

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 July 2007 (19.07.2007)

PCT

(10) International Publication Number  
WO 2007/080171 A1

- (51) International Patent Classification:  
G06F 17/30 (2006.01)
- (21) International Application Number:  
PCT/EP2007/050213
- (22) International Filing Date: 10 January 2007 (10.01.2007)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
11/330,485 12 January 2006 (12.01.2006) US
- (71) Applicant (for all designated States except US): **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, NY 10504 (US).
- (71) Applicant (for MG only): **IBM UNITED KINGDOM LIMITED** [GB/GB]; PO Box 41, Portsmouth Hampshire PO6 3AU (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **PALAPUDI, Sri-ram** [IN/US]; 2352 Nobili Avenue, Santa Clara, CA 95051 (US). **RAJAKANNIMARIYAN, Maria, Savarimuthu** [IN/US]; 3579 Klaus Drive, San Jose, CA 95121 (US). **SHANMUGAM, Ravisankar** [IN/US]; 77 Limewell

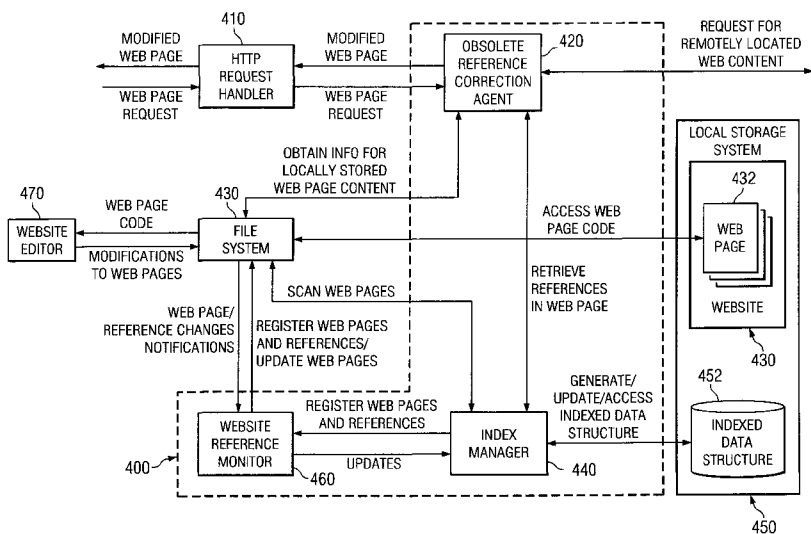
Court, San Jose, CA 95138 (US). **WOLAFKA, Rainer** [DE/DE]; Erlenweg 6, 65812 Bad Soden (DE).

- (74) Agent: **WATSON, Justine, Nicola**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester Hampshire SO21 2JN (GB).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**  
— with international search report

[Continued on next page]

(54) Title: MECHANISM TO TRAP OBSOLETE WEB PAGE REFERENCES AND AUTO-CORRECT INVALID WEB PAGE REFERENCES



(57) Abstract: A mechanism to trap obsolete web page references and auto-correct invalid Web page references is provided. With the mechanism, Web pages of a Web site are indexed in an indexed data structure having entries that list the references contained in the Web page. A Website reference monitor monitors changes to the Web pages and content referenced by these Web pages. If a change to the Web pages or referenced content is detected, other Web pages in the Web site that reference the modified content or Web pages are identified using the indexed data structure. The identified other Web pages may then be automatically updated. In addition, when a client device requests a Web page, the references in the Web page are checked to determine if they reference obsolete or invalid content and such references are modified to be non-selectable before providing the Web page to the client device.

WO 2007/080171 A1



---

*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.*

**MECHANISM TO TRAP OBSOLETE WEB PAGE REFERENCES AND  
AUTO-CORRECT INVALID WEB PAGE REFERENCES**

BACKGROUND

5 Technical Field

The present application relates generally to an improved data processing system and method. More specifically, the present application is directed to a mechanism for trapping obsolete Web page references and auto-correct  
10 invalid Web page references.

Description of Related Art

15 Generally, commercial Websites consist of a large amount of static and dynamic content such as Hypertext Markup Language (HTML) content, pictures, graphics, sound and video files, and Web applications. Due to the rapid and frequent changes to Website content, typically on a daily basis, Websites have to be modified accordingly in order to reflect the most up to date information. Such modifications include changing and  
20 relocating the content of the HTML, picture, graphics, audio, and video files, and deleting the old static and/or dynamic files.

Typically, such changes, relocation, and the like, is left up to individuals known as Webmasters. The Webmaster's primary role is to keep  
25 Websites up to date and manage the operation of the Website on a daily basis. When changes are to be made to a Website, it is up to the Webmaster to update the HTML, picture, graphics, audio, video files, and the like and to ensure that all references to the modified or relocated content are properly updated.

30 It can be seen that with rapid and frequent changes to Website content, even with very simple Websites, it may be difficult to completely identify every reference, e.g., hyperlinks and the like, to content that has been changed or relocated. Moreover, at present, web browsers and web servers  
35 do not know whether a reference to Website content is obsolete, i.e. no longer accessible by the reference, or invalid, i.e. not the correct content intended to be accessed by use of the reference, before the user of a client device tries to access the content. As a result, when a reference to content that has been changed or relocated is accessed by a  
40 user, the result may be an error due to the content no longer being

present at the particular location, with the same filename, or the like, identified in the reference. In some instances, such references, after changes to and/or relocating of content files has occurred, may point to the wrong content or out-of-date content, i.e. invalid content. This  
5 problem is made even more troublesome with the more complex Websites typically found in today's electronic businesses.

#### SUMMARY

10 According to a first aspect, there is provided a computer program product comprising a computer useable medium having a computer readable program, wherein the computer readable program, when executed on a computing device, causes the computing device to: generate an indexed data  
15 structure identifying Web pages of the Website and references to content that are present in the Web pages of the Website; receive a modification to content of the Website; search the indexed data structure to identify one or more Web pages of the Website that contain references to the modified content of the Website; and perform at least one operation based  
20 on the identification of the one or more Web pages of the Website that contain references to the modified content, wherein the at least one operation facilitates updating of the references to the modified content in the identified one or more Web pages of the Website.

25 There is preferably provided a mechanism for identifying obsolete or invalid references to Website or Web page content. There is preferably provided a mechanism for automatically correcting obsolete or invalid references in Web pages of Websites based on the identification of such obsolete or invalid references. There is further preferably provided a  
30 mechanism that renders obsolete or invalid references to Website or Web page content non-selectable by users of client devices via their Web browsers. The illustrative embodiments provide such mechanisms.

35 With the mechanisms of the illustrative embodiments, an indexing mechanism is provided for indexing each Web page of a Website and identifying all references to Website content present in the Web pages of the Website. In particular, an index manager is utilized that scans (i.e., crawls) the code of the Web pages of the entire Website and identifies references to Web page content, e.g., hyperlinks, references to image files, graphics files, sound files, video files, etc.. Entries in an indexed data  
40 structure for the Website are created for the Web pages with each entry identifying the references present in the corresponding Web page. The

crawling of the Website may be performed once to establish an initial indexed data structure that is subsequently maintained up-to-date by real time updates when the Website is modified. Alternatively, or in addition, the crawling of the Website may be performed periodically so as to ensure that the indexed data structure is correct.

In one embodiment, the indexed data structure is used to identify obsolete and invalid references to Web content in Web pages of a Website as the Website is modified. The index manager, according to one embodiment, registers the indexed Web pages and their corresponding references with a Website reference monitor that monitors real time modifications to the Website. Such modifications may include, for example, Website content deletion, Website content relocation, Website content renaming, Website content addition, or Web page modifications. The Website reference monitor, according to one embodiment, registers the Websites directory structures and files associated with the references in the Web pages to the operating system's file system so as to obtain real time updates regarding these directory structures and files from the file system.

That is, when a change to a registered directory or file occurs, e.g., the deletion, relocation, renaming or addition of a file or directory, the file system preferably notifies the Website reference monitor of this change. The Website reference monitor may then scan the indexed data structure to identify all references in all Web pages of the Website to the changed file or directory and may update these references accordingly in the code of these other Web pages. In addition, the indexed data structure may be updated to reflect the up-to-date modifications to the Website.

The manner by which these references are updated may be configured according to a preferences profile. For example, preferences may be set that indicate that references to modified Web page content may be automatically corrected in the code of the Web pages. Other preferences may include notifying a Webmaster or other administrator of the modification, providing a report of the references in the Web pages of the Website that need to be updated based on the modification to the Website content, marking obsolete or invalid references so that they are not selectable by a user of a client device, removing obsolete or invalid references in Web pages, and the like.

By way of the index data structure and the Website reference monitor, references to invalid or obsolete Web page content may be identified and automatically corrected so as to avoid having a user access a obsolete reference or the wrong Web page content. In addition, these mechanisms  
5 may reduce the network traffic by marking the obsolete or invalid references, or removing the obsolete or invalid references, such that they are not rendered by a Web browser of a client device or otherwise rendered such that they are not selectable by a user. In this way, a user is not able to select the reference to initiate a request for the obsolete or  
10 invalid Web page content. As a result, the network traffic associated with requesting obsolete or invalid Web page content is reduced.

In addition to the index manager and Website reference monitor, the illustrative embodiments also provide an obsolete reference correction  
15 agent that operates on client device requests for Web pages so as to remove or inactivate obsolete references to Web page content. When a client device sends a request to the Website for a particular Web page, a request handler receives the request and passes the request to the obsolete reference correction agent. The obsolete reference correction  
20 agent retrieves the requested Web page and checks the references within the Web page to determine if the references are to live Web page content.

This determination may involve retrieving information from the local file system for those references identifying locally stored Web page content.  
25 For references identifying remotely stored Web page content, such as on another server, a request for the Web page content may be sent to the remote system. If the local file system identifies the Web page content associated with the reference to be not present in the file system, or if the request for the Web page content results in an error message being  
30 returned, the reference in the requested Web page may be modified so as to make the reference non-selectable by a user of the client device. Such modification may involve modifying the code of the Web page to make the reference non-selectable, to remove the reference from the code  
35 altogether, or the like. The modified Web page code may then be sent to the client device so that it may be rendered on the client device via the client device's Web browser.

As stated above, there is provided, a computer program product comprising a computer useable medium having a computer readable program is provided.  
40 The computer readable program, when executed on a computing device, causes the computing device to generate an indexed data structure identifying Web

pages of the Website and references to content that are present in the Web pages of the Website. The computer readable program further may cause the computing device to receive a modification to content of the Website, search the indexed data structure to identify one or more Web pages of the Website that contain references to the modified content of the Website, and perform at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content. The references to content may comprise one or more of hyperlinks, uniform resource locators (URLs), references to image files, references to graphics files, references to sound files, or references to video files.

The at least one operation facilitates updating of the references to the modified content in the identified one or more Web pages of the Website. For example, the at least one operation may comprise automatically updating code of the identified one or more Web pages to change a reference to the modified content. The at least one operation may also comprise reporting the identified one or more Web pages having references to the modified content to an administrator. Moreover, the at least one operation may comprise marking the references to the modified content in the identified one or more Web pages such that they are not rendered by Web browsers of client devices in a manner that is selectable by a user.

The computer readable program may cause the computing device to perform at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content by retrieving a preferences profile identifying the at least one operation that is to be performed in response to an identification of one or more Web pages containing references to modified content and performing the at least one operation based on the at least one operation identified in the preferences profile. The computer readable program may cause the computing device to generate an indexed data structure by searching each Web page of the Website for references to content contained in each Web page and generating an entry in the indexed data structure for each Web page of the Website, wherein the entry is indexed by an identifier of the Web page and contains a listing of each reference to content contained in the corresponding Web page.

The computer readable program may further cause the computing device to register the indexed data structure with a Website reference monitor and parse the indexed data structure to identify references to content

identified in the indexed data structure. Moreover, the computer readable program may also cause the computing device to generate a monitor list comprising a list of the references to content identified in the indexed data structure that are to be monitored. The modification to content of the Website may be received based on a modification to content of the Website matching an entry in the monitor list.

The computer readable program may further cause the computing device to register the monitor list with a file system of a server computing device hosting the Website. The file system may notify the Website reference monitor of modifications to content corresponding to the references to content listed in the monitor list.

The computer readable program may further cause the computing device to update the indexed data structure based on results of performing the at least one operation. The computer readable program may cause the computing device to receive a request for a Web page from a client device and search the indexed data structure for an entry corresponding to the requested Web page. The computer readable program may also cause the computing device to check references to content identified in the entry of the indexed data structure corresponding to the requested Web page to identify one or more references to obsolete or invalid content, modify the one or more references to obsolete or invalid content in code of the requested Web page to generate modified code for the requested Web page, and provide the modified code for the request Web page to the client device.

The computer readable program may cause the computing device to check references to content identified in the entry of the indexed data structure by retrieving information, from a file system of a server computing device hosting the Web page, for those references to content that identify locally stored Web page content. Moreover, requests may be sent to remotely located computing devices hosting content associated with those references to content that identify remotely stored Web page content.

The computer readable program may cause the computing device to identify a reference to content to be a reference to obsolete or invalid content if the file system identifies the Web page content associated with the reference to be not present in a local storage system of the server computing device and registered with the file system or if a request for

the Web page content corresponding to the reference sent to a remote computing device results in an error message being returned.

5 According to another aspect, there is provided, a system is provided for updating a Website. The system may comprise a processor and a memory  
coupled to the processor. The memory may contain instructions that, when  
executed by the processor, implement an index manager and a Website  
reference monitor. The index manager may generate an indexed data  
10 structure identifying Web pages of the Website and references to content that are present in the Web pages of the Website. The Website reference  
monitor may receive a modification to content of the Website, search the  
indexed data structure to identify one or more Web pages of the Website  
that contain references to the modified content of the Website, and  
perform at least one operation based on the identification of the one or  
15 more Web pages of the Website that contain references to the modified  
content. The at least one operation may facilitate updating of the  
references to the modified content in the identified one or more Web pages  
of the Website.

20 For example, the at least one operation may comprise automatically updating code of the identified one or more Web pages to change a  
reference to the modified content. The at least one operation may also  
comprise reporting the identified one or more Web pages having references  
to the modified content to an administrator. Moreover, the at least one  
25 operation may comprise marking the references to the modified content in  
the identified one or more Web pages such that they are not rendered by  
Web browsers of client devices in a manner that is selectable by a user.

The Website reference monitor may perform at least one operation based on  
30 the identification of the one or more Web pages of the Website that  
contain references to the modified content by retrieving a preferences  
profile identifying the at least one operation that is to be performed in  
response to an identification of one or more Web pages containing  
references to modified content. The Website reference monitor may perform  
35 the at least one operation based on the at least one operation identified  
in the preferences profile.

The index manager may generate an indexed data structure by searching each  
40 Web page of the Website for references to content contained in each Web  
page and generating an entry in the indexed data structure for each Web  
page of the Website. The entry may be indexed by an identifier of the Web

page and may contain a listing of each reference to content contained in the corresponding Web page. The references to content may comprise one or more of hyperlinks, uniform resource locators (URLs), references to image files, references to graphics files, references to sound files, or references to video files.

The index manager may register the indexed data structure with a Website reference monitor. The Website reference monitor may parse the indexed data structure to identify references to content identified in the indexed data structure and generate a monitor list comprising a list of the references to content identified in the indexed data structure that are to be monitored. The modification to content of the Website may be received based on a modification to content of the Website matching an entry in the monitor list.

The Website reference monitor may register the monitor list with a file system of a server computing device hosting the Website. The file system may notify the Website reference monitor of modifications to content corresponding to the references to content listed in the monitor list. The index manager may update the indexed data structure based on results of performing the at least one operation.

The instructions in the memory may further implement a obsolete/invalid reference identification and correction engine. The obsolete/invalid reference identification and correction engine may receive a request for a Web page from a client device and search the indexed data structure for an entry corresponding to the requested Web page. The obsolete/invalid reference identification and correction engine may further check references to content identified in the entry of the indexed data structure corresponding to the requested Web page to identify one or more references to obsolete or invalid content, modify the one or more references to obsolete or invalid content in code of the requested Web page to generate modified code for the requested Web page, and provide the modified code for the request Web page to the client device.

The obsolete/invalid reference identification and correction engine may check references to content identified in the entry of the indexed data structure by retrieving information, from a file system of a server computing device hosting the Web page, for those references to content that identify locally stored Web page content and send requests to remotely located computing devices hosting content associated with those

references to content that identify remotely stored Web page content. The obsolete/invalid reference identification and correction engine may identify a reference to content to be a reference to obsolete or invalid content if the file system identifies the Web page content associated with the reference to be not present in a local storage system of the server computing device and registered with the file system or if a request for the Web page content corresponding to the reference sent to a remote computing device results in an error message being returned.

According to another aspect, there is provided, a method, in a data processing system, for updating a Website is provided. The method may comprise generating an indexed data structure identifying Web pages of the Website and references to content that are present in the Web pages of the Website. The method may further comprise receiving a modification to content of the Website, searching the indexed data structure to identify one or more Web pages of the Website that contain references to the modified content of the Website, and performing at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content. The at least one operation may facilitate updating of the references to the modified content in the identified one or more Web pages of the Website.

The at least one operation may comprise at least one of automatically updating code of the identified one or more Web pages to change a reference to the modified content, reporting the identified one or more Web pages having references to the modified content to an administrator, or marking the references to the modified content in the identified one or more Web pages such that they are not rendered by Web browsers of client devices in a manner that is selectable by a user.

The performing of at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content may comprise retrieving a preferences profile identifying the at least one operation that is to be performed in response to an identification of one or more Web pages containing references to modified content and performing the at least one operation based on the at least one operation identified in the preferences profile. The generating of an indexed data structure may comprise searching each Web page of the Website for references to content contained in each Web page and generating an entry in the indexed data structure for each Web page of the Website. The entry may be indexed by an identifier of the Web page and contains a

listing of each reference to content contained in the corresponding Web page.

5 The method may further comprise registering the indexed data structure with a Website reference monitor and parsing the indexed data structure to identify references to content identified in the indexed data structure. The method may also comprise generating a monitor list comprising a list of the references to content identified in the indexed data structure that are to be monitored. The modification to content of the Website may be  
10 received based on a modification to content of the Website matching an entry in the monitor list.

The method may comprise registering the monitor list with a file system of a server computing device hosting the Website. The file system may notify  
15 the Website reference monitor of modifications to content corresponding to the references to content listed in the monitor list. The method may further comprise updating the indexed data structure based on results of performing the at least one operation. Further, the method may comprise receiving a request for a Web page from a client device, searching the  
20 indexed data structure for an entry corresponding to the requested Web page, and checking references to content identified in the entry of the indexed data structure corresponding to the requested Web page to identify one or more references to obsolete or invalid content. The method may also comprise modifying the one or more references to obsolete or invalid  
25 content in code of the requested Web page to generate modified code for the requested Web page and providing the modified code for the request Web page to the client device.

The checking of references to content identified in the entry of the  
30 indexed data structure may comprise retrieving information, from a file system of a server computing device hosting the Web page, for those references to content that identify locally stored Web page content. The checking of references may further comprise sending requests to remotely located computing devices hosting content associated with those references  
35 to content that identify remotely stored Web page content.

These and other features and advantages of the present invention will be described in, or will become apparent to those of ordinary skill in the art in view of, the following detailed description of the exemplary  
40 embodiments of the present invention.

## BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the present invention will now be described, by way of example only, and with reference to the following drawings:

5

**Figure 1** is an exemplary block diagram of a distributed network data processing system in which exemplary aspects of the illustrative embodiments may be implemented;

10

**Figure 2** is an exemplary block diagram of a server data processing system in which exemplary aspects of the illustrative embodiments may be implemented;

15

**Figure 3** is an exemplary block diagram of a client data processing system in which exemplary aspects of the illustrative embodiments may be implemented;

20

**Figure 4** is an exemplary diagram illustrating a data flow between the primary operational elements of one illustrative embodiment;

**Figure 5** is an exemplary diagram illustrating an index structure in accordance with one illustrative embodiment;

25

**Figure 6** is a flowchart outlining an exemplary operation for scanning websites for obsolete Web page references and for auto-correcting Web page references in accordance with one illustrative embodiment; and

**Figure 7** is a flowchart outlining an exemplary operation for handling a client request in accordance with one illustrative embodiment.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

5 The illustrative embodiments provide a mechanism for identifying and automatically correcting obsolete and invalid references in Web pages. As such, the mechanisms of the illustrative embodiments are especially well suited for implementation in a distributed network data processing system in which a plurality of computing devices communicate with one another via one or more networks. **Figures 1-3** hereafter are provided as examples of data processing environments and devices in which the exemplary aspects of the illustrative embodiments may be implemented. **Figures 1-3** are only exemplary and are not intended to state or imply any limitation with regard to the types of environments or data processing systems in which the present invention may be implemented. Many modifications to the architectures illustrated in **Figures 1-3** may be made without departing from the spirit and scope of the present invention.

20 With reference now to the figures, **Figure 1** depicts a pictorial representation of a network of data processing systems in which a preferred embodiment of the present invention may be implemented. Network data processing system **100** is a network of computers in which the present invention may be implemented in accordance with a preferred embodiment. Network data processing system **100** contains a network **102**, which is the medium used to provide communications links between various devices and computers connected together within network data processing system **100**. Network **102** may include connections, such as wire, wireless communication links, or fiber optic cables.

30 In the depicted example, server **104** is connected to network **102** along with storage unit **106**. In addition, clients **108**, **110**, and **112** are connected to network **102**. These clients **108**, **110**, and **112** may be, for example, personal computers or network computers. In the depicted example, server **104** provides data, such as boot files, operating system images, and applications to clients **108-112**. Clients **108**, **110**, and **112** are clients to server **104**. Network data processing system **100** may include additional servers, clients, and other devices not shown. In the depicted example, network data processing system **100** is the Internet with network **102** representing a worldwide collection of networks and gateways that use the Transmission Control Protocol/Internet Protocol (TCP/IP) suite of protocols to communicate with one another. At the heart of the Internet is

a backbone of high-speed data communication lines between major nodes or host computers, consisting of thousands of commercial, government, educational and other computer systems that route data and messages. Of course, network data processing system **100** also may be implemented as a number of different types of networks, such as for example, an intranet, a local area network (LAN), or a wide area network (WAN). **Figure 1** is intended as an example, and not as an architectural limitation for the present invention.

Referring to **Figure 2**, a block diagram of a data processing system that may be implemented as a server, such as server **104** in **Figure 1**, is depicted in accordance with a preferred embodiment of the present invention. Data processing system **200** may be a symmetric multiprocessor (SMP) system including a plurality of processors **202** and **204** connected to system bus **206**. Alternatively, a single processor system may be employed. Also connected to system bus **206** is memory controller/cache **208**, which provides an interface to local memory **209**. I/O Bus Bridge **210** is connected to system bus **206** and provides an interface to I/O bus **212**. Memory controller/cache **208** and I/O Bus Bridge **210** may be integrated as depicted.

Peripheral component interconnect (PCI) bus bridge **214** connected to I/O bus **212** provides an interface to PCI local bus **216**. A number of modems may be connected to PCI local bus **216**. Typical PCI bus implementations will support four PCI expansion slots or add-in connectors. Communications links to clients **108-112** in **Figure 1** may be provided through modem **218** and network adapter **220** connected to PCI local bus **216** through add-in connectors.

Additional PCI bus bridges **222** and **224** provide interfaces for additional PCI local buses **226** and **228**, from which additional modems or network adapters may be supported. In this manner, data processing system **200** allows connections to multiple network computers. A memory-mapped graphics adapter **230** and hard disk **232** may also be connected to I/O bus **212** as depicted, either directly or indirectly.

Those of ordinary skill in the art will appreciate that the hardware depicted in **Figure 2** may vary. For example, other peripheral devices, such as optical disk drives and the like, also may be used in addition to or in place of the hardware depicted. The depicted example is not meant to imply architectural limitations with respect to the present invention.

The data processing system depicted in **Figure 2** may be, for example, an IBM® eServer™ pSeries® system, a product of International Business Machines Corporation in Armonk, New York, running the Advanced Interactive Executive (AIX®) operating system or LINUX® operating system. (IBM, eServer, pSeries and AIX are trademarks of International Business Machines Corporation in the United States, other countries, or both; and Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.)

With reference now to **Figure 3**, a block diagram illustrating a data processing system is depicted in which a preferred embodiment of the present invention may be implemented. Data processing system **300** is an example of a client computer. Data processing system **300** employs a peripheral component interconnect (PCI) local bus architecture. Although the depicted example employs a PCI bus, other bus architectures such as Accelerated Graphics Port (AGP) and Industry Standard Architecture (ISA) may be used. Processor **302** and main memory **304** are connected to PCI local bus **306** through PCI Bridge **308**. PCI Bridge **308** also may include an integrated memory controller and cache memory for processor **302**. Additional connections to PCI local bus **306** may be made through direct component interconnection or through add-in boards.

In the depicted example, local area network (LAN) adapter **310**, small computer system interface (SCSI) host bus adapter **312**, and expansion bus interface **314** are connected to PCI local bus **306** by direct component connection. In contrast, audio adapter **316**, graphics adapter **318**, and audio/video adapter **319** are connected to PCI local bus **306** by add-in boards inserted into expansion slots. Expansion bus interface **314** provides a connection for a keyboard and mouse adapter **320**, modem **322**, and additional memory **324**. SCSI host bus adapter **312** provides a connection for hard disk drive **326**, tape drive **328**, and CD-ROM drive **330**. Typical PCI local bus implementations will support three or four PCI expansion slots or add-in connectors.

An operating system runs on processor **302** and is used to coordinate and provide control of various components within data processing system **300** in **Figure 3**. The operating system may be a commercially available operating system, such as Windows® XP, which is available from Microsoft® Corporation. An object oriented programming system such as Java may run in

conjunction with the operating system and provide calls to the operating system from Java™ programs or applications executing on data processing system **300**. "Java" is a trademark of Sun Microsystems, Inc. Instructions for the operating system, the object-oriented programming system, and applications or programs are located on storage devices, such as hard disk drive **326**, and may be loaded into main memory **304** for execution by processor **302**. (Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.)

Those of ordinary skill in the art will appreciate that the hardware in **Figure 3** may vary depending on the implementation. Other internal hardware or peripheral devices, such as flash read-only memory (ROM), equivalent nonvolatile memory, or optical disk drives and the like, may be used in addition to or in place of the hardware depicted in **Figure 3**. Also, the processes of the present invention may be applied to a multiprocessor data processing system.

As another example, data processing system **300** may be a stand-alone system configured to be bootable without relying on some type of network communication interfaces. As a further example, data processing system **300** may be a personal digital assistant (PDA) device, which is configured with ROM and/or flash ROM in order to provide non-volatile memory for storing operating system files and/or user-generated data.

The depicted example in **Figure 3** and above-described examples are not meant to imply architectural limitations. For example, data processing system **300** also may be a notebook computer or hand held computer in addition to taking the form of a PDA. Data processing system **300** also may be a kiosk or a Web appliance.

Referring again to **Figure 1**, with the illustrative embodiments, server **104** provides one or more Websites that may be accessed by client devices **108-112**. In addition, server **104** includes an obsolete/invalid reference identification and correction engine that operates to monitor Websites to identify obsolete and/or invalid references to Web page content and automatically correct such references prior to Web pages being sent to client devices for rendering by client device Web browsers. In this way, frustration on the part of users of client devices when accessing obsolete and invalid references is reduced. Moreover, network traffic for retrieving obsolete or invalid Web page content is reduced.

**Figure 4** is an exemplary diagram illustrating a data flow between the primary operational elements of a obsolete/invalid reference identification and correction engine in accordance with one illustrative embodiment. In the illustrative embodiment, the operational elements shown in **Figure 4** are provided as part of a server computing device that hosts one or more Websites. For example, the server computing device may be server **104** in **Figure 1** that provides Website Web page content to client devices **108-112**.

As shown in **Figure 4**, a obsolete/invalid reference identification and correction engine **400** includes a obsolete reference correction agent **420**, an index manager **440**, and a website reference monitor **460**. The elements **420**, **440** and **460** interfaces with a file system **480** of the server computing device to obtain access to Web pages **432** of Website **430** stored in local storage system **450**. The index manager **440** further interfaces with an index data structure **452** stored in the local storage system **450**. Obsolete reference correction agent **420** further interfaces with HTTP request handler **410** to handle requests for Web pages from client computing devices.

The obsolete/invalid reference identification and correction engine **400** (hereafter referred to as the "reference engine") has two main modes of operation. In a first mode of operation, the reference engine **400** monitors modifications to a Website, such as through Website editor **470**, in order to identify obsolete/invalid references to Web page content and automatically correct such references. In a second mode of operation, the reference engine **400** operates on requests from client devices for Web pages so as to identify obsolete references in the requested Web pages and rendering these obsolete references non-selectable prior to providing the Web pages to the client devices. Each of these modes of operation will now be described with reference to **Figure 4**.

In both modes of operation, the reference engine **400** uses an indexed data structure **452** corresponding to the Website **430** for identifying references present in the Web pages **432** that make up the Website **430**. This indexed data structure **452** is generated and maintained up-to-date by the index manager **440**.

The index manager **440** indexes each Web page of a Website and identifies all references to Website content present in the Web pages **432** of the Website **430**. In particular, an index manager **440** scans (i.e., crawls) the code of the Web pages **432** of the entire Website **430** and identifies  
5 references to Web page content, e.g., hyperlinks, references to image files, graphics files, sound files, video files, etc. For example, the index manager **440** looks at the markup language code, e.g., HyperText Markup Language (HTML), for the Web pages **432** and, based on HTML tags, recognizable HTML code terms, or the like, identifies hyperlinks, file  
10 references, and the like, in the markup language code of the Web pages **432**. In one illustrative embodiment, references are provided as Uniform Resource Locators (URLs) and the index manager **440** searches the code of the Web pages **432** for URLs.

Based on the results of the search of a Web page in the Web pages **432** of the Website **430**, an entry for the Web page is added to the indexed data structure **452**. The entry in the indexed data structure **452** is indexed by the Web page reference, e.g., the URL of the Web page, and identifies the  
15 references present in the corresponding Web page. Other indexing mechanisms may be used as well, including indexed hash tables, such as for secure Web sites. This searching, or crawling, of a Web page is repeated  
20 for each Web page in the plurality of Web pages **432** that together comprise the Website **430** such that an indexed data structure **452** for the entire Website **430** is generated. As a result, the indexed data structure **452**  
25 will have a separate entry for each Web page in the Website **430** and each entry will identify what Web content references are present in the code of the corresponding Web page.

The searching or crawling of the Website **430** may be performed once, such  
30 as upon deployment of the Website **430**, to establish an initial indexed data structure **452** that is subsequently maintained up-to-date by real time updates when the Website **430** is modified, as discussed in greater detail hereafter. Alternatively, or in addition, the searching or crawling of the Website **430** may be performed periodically so as to ensure that the  
35 indexed data structure **452** is correct and was not inadvertently corrupted or otherwise not kept up-to-date.

The indexed data structure **452** is used to identify obsolete and invalid references to Web content in Web pages of a Website as the Website is

modified. Once the index manager **440** generates the indexed data structure **452**, the index manager **440** registers the indexed Web pages and their corresponding references with the Website reference monitor **460**.

Essentially, the indexed data structure **452** is provided to the Website

5 reference monitor **460** which parses the indexed data structure **452** and identifies which files are to be monitored by the Website reference monitor **460**. The identification of these files is then added to a monitor list maintained by the Website reference monitor **460**. The monitor list is registered with the file system **480** which provides notifications of

10 modifications to the Website reference monitor **460** when any of the files referenced in the monitor list are modified, i.e. deleted, renamed, relocated, new file references added to these files, or the like.

Notifications of modifications to files are provided by the file system **480** to the Website reference monitor **460**. The file system **480** informs the

15 Website reference monitor **460**, through standard file system notification mechanisms, of the particular file that is modified and the nature of the modification, e.g., deletion, renaming, relocation, addition, etc. Based on the notification, the Website reference monitor **460** may search the indexed data structure **452** for the references to the file that was

20 modified. In this way, the Website reference monitor **460** may identify which Web pages **432** of the Website **430** need to be modified based on the modifications to the file.

For example, a user of a Website editor **470** may access a Web page in the

25 set of Web pages **432** and modify it. In the process, the Web page **432** may be stored in a different location of the local storage system **450**, i.e. at a different hyperlink location. Thus, the old hyperlinks to the Web page in other Web pages **432** of the Website **430** will either be obsolete (not have an associated Web page file at the location specified by the

30 hyperlink) or may reference the old, invalid, version of the Web page. Accordingly, these hyperlinks in the other Web pages **432** must be updated to reference the new, modified, version of the Web page at the new location.

35 The modification performed by the user of the Website editor **470** is reported by the file system **480** to the Website reference monitor **460** and indicates both the file modified and the nature of the modification, e.g., the new location of the modified file in the above example. The Website reference monitor **460** searches all entries of the indexed data structure

**452**, via the index manager **440**, to identify all references to the file that was modified. The references to the modified file may be quickly and easily identified by virtue of the indexed data structure since each entry in the indexed data structure identifies the references included in the Web page associated with the entry. Thus, by searching each entry, all of the references to files, Web pages, and the like, may be identified for the entire Website **430**.

Based on the results of the search, one or more of a plurality of operations may be performed. These operations may include automatically updating the references in the other Web pages **432**, notifying a Webmaster or other administrator of the Web pages that need to be updated along with the identifier of the file that was modified and the nature of the modification, marking the references in the other Web pages as being invalid or obsolete depending upon the nature of the modification such that they are not rendered by Web browsers in a manner that is selectable by a user, and the like. Such marking of references may be performed, for example, by inserting appropriate tags into the code of the Web pages that, when interpreted by a Web browser, cause the Web browser to render the reference in a non-selectable manner, such as by graying out the reference, removing the hyperlink aspect of the reference and leaving it as text only, or the like.

The manner by which these references are updated may be configured according to a preferences profile stored in the Website reference monitor **460** which is modifiable by a Website operator, owner, or the like. For example, preferences may be set that indicate that references to modified Web page content, e.g., files, directories, or the like, may be automatically corrected in the code of the Web pages. Other preferences may include notifying a Webmaster or other administrator of the modification, providing a report of the references in the Web pages of the Website that need to be updated based on the modification to the Website content, marking obsolete or invalid references so that they are not selectable by a user of a client device, removing obsolete or invalid references in Web pages, and the like.

If the other Web pages **432** are to be modified such that the references to the modified files are updated, then the Website reference monitor **460** edits the code of the Web pages **432** to change references to the old, obsolete, or invalid version of the file. The references are updated based on the nature of the modification performed to the file. For

example, if the file is modified and relocated, then the references are updated to reference the new location of the modified file. If the file is modified and renamed, then the references to the file are updated to refer to the new renamed file. If the file is deleted, then the  
5 references to the file in the Web pages **432** is removed or marked as obsolete or invalid.

Based on the updates to the actual code of the Web pages **432** that include references to the file that was modified, the Website reference monitor  
10 **460** informs the index manager **440** of the Web pages **432** that were updated and the manner by which they were updated, e.g., the changes to the file names, the changes to the storage locations, the removal of a reference to a file, the addition of a reference to a file, and the like. Based on the update information sent from the Website reference monitor **460** to the  
15 index manager **440**, the index manager **440** updates the entries in the index data structure **452** for the Web pages **432** that were updated. In this way, the indexed data structure **452** is automatically kept up-to-date as modifications to the Website **430** are made by a user of the Website editor **470**. Furthermore, references to the modified files of a Website **430** are  
20 automatically updated throughout the Website **430** so as to eliminate obsolete or invalid references.

It should be noted that, in addition to detecting modifications to existing files, directories, Web pages, and the like, the file system **480**  
25 may further notify the Website reference monitor of additions to the Website **430**. For example, if a new Web page is generated, new files or directories are generated, and added to the Website, such additions will be notified to the Website reference monitor **460**. Typically, to integrate such new files, directories, or Web pages into the Website **430**, existing  
30 Web pages **432** of the Website **430** will need to be modified to include a reference to these new files, directories, or Web pages and thus, the new elements may be integrated into the indexed data structure at this time. Alternatively, the file system **480** may inform the Website reference monitor **460** of the generation of these new elements when they are created,  
35 even though they are not part of the registered list of Web pages and references yet, such that they may be integrated into the indexed data structure and registered with the Website reference monitor **460** and file system **480**.

In addition to the index manager **440** and Website reference monitor **460**, the obsolete/invalid reference identification and correction engine **400** of the illustrative embodiments also provides a obsolete reference correction agent **420** that, in the second mode of operation, operates on client device requests for Web pages so as to remove or inactivate obsolete references to Web page content. When a client device sends a request to the Website **430** for a particular Web page **432**, the request handler **410** receives the request and passes the request to the obsolete reference correction agent **420**. The obsolete reference correction agent **420** retrieves the requested Web page **432** via the file system **480** and information for the requested Web page **432** from a corresponding entry in the indexed data structure **452**. Based on the information retrieved from the indexed data structure **452**, the obsolete reference correction agent **420** checks the references within the Web page **432** to determine if the references are to live Web page content, i.e. existing and valid files in the local storage system **450**.

This determination may involve retrieving information from the local file system **480** for those references identifying locally stored Web page content, e.g., files in the local storage system **450**. For references identifying remotely stored Web page content, such as files on another server, a request for the Web page content may be sent to the remote system. If the local file system **480** identifies the Web page content associated with the reference to be not present in the local storage system **450** and registered with file system **480**, or if the request for the Web page content sent to the remote system results in an error message being returned, the reference in the requested Web page may be modified so as to make the reference non-selectable by a user of the client device. For example, the obsolete reference correction agent **420** may modify the code of the Web page by inserting an appropriate tag in the code of the Web page that causes a Web browser of the client device to render the reference in a non-selectable manner, e.g., rendering the reference in a "grayed-out" manner and removing the selectable hyperlink such that the reference is provided as text only. Alternatively, the reference may be removed from the code altogether. The modified Web page code may then be sent, by the obsolete reference correction agent **420**, to the client device via the request handler **410** so that it may be rendered on the client device via the client device's Web browser.

**Figure 5** is an exemplary diagram illustrating an index structure in accordance with one illustrative embodiment. As shown in **Figure 5**, the

index structure **500** includes entries, such as entry **510**, for each Web page of a Website. The entries have an index key **520** and a listing **530** of the references included in the corresponding Web page. The listing of references **530** may be used to identify which Web pages have references to Web page content, e.g., files, that are modified by a user using a Website editor. The index key **520** corresponding to the entries that are identified as having references to Web page content that is modified may be used to identify the Web pages that need to be modified to reflect the modifications to the Web page content, as previously discussed above. The index key **520** may further be used to identify entries in the index data structure **500** that need to be updated based on changes to references in a corresponding Web page.

Thus, by way of the index data structure **452** and the Website reference monitor **460**, references to invalid or obsolete Web page content may be identified and automatically corrected so as to avoid having a user access a obsolete reference or the wrong Web page content. In addition, these mechanisms may reduce the network traffic by marking the obsolete or invalid references, or removing the obsolete or invalid references, such that they are not rendered by a Web browser of a client device or otherwise rendered such that they are not selectable by a user. In this way, a user is not able to select the reference to initiate a request for the obsolete or invalid Web page content. As a result, the network traffic associated with requesting obsolete or invalid Web page content is reduced.

**Figures 6** and **7** outline exemplary operations in accordance with illustrative embodiments of the present invention. It will be understood that each block of the flowchart illustrations, and combinations of blocks in the flowchart illustrations, can be implemented by computer program instructions. These computer program instructions may be provided to a processor or other programmable data processing apparatus to produce a machine, such that the instructions which execute on the processor or other programmable data processing apparatus create means for implementing the functions specified in the flowchart block or blocks. These computer program instructions may also be stored in a computer-readable memory or storage medium that can direct a processor or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory or storage medium produce an article of manufacture including instruction means which implement the functions specified in the flowchart block or blocks.

Accordingly, blocks of the flowchart illustrations support combinations of means for performing the specified functions, combinations of steps for performing the specified functions and program instruction means for performing the specified functions. It will also be understood that each block of the flowchart illustrations, and combinations of blocks in the flowchart illustrations, can be implemented by special purpose hardware-based computer systems which perform the specified functions or steps, or by combinations of special purpose hardware and computer instructions.

**Figure 6** is a flowchart outlining an exemplary operation for scanning websites for obsolete Web page references and for auto-correcting Web page references in accordance with one illustrative embodiment. As shown in **Figure 6**, the operation starts by scanning Web pages of a Website to identify references present in the Web pages (step **610**). Entries for each Web page of the Website are created in an indexed data structure identifying the Web page and the references present in the Web page (step **620**). The operation then registers the indexed Web pages and references with a Website reference monitor (step **630**). The Website reference monitor registers the indexed Web pages and references with the file system such that modifications to the Web pages, directories, and reference files will be notified to the Website reference monitor (step **640**).

The operation then waits for a modification to a file, directory, or Web page of the Website (step **650**). A determination is made as to whether a modification is detected (step **660**). If not, the operation returns to step **650** and continues to wait. If a modification is detected, a notification of the subject of the modification and the nature of the modification is provided to the Website reference monitor (step **670**). The Website reference monitor then searches the indexed data structure for references to the subject of the modification (step **680**).

For each reference to the subject of the modification found in the indexed data structure, the Website reference monitor performs an operation corresponding to a profile identifying the operations to perform when references to modified contents of the Website are identified (step **690**). Such operations may include updating code of the Web pages corresponding to the identified references based on the nature of the modification,

reporting the Web pages that need to be modified to an administrator, and the like. The index manager is then informed of the changes, if any, to the structure of the Website such that the indexed data structure is updated (step 695). The operation then terminates.

5

**Figure 7** is a flowchart outlining an exemplary operation for handling a client request in accordance with one illustrative embodiment. As shown in **Figure 7**, the operation starts by receiving the request for a Web page from a client device (step 710). The Web page is retrieved (step 720) and a corresponding indexed data structure entry is retrieved (step 730). The references identified in the indexed data structure entry are checked to determine if any of the references are to obsolete or invalid content, e.g., files (step 740).

10

15

A determination is made as to whether obsolete or invalid content is found (step 750). If not, the Web page is sent to the client device without modification (step 760). If obsolete or invalid content is found, the code of the Web page is modified to make such references to the obsolete or invalid content non-selectable when rendered by a Web browser on the client device (step 770). The modified Web page is then sent to the client device (step 780) and the operation terminates.

20

Thus, by operation of the mechanisms of the illustrative embodiments, obsolete or invalid references in Web pages of a Website may be automatically identified and modified prior to the Web pages being accessed by a user of a client device. In addition, the mechanisms of the illustrative embodiments provide an automated way to update references to modified content throughout a Website. This helps in reducing the frustration level of users of client devices when accessing obsolete or invalid links to Website content and helps Webmasters or administrators in identifying the portions of the Website that need to be modified when content of the Website that is referenced by these portions is modified. Furthermore, by reducing the occurrence of obsolete or invalid references in Websites, the illustrative embodiments reduce unnecessary network traffic.

25

30

35

It is important to note that while the present invention has been described in the context of a fully functioning data processing system, those of ordinary skill in the art will appreciate that the processes of the present invention are capable of being distributed in the form of a computer readable medium of instructions and a variety of forms and that the present invention applies equally regardless of the particular type of

40

signal bearing media actually used to carry out the distribution. Examples of computer readable media include recordable-type media, such as a floppy disk, a hard disk drive, a RAM, CD-ROMs, DVD-ROMs, and transmission-type media, such as digital and analog communications links, 5 wired or wireless communications links using transmission forms, such as, for example, radio frequency and light wave transmissions. The computer readable media may take the form of coded formats that are decoded for actual use in a particular data processing system.

10 The description of the present invention has been presented for purposes of illustration and description, and is not intended to be exhaustive or limited to the invention in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art. The embodiment was chosen and described in order to best explain the 15 principles of the invention, the practical application, and to enable others of ordinary skill in the art to understand the invention for various embodiments with various modifications as are suited to the particular use contemplated.

**CLAIMS**

1. A computer program product comprising a computer useable medium having a computer readable program, wherein the computer readable program, when executed on a computing device, causes the computing device to:  
5 generate an indexed data structure identifying Web pages of the Website and references to content that are present in the Web pages of the Website;

10 receive a modification to content of the Website;

search the indexed data structure to identify one or more Web pages of the Website that contain references to the modified content of the Website; and

15 perform at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content, wherein the at least one operation facilitates updating of the references to the modified content in the identified one or more  
20 Web pages of the Website.

2. The computer program product of claim 1, wherein the at least one operation comprises automatically updating code of the identified one or more Web pages to change a reference to the modified content.

3. The computer program product of claim 1, wherein the at least one operation comprises reporting the identified one or more Web pages having references to the modified content to an administrator.

4. The computer program product of claim 1, wherein the at least one operation comprises marking the references to the modified content in the identified one or more Web pages such that they are not rendered by Web browsers of client devices in a manner that is selectable by a user.

5. The computer program product of any preceding claim, wherein the computer readable program causes the computing device to perform at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content by:

retrieving a preferences profile identifying the at least one operation that is to be performed in response to an identification of one or more Web pages containing references to modified content; and

performing the at least one operation based on the at least one operation identified in the preferences profile.

6. The computer program product of any preceding claim, wherein the computer readable program causes the computing device to generate an indexed data structure by:

searching each Web page of the Website for references to content contained in each Web page; and

generating an entry in the indexed data structure for each Web page of the Website, wherein the entry is indexed by an identifier of the Web page and contains a listing of each reference to content contained in the corresponding Web page.

7. The computer program product of any preceding claim, wherein the references to content comprise one or more of hyperlinks, uniform resource locators (URLs), references to image files, references to graphics files, references to sound files, or references to video files.

8. The computer program product of any preceding claim, wherein the computer readable program further causes the computing device to:

register the indexed data structure with a Website reference monitor;

parse the indexed data structure to identify references to content identified in the indexed data structure; and

generate a monitor list comprising a list of the references to content identified in the indexed data structure that are to be monitored, wherein the modification to content of the Website is received based on a modification to content of the Website matching an entry in the monitor list.

9. The computer program product of claim 8, wherein the computer readable program further causes the computing device to:

register the monitor list with a file system of a server computing device hosting the Website, wherein the file system notifies the Website reference monitor of modifications to content corresponding to the references to content listed in the monitor list.

5

10. The computer program product of any preceding claim, wherein the computer readable program further causes the computing device to:

10

update the indexed data structure based on results of performing the at least one operation.

11. The computer program product of any preceding claim, wherein the computer readable program further causes the computing device to:

15

receive a request for a Web page from a client device;

search the indexed data structure for an entry corresponding to the requested Web page;

20

check references to content identified in the entry of the indexed data structure corresponding to the requested Web page to identify one or more references to obsolete or invalid content;

25

modify the one or more references to obsolete or invalid content in code of the requested Web page to generate modified code for the requested Web page; and

provide the modified code for the request Web page to the client device.

30

12. The computer program product of claim 11, wherein the computer readable program causes the computing device to check references to content identified in the entry of the indexed data structure by:

35

retrieving information, from a file system of a server computing device hosting the Web page, for those references to content that identify locally stored Web page content; and

40

sending requests to remotely located computing devices hosting content associated with those references to content that identify remotely stored Web page content.

13. The computer program product of claim 12, wherein the computer readable program causes the computing device to identify a reference to content to be a reference to obsolete or invalid content if the file system identifies the Web page content associated with the reference to be  
5 not present in a local storage system of the server computing device and registered with the file system or if a request for the Web page content corresponding to the reference sent to a remote computing device results in an error message being returned.

10 14. A system for updating a Website, comprising:

a processor; and

15 a memory coupled to the processor, wherein the memory contains instructions that, when executed by the processor, implement an index manager and a Website reference monitor, wherein the index manager generates an indexed data structure identifying Web pages of the Website and references to content that are present in the Web pages of the Website, and wherein the Website reference monitor:

20 receives a modification to content of the Website;

searches the indexed data structure to identify one or more Web pages of the Website that contain references to the modified content of the  
25 Website; and

performs at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content, wherein the at least one operation facilitates updating  
30 of the references to the modified content in the identified one or more Web pages of the Website.

15. The system of claim 14, wherein the at least one operation comprises automatically updating code of the identified one or more Web pages to  
35 change a reference to the modified content.

16. The system of claim 14, wherein the at least one operation comprises reporting the identified one or more Web pages having references to the modified content to an administrator.

17. The system of claim 14, wherein the at least one operation comprises marking the references to the modified content in the identified one or more Web pages such that they are not rendered by Web browsers of client devices in a manner that is selectable by a user.

5

18. The system of any of claims 14 to 17, wherein the Website reference monitor performs at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content by:

10

retrieving a preferences profile identifying the at least one operation that is to be performed in response to an identification of one or more Web pages containing references to modified content; and

15

performing the at least one operation based on the at least one operation identified in the preferences profile.

19. The system of any of claims 14 to 18, wherein the index manager generates an indexed data structure by:

20

searching each Web page of the Website for references to content contained in each Web page; and

25

generating an entry in the indexed data structure for each Web page of the Website, wherein the entry is indexed by an identifier of the Web page and contains a listing of each reference to content contained in the corresponding Web page.

30

20. The system of any of claims 14 to 19, wherein the references to content comprise one or more of hyperlinks, uniform resource locators (URLs), references to image files, references to graphics files, references to sound files, or references to video files.

35

21. The system of any of claims 14 to 20, wherein the index manager registers the indexed data structure with a Website reference monitor, and wherein the Website reference monitor:

40

parses the indexed data structure to identify references to content identified in the indexed data structure; and

generates a monitor list comprising a list of the references to content identified in the indexed data structure that are to be monitored, wherein the modification to content of the Website is received based on a modification to content of the Website matching an entry in the monitor list.

22. The system of claim 21, wherein the Website reference monitor registers the monitor list with a file system of a server computing device hosting the Website, and wherein the file system notifies the Website reference monitor of modifications to content corresponding to the references to content listed in the monitor list.

23. The system of any of claims 14 to 22, wherein the index manager updates the indexed data structure based on results of performing the at least one operation.

24. The system of any of claims 14 to 23, wherein the instructions further implement a obsolete/invalid reference identification and correction engine, and wherein the obsolete/invalid reference identification and correction engine:

receives a request for a Web page from a client device;

searches the indexed data structure for an entry corresponding to the requested Web page;

checks references to content identified in the entry of the indexed data structure corresponding to the requested Web page to identify one or more references to obsolete or invalid content;

modifies the one or more references to obsolete or invalid content in code of the requested Web page to generate modified code for the requested Web page; and

provides the modified code for the request Web page to the client device.

25. The system of claim 24, wherein the obsolete/invalid reference identification and correction engine checks references to content identified in the entry of the indexed data structure by:

retrieving information, from a file system of a server computing device hosting the Web page, for those references to content that identify locally stored Web page content; and

5 sending requests to remotely located computing devices hosting content associated with those references to content that identify remotely stored Web page content.

10 26. The system of claim 25, wherein the obsolete/invalid reference identification and correction engine identifies a reference to content to be a reference to obsolete or invalid content if the file system identifies the Web page content associated with the reference to be not present in a local storage system of the server computing device and registered with the file system or if a request for the Web page content  
15 corresponding to the reference sent to a remote computing device results in an error message being returned.

20 27. A method, in a data processing system, for updating a Website, comprising:

generating an indexed data structure identifying Web pages of the Website and references to content that are present in the Web pages of the Website;

25 receiving a modification to content of the Website;

searching the indexed data structure to identify one or more Web pages of the Website that contain references to the modified content of the Website; and

30 performing at least one operation based on the identification of the one or more Web pages of the Website that contain references to the modified content, wherein the at least one operation facilitates updating of the references to the modified content in the identified one or more  
35 Web pages of the Website.

40 28. The method of claim 27, wherein the at least one operation comprises at least one of automatically updating code of the identified one or more Web pages to change a reference to the modified content, reporting the identified one or more Web pages having references to the modified content to an administrator, or marking the references to the modified content in

the identified one or more Web pages such that they are not rendered by Web browsers of client devices in a manner that is selectable by a user.

29. The method of claim 27 or 28, wherein performing at least one  
5 operation based on the identification of the one or more Web pages of the Website that contain references to the modified content comprises:

10 retrieving a preferences profile identifying the at least one operation that is to be performed in response to an identification of one or more Web pages containing references to modified content; and

performing the at least one operation based on the at least one operation identified in the preferences profile.

30. The method of claim 27, 28 or 29, wherein generating an indexed data  
15 structure comprises:

20 searching each Web page of the Website for references to content contained in each Web page; and

25 generating an entry in the indexed data structure for each Web page of the Website, wherein the entry is indexed by an identifier of the Web page and contains a listing of each reference to content contained in the corresponding Web page.

31. The method of any of claims 27 to 30, further comprising:

30 registering the indexed data structure with a Website reference monitor;

parsing the indexed data structure to identify references to content identified in the indexed data structure; and

35 generating a monitor list comprising a list of the references to content identified in the indexed data structure that are to be monitored, wherein the modification to content of the Website is received based on a modification to content of the Website matching an entry in the monitor list.

40 32. The method of claim 31, further comprising:

registering the monitor list with a file system of a server computing device hosting the Website, wherein the file system notifies the Website reference monitor of modifications to content corresponding to the references to content listed in the monitor list.

5

33. The method of any of claims 27 to 32, further comprising updating the indexed data structure based on results of performing the at least one operation.

10 34. The method of any of claims 27 to 33, further comprising:

receiving a request for a Web page from a client device;

15 searching the indexed data structure for an entry corresponding to the requested Web page;

checking references to content identified in the entry of the indexed data structure corresponding to the requested Web page to identify one or more references to obsolete or invalid content;

20

modifying the one or more references to obsolete or invalid content in code of the requested Web page to generate modified code for the requested Web page; and

25 providing the modified code for the request Web page to the client device.

35. The method of claim 34, wherein checking references to content identified in the entry of the indexed data structure comprises:

30 retrieving information, from a file system of a server computing device hosting the Web page, for those references to content that identify locally stored Web page content; and

35 sending requests to remotely located computing devices hosting content associated with those references to content that identify remotely stored Web page content.

36. The method of claim 35 comprising the steps of:

40 identifying a reference to content to be a reference to obsolete or invalid content if the file system identifies the Web page content associated with the reference to be not present in a local storage system

of the server computing device and registered with the file system or if a request for the Web page content corresponding to the reference sent to a remote computing device results in an error message being returned.

- 5 37. A computer program comprising program code means adapted to perform the method of any of claims 27 to 36 when said program is run on a computer.

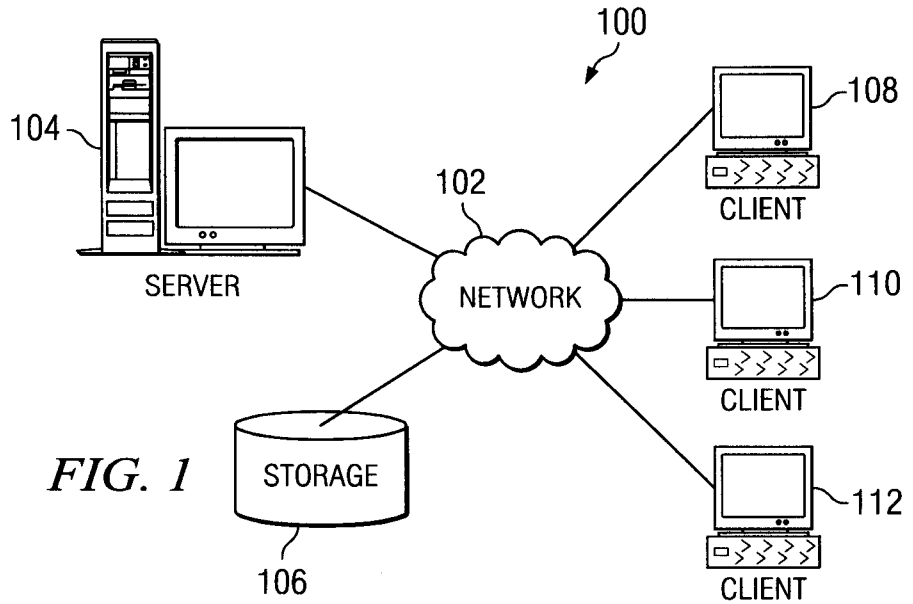


FIG. 1

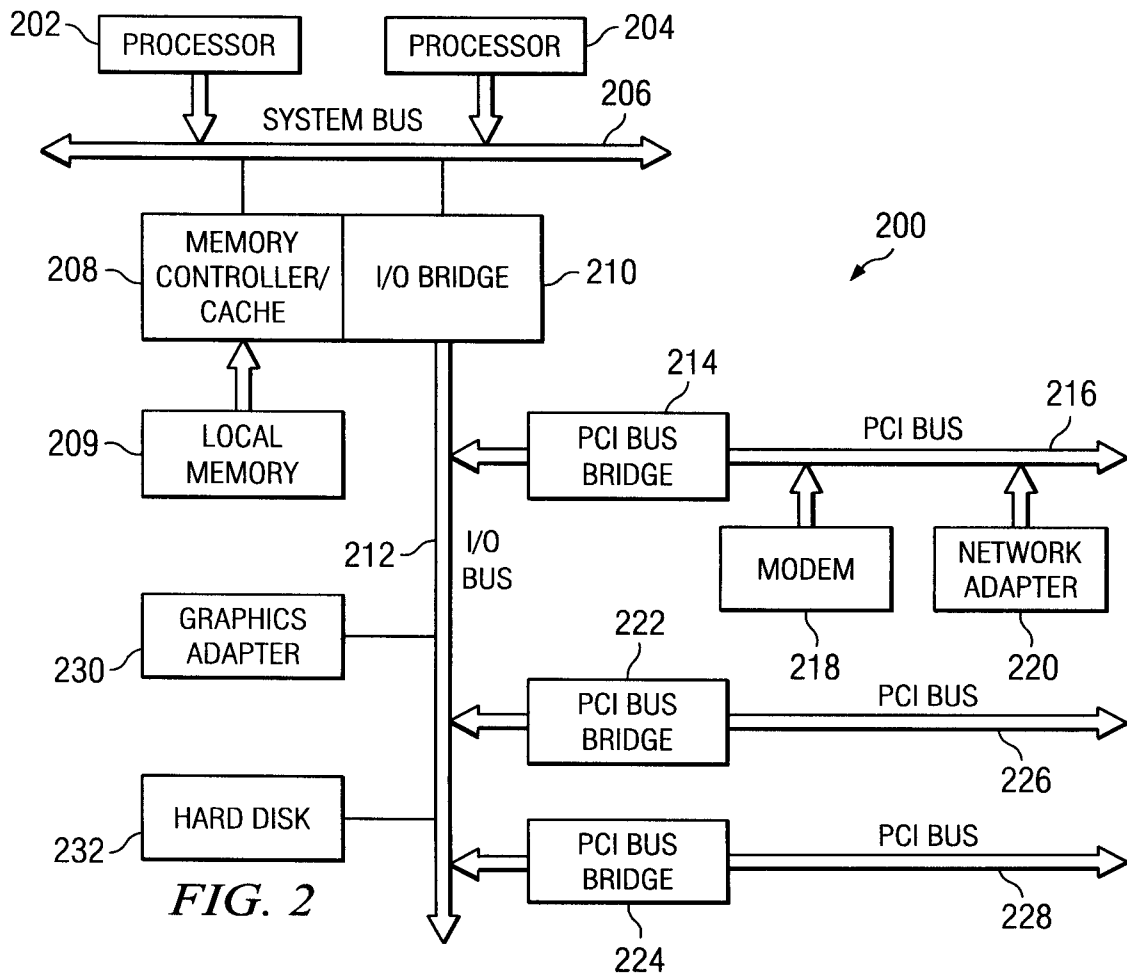


FIG. 2

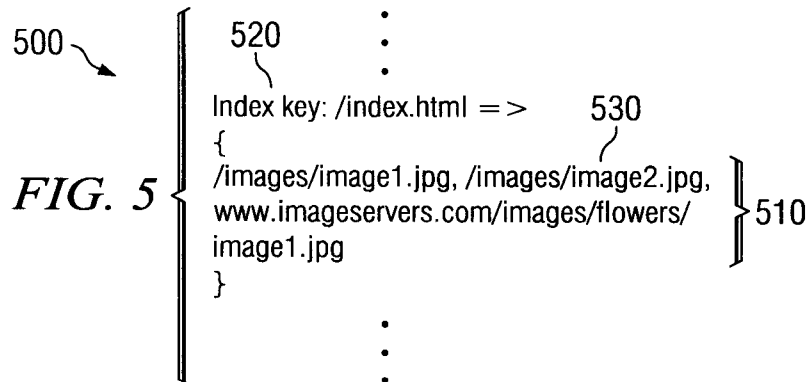
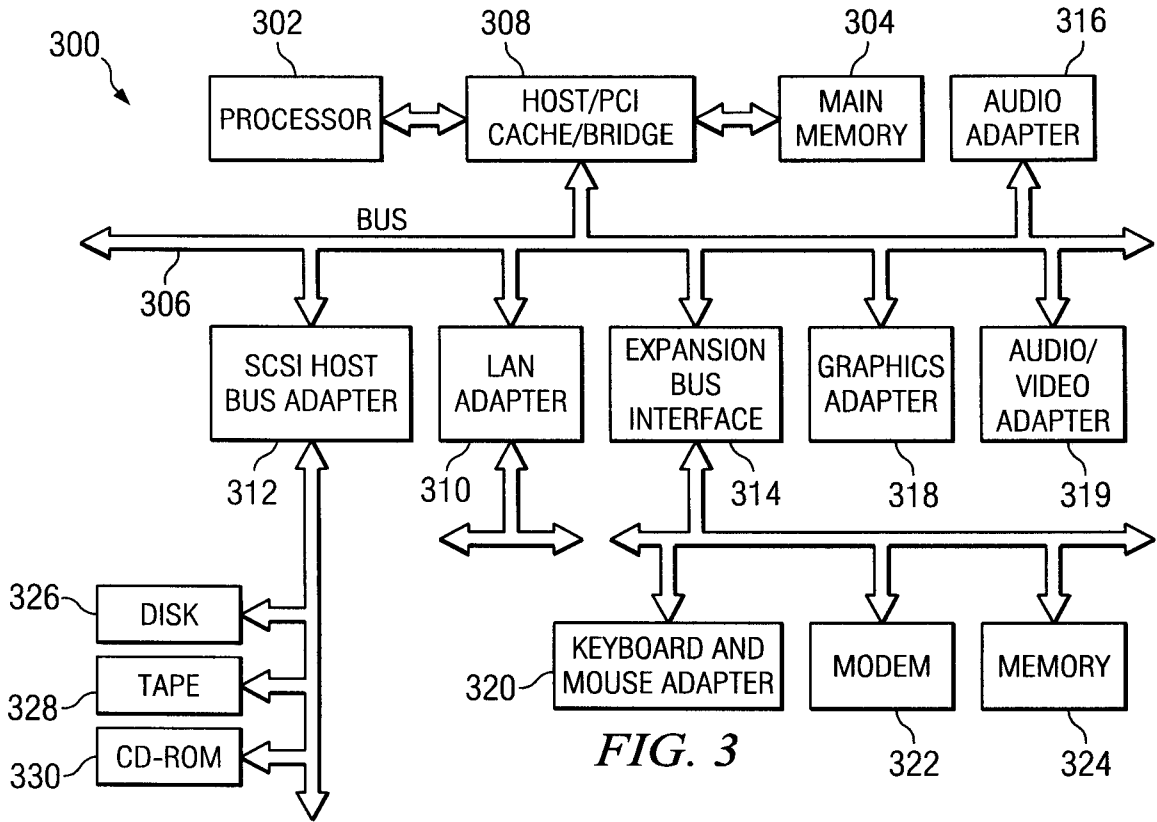


FIG. 4

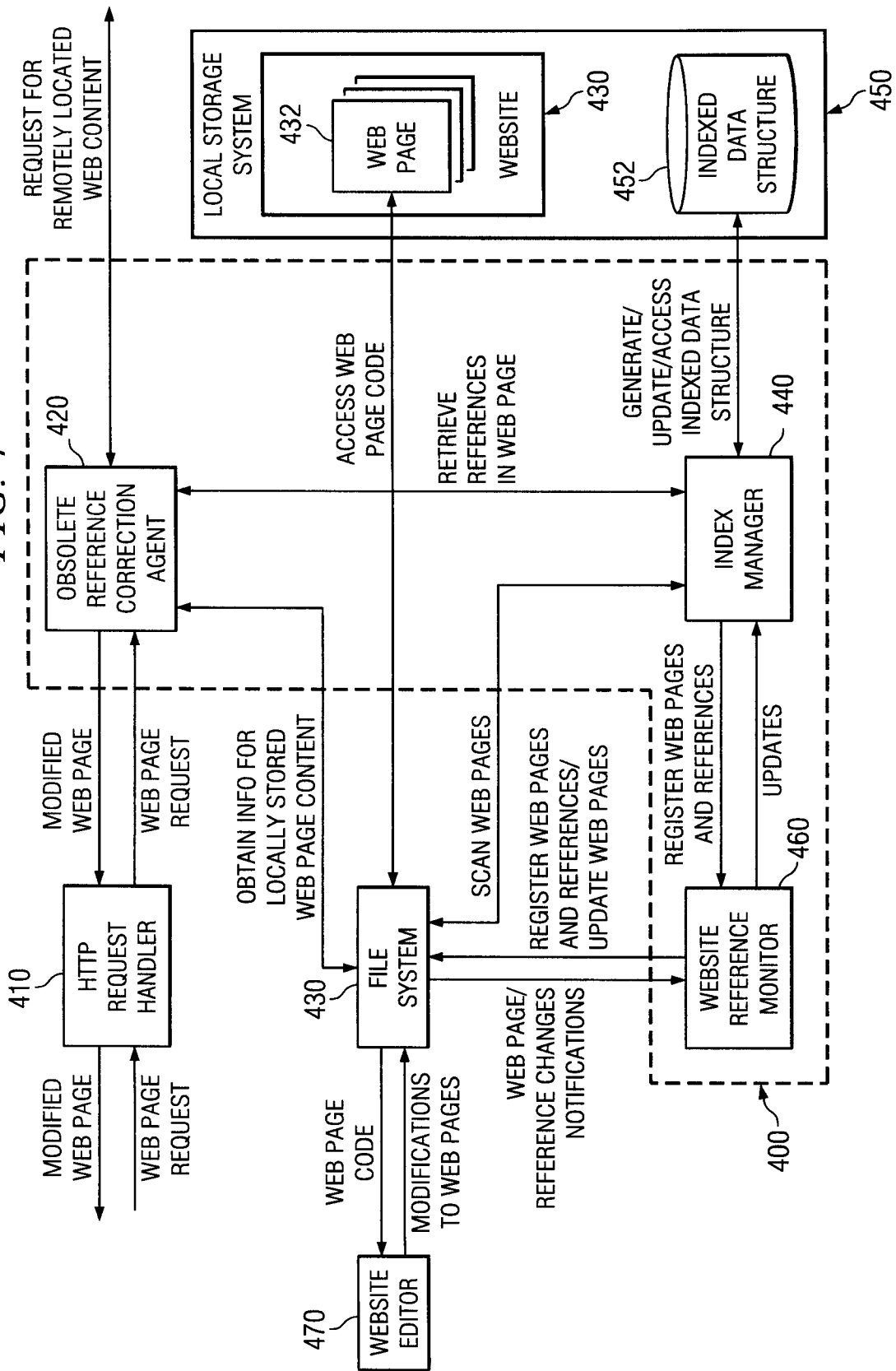


FIG. 6

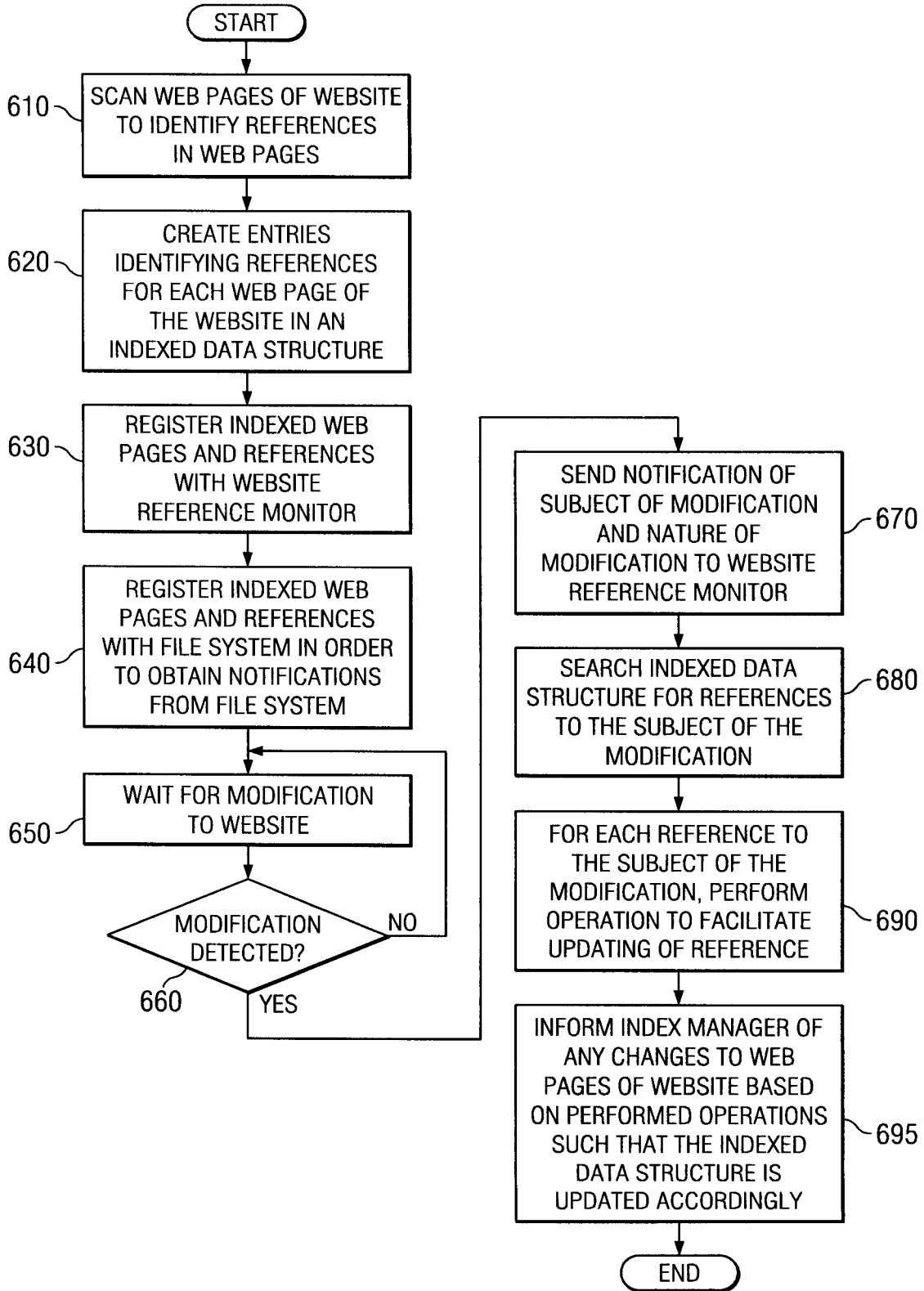
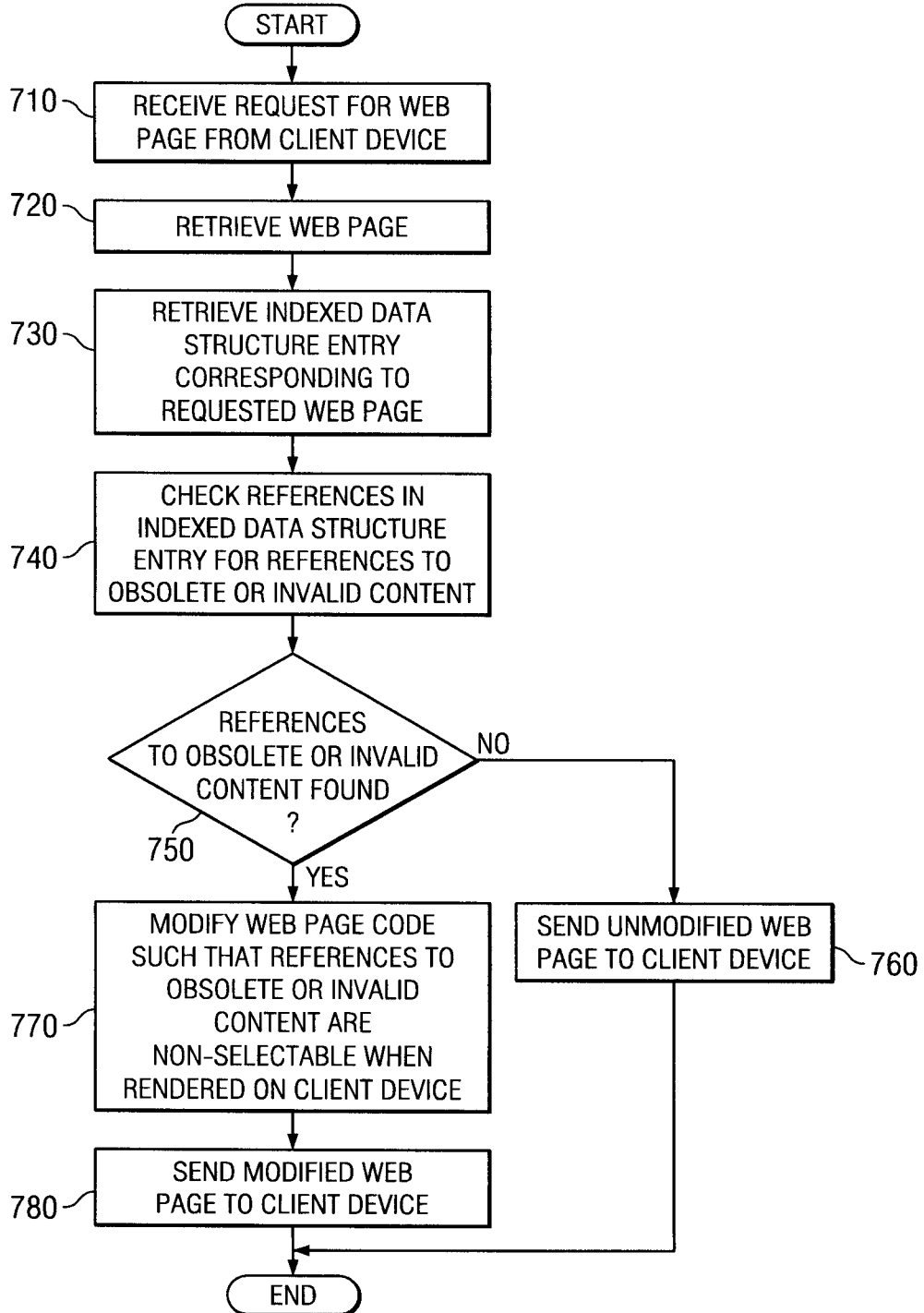


FIG. 7



## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2007/050213

A. CLASSIFICATION OF SUBJECT MATTER  
 INV. G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 578 078 B1 (SMITH MICHAEL D [US] ET AL) 10 June 2003 (2003-06-10) figures 9A-9C column 16, line 49 - column 18, line 17 ----- -/--	1-37



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\* & \* document member of the same patent family

Date of the actual completion of the international search

20 March 2007

Date of mailing of the international search report

03/04/2007

Name and mailing address of the ISA/

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

Authorized officer

Michalski, Stéphane

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2007/050213

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DAVIS H C ED - GRONBAEK K ET AL ASSOCIATION FOR COMPUTING MACHINERY: "REFERENTIAL INTEGRITY OF LINKS IN OPEN HYPERMEDIA SYSTEMS" HYPERTEXT '98. THE 9TH ACM CONFERENCE ON HYPERTEXT AND HYPERMEDIA. PITTSBURGH, PA, JUNE 20 - 24, 1998, ACM CONFERENCE ON HYPERTEXT AND HYPERMEDIA, NEW YORK, NY : ACM, US, 24 June 1998 (1998-06-24), pages 207-216, XPO01197248 ISBN: 0-89791-972-6 figures 2,3 page 210, left-hand column, paragraph 4 page 211, left-hand column, paragraph 7 - page 212, left-hand column, paragraph 1 page 212, right-hand column, last paragraph - page 213, left-hand column, paragraph 1</p>	1-37
X	<p>----- PITKOW J E ET AL: "Supporting the Web: A distributed hyperlink database system" COMPUTER NETWORKS AND ISDN SYSTEMS, NORTH HOLLAND PUBLISHING. AMSTERDAM, NL, vol. 28, no. 11, May 1996 (1996-05), pages 981-991, XPO04018201 ISSN: 0169-7552 abstract page 984, left-hand column - page 986, right-hand column figure 2</p>	1-37
X	<p>----- CREECH M L: "Author-oriented link management" COMPUTER NETWORKS AND ISDN SYSTEMS, NORTH HOLLAND PUBLISHING. AMSTERDAM, NL, vol. 28, no. 11, May 1996 (1996-05), pages 1015-1025, XPO04018204 ISSN: 0169-7552 page 1017, right-hand column, last paragraph - page 1018, left-hand column, paragraph 1 page 1016, right-hand column - page 1017, left-hand column</p> <p>-----</p>	1-37

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No  
PCT/EP2007/050213

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6578078	B1 10-06-2003	US 2004024848 A1	05-02-2004