first_archive," or "final_archive," respectively. The server will search through the archive to find the required document. For example, if a currently viewed document has a URL in the form:

http://www.lucent.com/doc.html?rtime=1997_5_12

and there are different versions of the same document archived on 7/2/97, 6/1/97, 4/1/97 and 3/1/97. The following link can be used in the current document to refer to one of these archives:

doc.html?rtime=next_archive

These relative archive time stamps make moving between different version of the same documents more efficient.

A time base parameter can be used to specify the reference date for the relative timestamp. For example,

timbase=1998_6_11&time=+100D

indicates the date that is 100 days after June 11, 1998. Similarly,

timbase=1998_6_11&time=next_Monday

indicates the date that is the Monday after June 11, 1998.

According to another feature of the present invention, wildcard characters and date ranges in the time stamp can be used to implement a variable time stamp in a URL when a user is not sure of the date for a specific web resource. In this manner, the server can display a list of the specific web resources

that match the time stamp pattern. In the illustrative implementation, the following time stamp patterns are used:

TIME STAMP PATTERN	MEANING
*	wildcard character matching 0 or more digits
.	wildcard character matching one digit
m--n	from m to n
,	or
( )	character grouping
---	date range specifier to specify a range between two dates (either absolute or relative dates)

The above time stamp patterns can be used in the time= or rtime= (relative time stamp) fields of the URL to specify, for example, an unknown year, month, day, hour, minute or second. For example,

EXAMPLE	MEANING
199(1,2)	1991 or 1992
199(0-9)	1990 through 1999
19(2-3)(4-5)	1924, 1925, 1934, or 1935
19(20-30, 88, 90)	1920 to 1930, 1988, or 1990
190.	1900 to 1909
19..	1900 to 1999
19*	19, 190 to 199, 1900 to 1999, ...

Thus, the time stamp patterns can be used to request a list of resources having a matching time stamp. For example

EXAMPLE	MEANING
http://www.a.com/res.html?rtime=1998_*	all res.html pages in 1998
http://www.a.com/res.html?rtime=1998_10_*	all res.html pages in October, 1998
http://www.a.com/res.html?rtime=*	all archived res.html pages
http://www.a.com/res.html?rtime=1998_10_(2,3)	all res.html pages on October 2 and 3, 1998

When a server receives a URL request containing a variable time stamp, the server recognizes that the client is requesting a list of different versions of the same resource. The server will search through all the archives to identify all matched resources and return an HTML page with hyperlinks pointing to all matched resources. Since some web resources might have many versions archived, the user can optionally specify how to present the links. Thus, according to a further feature of the invention, a "timeorder" parameter allows the user to specify how to display the links corresponding to the matched resources. For example, timeorder=increase will present the links in increasing time order.

In addition, the links can be presented in a calendar-like format for easy navigation and selection. For example, if the links for the matching resources expand through several years, the links can be displayed in the following manner, with the month number underlined to indicate existing versions of the matching resources:

1996	1	2	3	4	5	6	7	8	9	10	11	12
1997	1	2	3	4	5	6	7	8	9	10	11	12
1998	1	2	3	4	5	6	7	8	9	10	11	12

The URL corresponding to the link for March, 1998 would have the form http://www.a.com/res.html?rtime=1998_3_*&timeorder=calendar. The time order can also be specified in terms of units of time. For example, timeorder=+D means to list the links of the matching resources in increasing day order, with the links corresponding to the first day of each month first, followed by links corresponding to the second day of each month and so on. Likewise, timeorder=+DY means to

list the links of the matching resources in increasing day order, then in increasing year order.

Relative time stamps can be extended using the "*" wildcard at the end of the time value. For example, if the current day is December 9, 1998, rtime=next_month* means rtime=1999_1_* (any day in January, 1999). Similarly, if the current day is December 9, 1998, rtime=+1y* means rtime=1999_12_9_* (any time on December 9, 1999).

In an alternate implementation, referred to herein as the "request-header scheme," the time stamp can be indicated as one of the HTTP request headers, such as:

Time-Stamp: June 9, 1998.

In another embodiment, referred to herein as the " special character scheme," special characters can be utilized to indicate the inclusion of a time stamp in the URL, such as:

http://www.cnnfn.com;time=2+2+1998&timezone=server.

WEB BROWSER

The Web browser 100 may be embodied as a conventional browser, such as Microsoft Internet Explorer™ or Netscape Navigator™, as modified herein to incorporate the features and functions of the present invention. As discussed further below, the Web browser 100 only needs to incorporate a new options selection panel to permit the user to specify the desired date and time. In fact, a conventional Web browser 100 can be utilized, with the user manually appending the time stamp to the URL indicated in the "Location" window of the browser 100.

In one implementation, the user has the option to turn the time stamp on or off. If the time stamp is activated, the browser 100 will change the URL accordingly before sending the URL out to the Web 130. Since there is no guarantee that the corresponding web server 140, 150 recognizes a time stamp, the

document returned by the server 140, 150 might contain embedded hyperlinks that do not contain time stamps. Thus, in this situation, the web browser 100 can automatically convert the URL associated with an embedded hyperlink to add an appropriate time stamp when the user clicks on the hyperlink if the time stamp option is activated. The Web browser 100 should convert the URL in accordance with the selected time stamp format. In a request-header-scheme implementation, the browser 100 should be modified to send the special request header ("Time-Stamp: June 9, 1998"). In addition, the HTML should be modified to include a new time stamp tag for any embedded hyberlink with a specific time stamp. For example, for a hyperlink such as:

<A HREF="www.lucent.com">Lucent Web Site</A>

the HTML should be modified to indicate the time stamp of Feb. 2, 1998 as follows:

<A HREF="www.lucent.com"><TIMESTAMP TIME="2+2+1998"

TIMEZONE=server></TIMESTAMP>Lucent Web Site</A>.

PERSISTENT WEB SERVERS

The persistent Web servers 140, 150 may be embodied as conventional hardware and software, as modified herein to carry out the functions and operations described below. Specifically, the persistent Web servers 140, 150 need to know how to (i) receive URLs containing a time stamp or relative time-stamp, (ii) extract the time stamp, (iii) retrieve the Web page corresponding to the appropriate time-stamp, and (iv) return the requested page to the client. The persistent Web servers 140, 150 should interpret the extracted URL in accordance with the selected time stamp format.

If a version of the Web resource corresponding to the requested time does not exist, the present invention provides a version of the document stored time-wise in the vicinity of the requested target time. For example, the

present invention may assume the Web resource has not changed from the previous archived version, and the version of the Web resource with the most recent time-stamp preceding the requested time is provided. Alternatively, the version of the Web resource with the next immediate time-stamp after the requested time is provided.

In addition, the persistent Web servers 140, 150 need to preserve all the information in their history of serving the Web. Thus, as shown in FIG. 1, each persistent Web server, such as the servers 140, 150, includes a persistent archive 145, 155, respectively, for storing all of the versions of Web resources that will be persistently available to Web users. The persistent archives 145, 155 may be embodied as any storage device, although a persistent (non-erasable) storage device such as CD-ROM, CR-R, WORM or DVD-ROM may be preferred.

For the persistent Web servers 140, 150 to support dated URLs, the persistent Web servers 140, 150 need to store all of their contents in a chronicle fashion to enable the retrieval of timely information. In one implementation, shown in FIG. 2, the persistent archives 145, 155 store the entire web site contents on permanent storage devices according to some sort of chronological directory structure. FIG. 2 shows a directory structure 200 that arranges the contents of the Web site chronologically. Thus, each leaf, such as the leaf 210, in the directory structure 200 corresponds to a dated URL. For example, a dated URL such as:

http://www.nytimes.com?time=24+2+1998

is conceptually equivalent to:

http://www.nytimes.com/archive/1998/2/24/.

Of course, storing the entire web site contents is inefficient in terms of storage usage. Many Web pages exhibit few, if any, changes from day to day. Thus, significant storage efficiencies can be achieved by simply removing

redundancy in the archive. Once the redundancy is removed, the storage requirement in addition to the regular web site storage is usually not very large.

FIG. 3 illustrates an archival process 300 for reducing the redundancy of the persistent archive 145, 155. All the files or subdirectories mentioned in the algorithm are under the archive subdirectory 220 of the illustrative directory structure 200 of FIG. 2. As shown in FIG. 3, the archival process 300 initially performs a test during step 310 for each subdirectory, such as subdirectory A, to determine whether there exists a subdirectory B that is created earlier and has identical contents as subdirectory A. If it is determined during step 310 that there is no subdirectory B created earlier and having identical contents as subdirectory A, then it is not possible to reduce the redundancy on the subdirectory level of the persistent archive 145, 155 and program control proceeds to step 330.

If, however, it is determined during step 310 that there exists a subdirectory B that is created earlier and has identical contents as subdirectory A, then subdirectory A becomes an alias during step 320 pointing to subdirectory B. For example, as shown in FIG. 4A, if the current contents of a Web site is identical to the contents of the previous day, an alias is created for today pointing to yesterday's subdirectory. Likewise, as shown in FIG. 4B, if the current month's content is the same as the contents of the previous month, an alias is created for this month pointing to last month's subdirectory.

Thereafter, a test is performed during step 330 for each file, such as file A, to determine whether there exists a file B that is created earlier and has identical contents as file A. If it is determined during step 330 that there is no file B created earlier and having identical contents as file A, then it is not possible to reduce the redundancy of the persistent archive 145, 155 on the file level. Thus, program control terminates during step 350.

If, however, it is determined during step 330 that there exists a file B that is created earlier and has identical contents as file A, then file A becomes an alias during step 340 pointing to file B. Thereafter, program control terminates during step 350.

The archival process 300 may be impractical, since it needs to search for match files or directories. The run time increases exponentially with the number of entities in the archive. Many sub-optimal solutions are possible, as would be apparent to a person of ordinary skill in the art. A very simple solution is just checking what you want to archive today against the most recently added archive (like yesterday's contents). Since most of the web sites only differ from their previous archived ones slightly, this approach is quite reasonable. This approach is similar to the well-known incremental backup of a file system.

If a Web server is not persistent, it should only have minimal impact. In one embodiment, if a request includes a time stamp that is not recognized by a Web server, the server should deliver the most recent version of the requested Web resource.

Another way to reduce storage requirements of the persistent archive is to make the Web server smarter in terms of searching the correct archived data. For example, persistent storage of a web resource can be limited to versions that have some difference relative to previously saved versions of the web resource. For example, if an illustrative archive contains the following five different versions of a web resource: 6/4/1996, 6/12/1996, 3/23/1997, 2/1/1998 and 2/3/1998, the web server assumes that if the requested date does not equal any of the archived versions, then the requested date is identical to the version with the closest earlier date. In addition, a special symbolic link (or alias on MacOS, short cut on MS Windows) can be used in a directory to represent where to looks for files or directories that are not found under the current directory. In this manner,

only the changed parts are stored under appropriate directories. All the unchanged data can be referred through a chain of such special links.

DNS SERVER

The domain name server (DNS) may be embodied as conventional hardware and software, as modified herein to carry out the functions and operations described below. Conventional DNS servers will reject any domain name reference which is not in the DNS database. One benefit of dated URL in accordance with the present invention is that it can be used to refer to historical Web resources. For example, if company A is merged into company B, all the web pages referred through "www.A.com" may no longer be valid. For users who want to access some documents from company A, they need to change all the reference to some place in company B's web site.

The historical information of company A can still be accessed if the DNS server does not reject the name reference, but instead consults an archive service company that knows where the historical information of company A is located. The DNS server itself can also store some historical data to resolve the name to IP address process faster. FIGS. 5A and 5B provide examples of data stored in a DNS server database before and after the merger of companies A and B, respectively. As shown in FIG. 5B, if a user wants to find www.A.com after the merger, the DNS server has enough information to redirect the user's request to a new IP address associated with company B. The dates listed in the database are the valid periods for the corresponding domain name. Thus, a dated domain name reference like "www.A.com 2/2/1999" is invalid, while "www.A.com 2/2/1992" is valid.

FIG. 6 illustrates a DNS server process 600 in accordance with the present invention. As shown in FIG. 6, the DNS server process 600 initially receives a domain name request during step 610. A test is performed during step 620 to determine if the domain name request is dated. If it is determined during

step 620 that the domain name request is not dated, the regular name searching result is returned during step 630.

If, however, it is determined during step 620 that the domain name request is dated, the DNS server process 600 searches the DNS database for the domain name with the date constraint during step 640. A further test is performed during step 650 to determine if the dated domain name is found. If it is determined during step 650 that the dated domain name is not found, then the DNS server consults with an archive service company during step 660 for further searching before program control proceeds to step 670.

If, however, it is determined during step 650 that the dated domain name is not found, then the searching result and indication, if redirect, are returned during step 670, before program control terminates.

After the domain name is resolved by the DNS server, the Web browser 100 needs to send the request to the web server 140, 150 according to what is returned from the DNS server. For example, a request from the user for the following URL, "http://www.A.com?time=2+2+1992," will cause the browser 100 to send a domain name resolving request to the DNS server in a format such as "www.A.com 2/2/1992." Since company A is now part of company B, the results will look like "123.2.3.222 redirect." The Web browser 100 now has the IP address of the server and also knows it is a redirect one. Thus, the Web browser 100 will effectively send a request to the Web server 140, 150 of Company B in a form such as "http://123.2.3.222?http://www.A.com&time=2+2+1992." The Web server 140, 150 of Company B will know how to map this old address of company A's to the appropriate place and get the correct information.

One side benefit of this new DNS server is that some names can be reused once they are history. For example, another company named Company A

can utilize the www.A.com domain name after a predefined period, by updating the DNS database with the following entries:

URL	STARTING DATE	ENDING DATE	IP ADDRESS	REDIRECTED URL
WWW.A.COM	13/2/1980	1/5/1998	123.2.3.222	WWW.B.COM
WWW.A.COM	1/1/2000	PRESENT	234.2.2.12	--
WWW.B.COM	23/1/1985	PRESENT	123.2.3.222	--

In this manner, domain names can be reused without wasting them forever.

PERSISTENT ACCESS TO DYNAMIC WEB CONTENT

The Web is now full of dynamic content, including real time video, for example, from a WebCam, and audio streams, for example, from a WebCast event, as well as Java, Javascript or Active-X enabled web pages. Depending on the application, it may not be necessary or desirable to archive the continuous dynamic content. In addition, it may not be necessary or desirable to archive all the advertising portions of a web pages.

For chronological data, such as bank or stock broker transactions, it is easy to extract part of the record for a given time stamp restriction. For example, to check the account balance of a give date, the browser 100 can send a request in the form:

http://bank.com/?id=12345&time=2+3+1998.

The server 140, 150 only needs to retrieve or recalculate the data up to March 2, 1998 and return the results. Since all the transactions in such application environments have time stamps anyway, it is straightforward to add this function to the service.

For real time contents, the only restriction in appending a time stamp is the storage requirement. If a lot of storage space is available compared to the amount of information to be archived, the Web site administrator can choose to

archive the real time contents or to archive some of them such as one day, one week or one year's worth of data.

For dynamically created advertisements, the Web site administrator must decide whether it is reasonable to 'reshow' the old advertisement (for some special reason) or whether the old advertisement can be replaced with a new, up-to-date commercial which is not relevant to the 'real' archived web contents.

For a dynamically executed script on a Web page, there are two possible situations. If the script is not related to any time function, then the script can be archived safely. If, however, the script will change its behavior according to some time function, such as a clock or calendar display, then the script needs to be modified to deal with dated URL, requiring modifications to Java, Javascript or Active-X standards, as appropriate. Web pages generated dynamically by CGI applications are treated in a similar manner.

Internet Archive Service

The present invention provides persistent Web access, without requiring all web servers to directly handle time-stamps in a URL. Generally, a user enters a dated URL, for example, using a web browser, in the manner described above. The browser transmits the dated URL to the appropriate web server. As previously indicated, the disclosed Web browser can optionally include a mechanism to facilitate the specification of the desired date and time, or the user can manually append the time stamp to the URL indicated in the "Location" window of the browser. The Web server receives the URL (containing a time stamp, a relative time-stamp or a variable time-stamp). Thereafter, the Web server retrieves the correct Web page(s) from its own archive, or redirects the user to a predefined Internet archive site, in accordance with the present invention.

For example, a user enters the following dated URL to access the main page of a news web site for the last day of 1999:

http://www.news-site.com/?rtime=1999_12_31

In accordance with the present invention, the news web site will redirect the user request to the following URL:

http://www.archive.com/news-site.com?rtime=1999_12_31

The Internet archive service that hosts the archive for the news web site will know the redirected URL is for content from the news web site and will retrieve the appropriate resource.

FIG. 7A illustrates the various communications between the user's browser 100, the web server 140 and the Internet archive service 160, in an embodiment where the web server 140 knows the corresponding Internet archive service 160, where the requested resource has been archived. As shown in FIG. 7A, the user initially sends a request 710 (using the browser 100) to the web server 140. The web server 140 recognizes the request is for an archived resource, and returns a URL redirect response 720. The browser 100 uses the URL redirect response 720, such as the following example provided above:

http://www.archive.com/news-site.com?rtime=1999_12_31,

to contact the Internet archive service 160, using a message 730. The Internet archive service 160 uses the dated URL to retrieve the desired resource or to form a page of links for the corresponding resources, in the case of an Internet query, which is returned to the browser 100 in a message 740.

FIG. 7B illustrates the various communications between the user's browser 100, the web server 140, an archive finder 750, and the Internet archive service 160, in an embodiment where the web server 140 does not know the corresponding Internet archive service 160 where the requested resource has been archived. As shown in FIG. 7B, the user initially sends a request 755 (using the browser 100) to the web server 140. The web server 140 recognizes the request is for an archived resource, and sends a request 760 to the archive finder 750 to identify the particular Internet archive service 160 where the requested resource has been archived. The archive finder 750 evaluates the request and finds the best

Internet archive service 160 for the request according to various criteria, such as site load, site location, and nature of the request. The archive finder 750 returns the identity of the best Internet archive service 160 to the web server 140 in a message 765. The web server 140 then forwards the identified best Internet archive service 160 to the browser 100 as a URL redirect response 770. The browser 100 uses the URL redirect response 770, such as the following example provided above:

http://www.archive.com/news-site.com?rtime=1999_12_31,

to contact the identified Internet archive service 160, using a message 775. The identified Internet archive service 160 uses the dated URL to retrieve the desired resource or to form a page of links for the corresponding resources, in the case of an Internet query, which is returned to the browser 100 in a message 780.

It is to be understood that the embodiments and variations shown and described herein are merely illustrative of the principles of this invention and that various modifications may be implemented by those skilled in the art without departing from the scope and spirit of the invention.

The claims defining the invention are as follows:

~~Claims~~:

1. A method for providing an electronic document, said electronic document having multiple versions, each of said versions having a time-stamp, said method comprising the steps of:

receiving a request for one of said versions of said electronic document, said request including an indication of said time-stamp; and

transmitting a network address of a server where said requested version of said electronic document is archived.

2. The method according to claim 1, further comprising the step of storing a pointer to a previous version of said electronic document when said previous version of said electronic document is substantially identical.

3. The method according to claim 1, wherein an address identifying said electronic document includes said time-stamp.

4. The method according to claim 3, wherein said address is a Uniform Resource Locator ("URL").

5. The method according to claim 4, wherein said Uniform Resource Locator ("URL") has an associated request header for indicating said time stamp.

6. The method according to claim 1, wherein said request is received from a browser.

7. The method according to claim 1, wherein said transmitted network address is a URL redirect request that initiates a receiver of said network address to obtain said requested version of said electronic document from said server where said requested version of said electronic document is archived.

8. The method according to claim 1, wherein said transmitting step further comprises the step of transmitting a redirect URL.

9. The method according to claim 1, further comprising the step of identifying said server where said requested version of said electronic document is archived.

10. The method according to claim 9, wherein said identifying step further comprising the step of identifying a server satisfying predefined criteria.

11. A system for storing an electronic document having multiple versions, said system comprising:

a memory for storing computer readable code; and

a processor operatively coupled to said memory, said processor configured to:

receive a request for one of said versions of said electronic document, said request including an indication of said time-stamp; and

transmit a network address of a server where said requested version of said electronic document is archived.

12. The system according to claim 11, wherein said processor is further configured to store a pointer to a previous version of said electronic document when said previous version of said electronic document is substantially identical.

13. The system according to claim 11, wherein an address identifying said electronic document includes said time-stamp.

14. The system according to claim 13, wherein said address is a Uniform Resource Locator ("URL").

15. The system according to claim 14, wherein said Uniform Resource Locator ("URL") has an associated request header for indicating said time stamp.

16. The system according to claim 11, wherein said request is specified using a browser.

17. The system according to claim 11, wherein said transmitted network address is a URL redirect request that initiates a receiver of said network address to obtain said requested version of said electronic document from said server where said requested version of said electronic document is archived.

18. The system according to claim 11, wherein said transmitting step further comprises the step of transmitting a redirect URL.

19. The system according to claim 11, wherein said processor is further configured to identify said server where said requested version of said electronic document is archived.

20. The system according to claim 19, wherein said processor is further configured to identify said server satisfying predefined criteria.

21. An article of manufacture for accessing an electronic document, said article of manufacture comprising:

a computer readable medium having computer readable program code means embodied thereon, said computer readable program code means comprising program code means for causing a computer to:

receive a request for one of said versions of said electronic document, said request including an indication of said time-stamp; and

transmit a network address of a server where said requested version of said electronic document is archived.

22. A method of providing an electronic document, said method being substantially as described herein in relation to any one embodiment with reference to Figs. 1-7B.

23. A system for storing an electronic document, said system being substantially as described herein in relation to any one embodiment with reference to Figs. 1-7B.

24. An article of manufacture for assessing an electronic document, said article of manufacture being substantially as described herein in relation to any one embodiment with reference to Figs. 1-7B.

DATED this Twenty-second Day of May, 2001

Lucent Technologies Inc.

Patent Attorneys for the Applicant

SPRUSON & FERGUSON

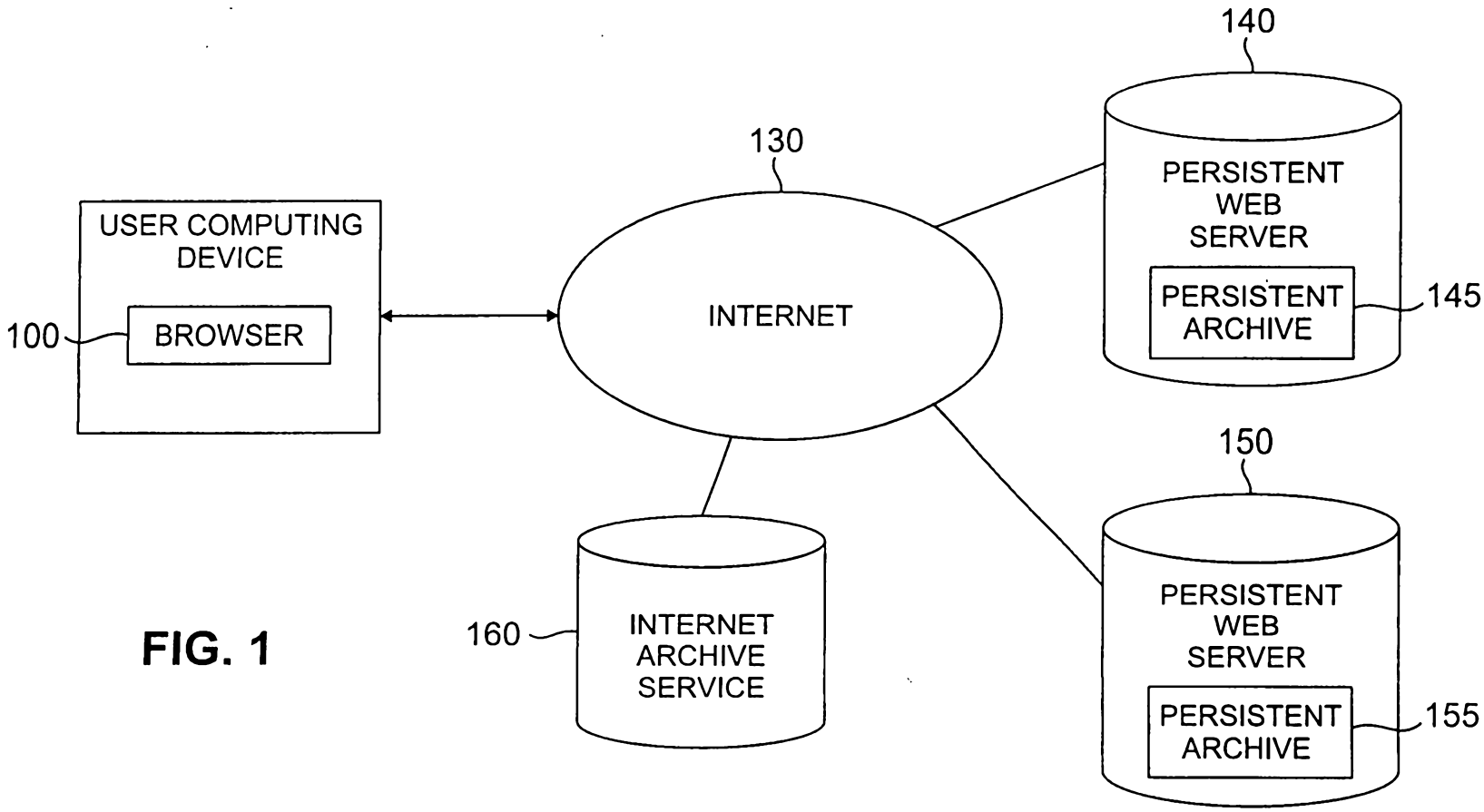
140
130
USER COMPUTING DEVICE
100
BROWSER
INTERNET
PERSISTENT WEB SERVER
PERSISTENT ARCHIVE
145
150
160
INTERNET ARCHIVE SERVICE
PERSISTENT WEB SERVER
PERSISTENT ARCHIVE
155

FIG. 1

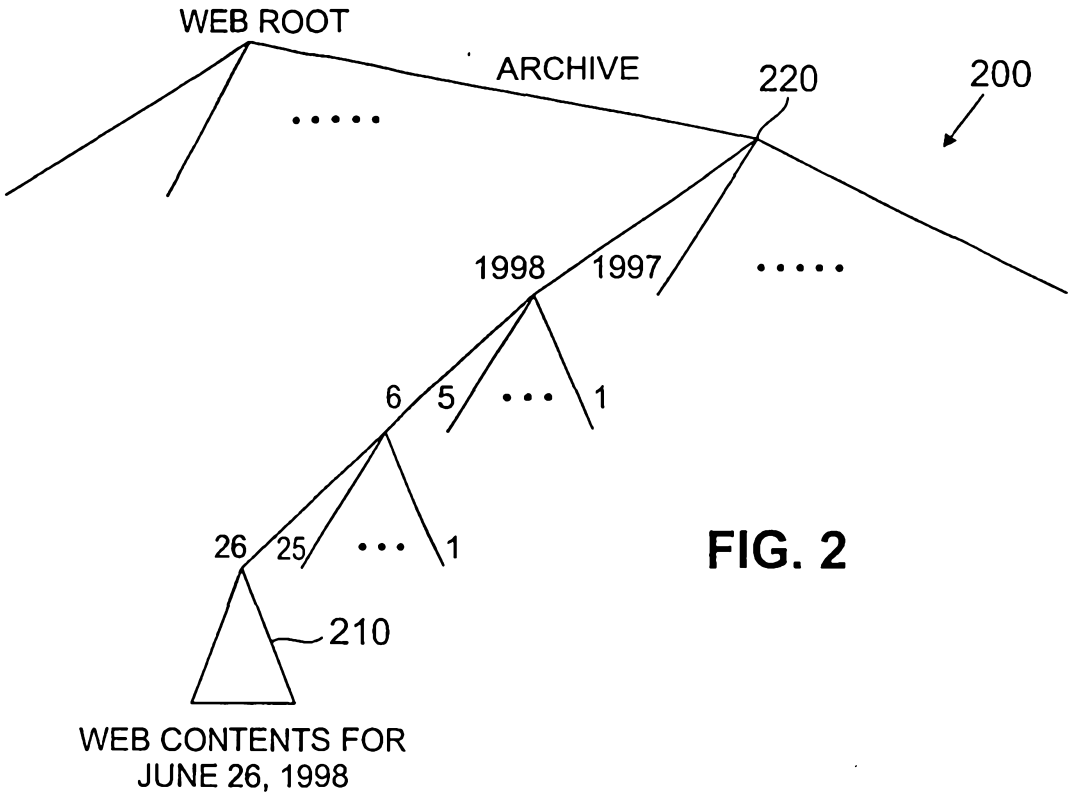

FIG. 2

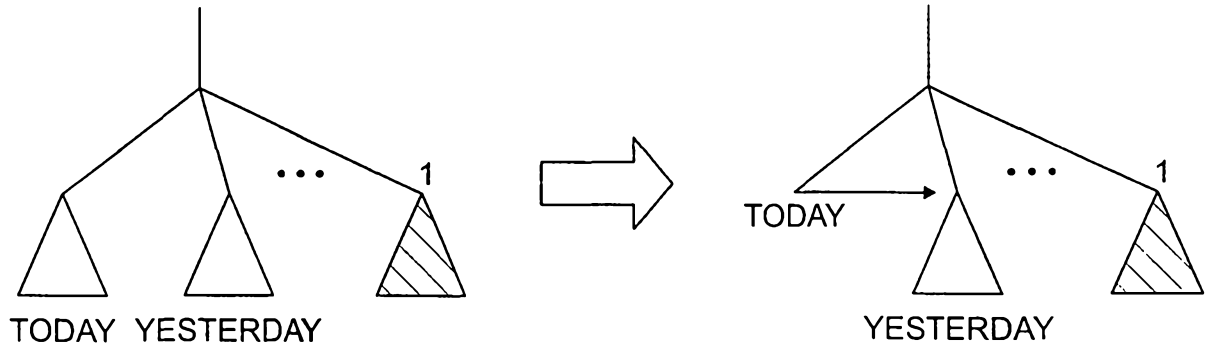

FIG. 4A

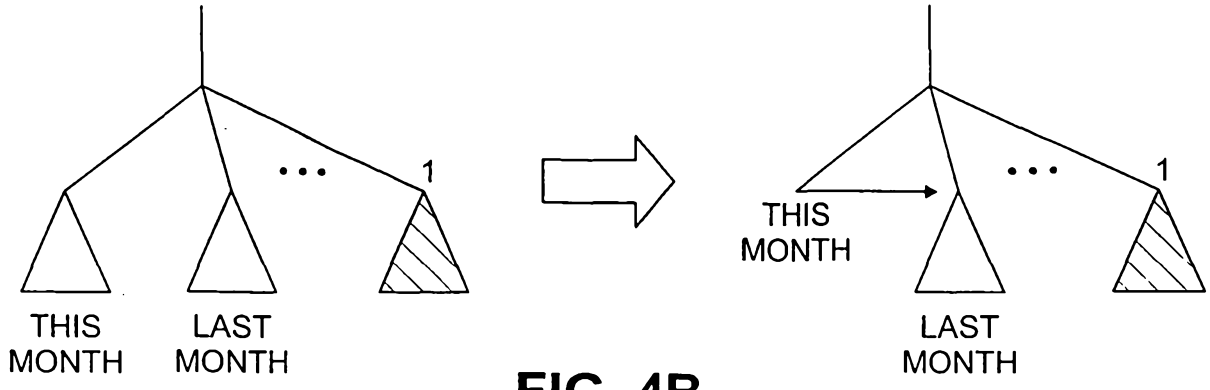

FIG. 4B

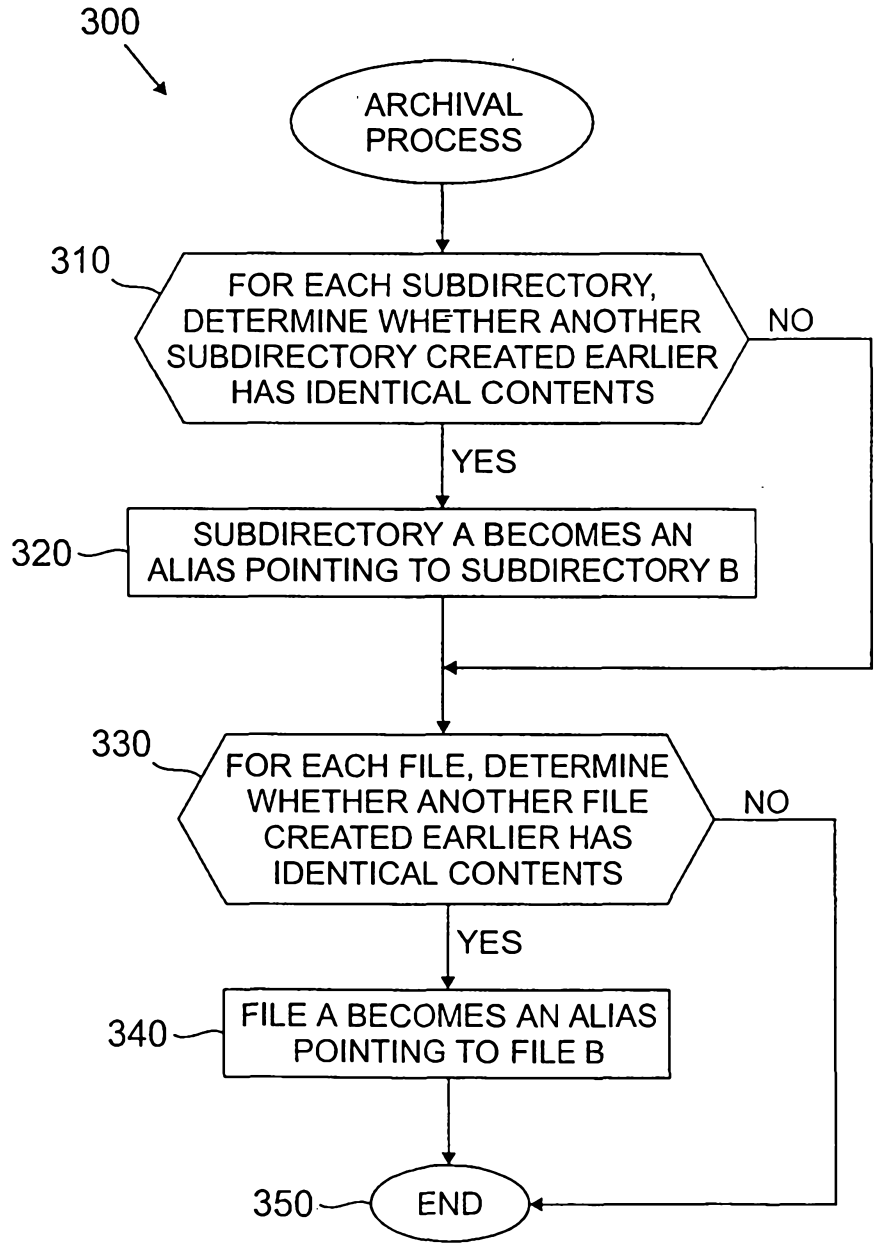
300
ARCHIVAL PROCESS
310
FOR EACH SUBDIRECTORY, DETERMINE WHETHER ANOTHER SUBDIRECTORY CREATED EARLIER HAS IDENTICAL CONTENTS
NO
YES
320
SUBDIRECTORY A BECOMES AN ALIAS POINTING TO SUBDIRECTORY B
330
FOR EACH FILE, DETERMINE WHETHER ANOTHER FILE CREATED EARLIER HAS IDENTICAL CONTENTS
NO
YES
340
FILE A BECOMES AN ALIAS POINTING TO FILE B
350
END

FIG. 3

URL	STARTING DATE	ENDING DATE	IP ADDRESS	REDIRECTED URL
WWW.A.COM	13/2/1980	PRESENT	222.1.33.22	--
WWW.B.COM	23/1/1985	PRESENT	123.2.3.222	--

FIG. 5A

URL	STARTING DATE	ENDING DATE	IP ADDRESS	REDIRECTED URL
WWW.A.COM	13/2/1980	1/5/1998	123.2.3.222	WWW.B.COM
WWW.B.COM	23/1/1985	PRESENT	123.2.3.222	--

FIG. 5B

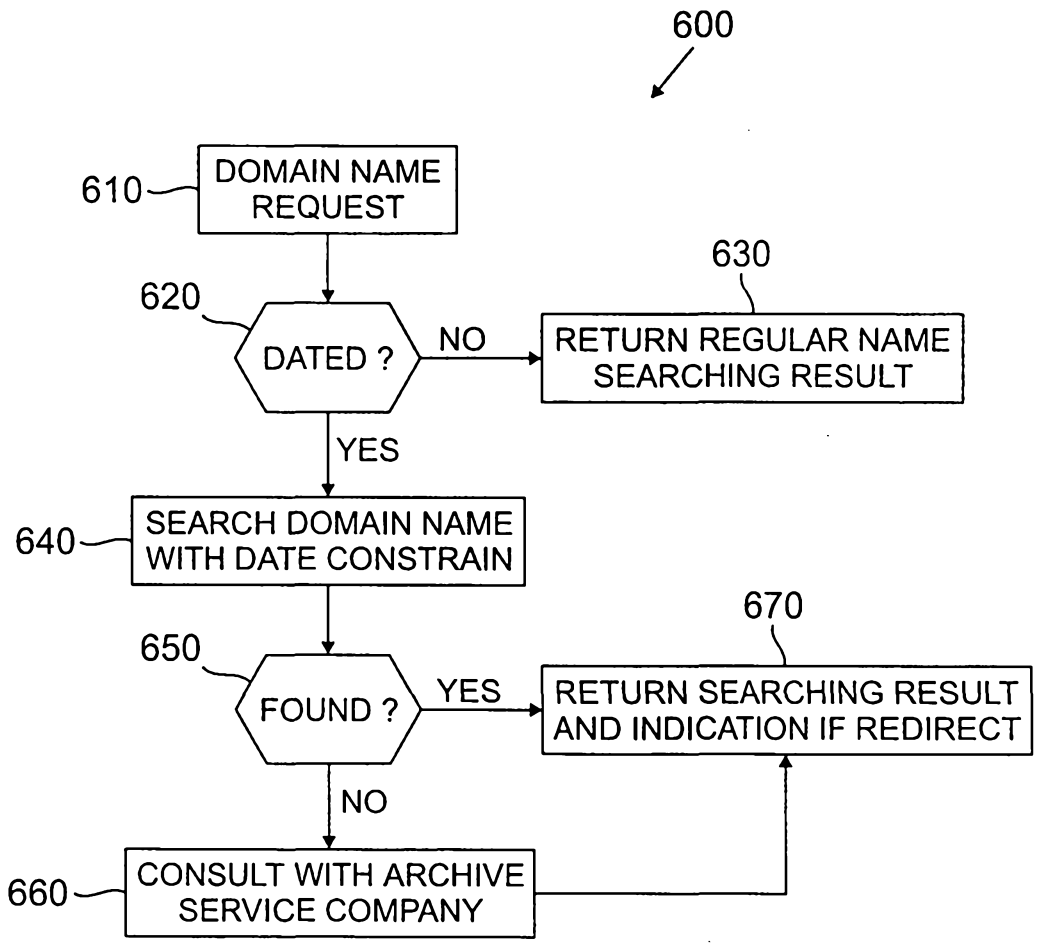
600
610
DOMAIN NAME REQUEST
620
DATED ?
NO
630
RETURN REGULAR NAME SEARCHING RESULT
YES
640
SEARCH DOMAIN NAME WITH DATE CONSTRAIN
650
FOUND ?
YES
670
RETURN SEARCHING RESULT AND INDICATION IF REDIRECT
NO
660
CONSULT WITH ARCHIVE SERVICE COMPANY

FIG. 6

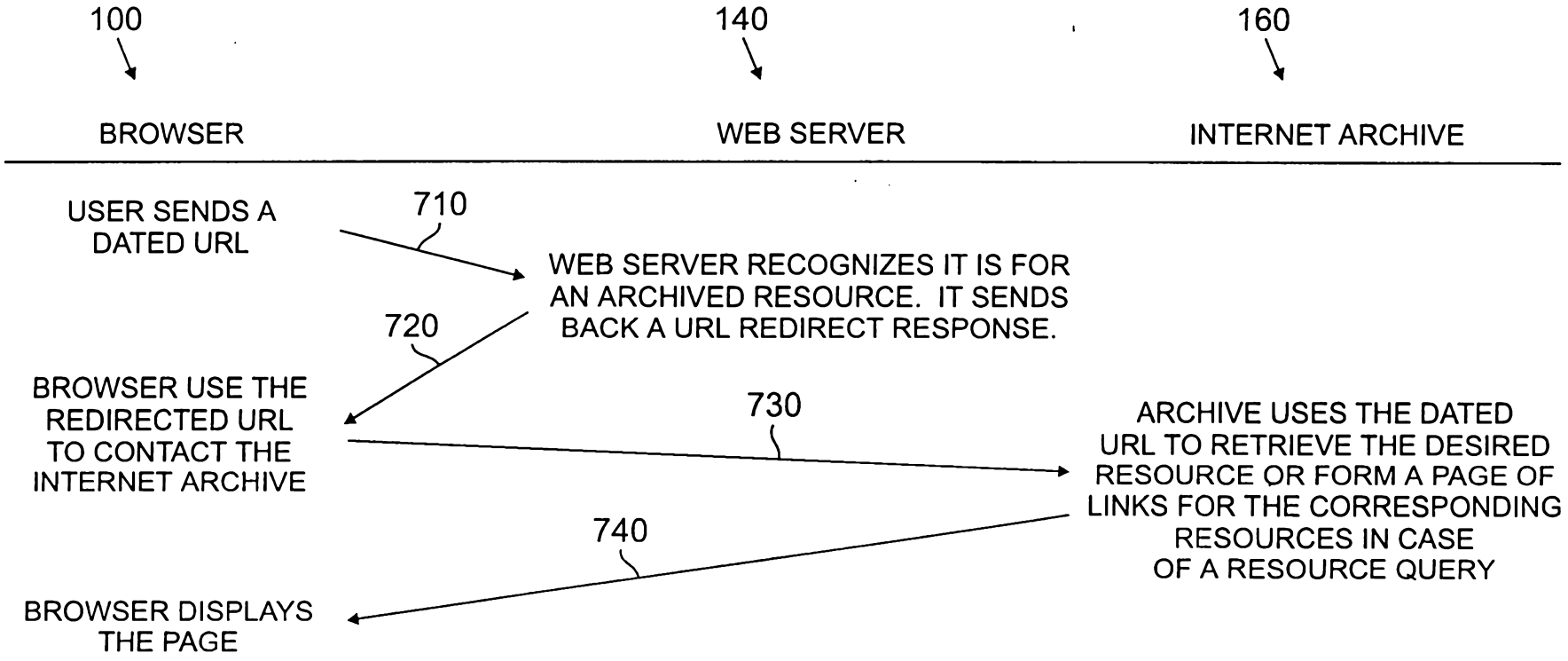
100
140
160
BROWSER
WEB SERVER
INTERNET ARCHIVE
USER SENDS A DATED URL
710
WEB SERVER RECOGNIZES IT IS FOR AN ARCHIVED RESOURCE. IT SENDS BACK A URL REDIRECT RESPONSE.
720
BROWSER USE THE REDIRECTED URL TO CONTACT THE INTERNET ARCHIVE
730
ARCHIVE USES THE DATED URL TO RETRIEVE THE DESIRED RESOURCE OR FORM A PAGE OF LINKS FOR THE CORRESPONDING RESOURCES IN CASE OF A RESOURCE QUERY
740
BROWSER DISPLAYS THE PAGE

FIG. 7A

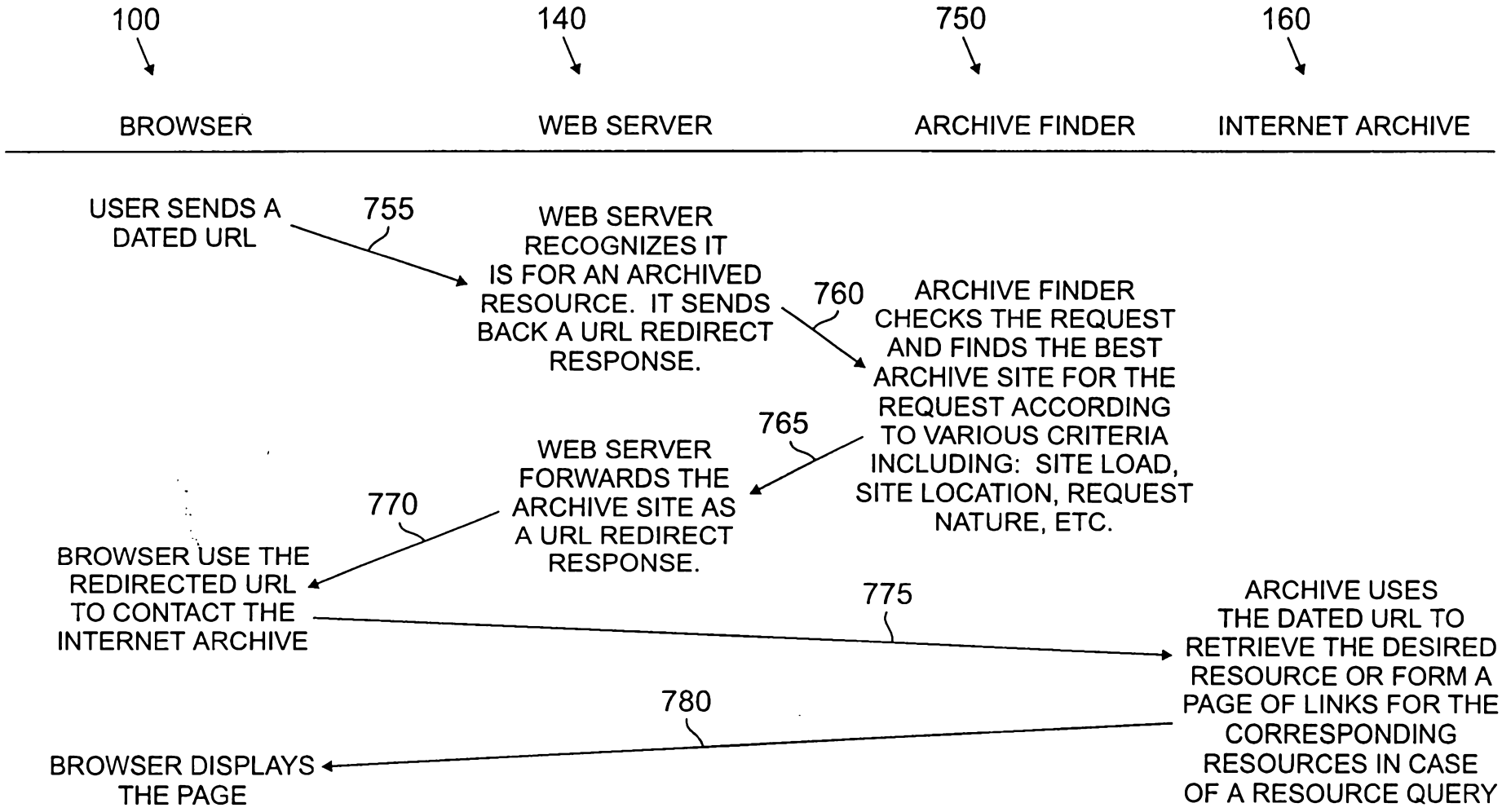
100
140
750
160
BROWSER
WEB SERVER
ARCHIVE FINDER
INTERNET ARCHIVE
USER SENDS A DATED URL
755
WEB SERVER RECOGNIZES IT IS FOR AN ARCHIVED RESOURCE. IT SENDS BACK A URL REDIRECT RESPONSE.
760
ARCHIVE FINDER CHECKS THE REQUEST AND FINDS THE BEST ARCHIVE SITE FOR THE REQUEST ACCORDING TO VARIOUS CRITERIA INCLUDING: SITE LOAD, SITE LOCATION, REQUEST NATURE, ETC.
765
WEB SERVER FORWARDS THE ARCHIVE SITE AS A URL REDIRECT RESPONSE.
770
BROWSER USE THE REDIRECTED URL TO CONTACT THE INTERNET ARCHIVE
775
ARCHIVE USES THE DATED URL TO RETRIEVE THE DESIRED RESOURCE OR FORM A PAGE OF LINKS FOR THE CORRESPONDING RESOURCES IN CASE OF A RESOURCE QUERY
780
BROWSER DISPLAYS THE PAGE

FIG. 7B