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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 June 2005 (23.06.2005)

PCT

(10) International Publication Number
WO 2005/057826 A2

(51) International Patent Classification^7: H04L

(21) International Application Number: PCT/US2004/021107

(22) International Filing Date: 30 June 2004 (30.06.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/610,534 3 July 2003 (03.07.2003) US

(71) Applicants *(for all designated States except US)*: **BBNT SOLUTIONS LLC** [US/US]; 10 Moulton Street, Cambridge, MA 02138 (US). **VERIZON CORPORATE SERVICES GROUP INC.** [US/US]; 1095 Avenue of the Americas, New York, NY 10036 (US).

(71) Applicants and
(72) Inventors: **SHEPHARD, Scott** [US/US]; 552 Lincoln Street, Waltham, MA 02451 (US). **COLBATH, Sean** [US/US]; 12 Lilac Court, Cambridge, MA 02141 (US). **KUBALA, Francis, G.** [US/US]; 38 Hillcroft Road, Boston, MA 02130 (US).

(74) Agent: **SUCHTYA, Leonard, C.**; c/o Christian R. Andersen, Verizon Corporate Services Group Inc., 600 Hidden Ridge, Mailcode HQE03H14, Irving, TX 75038 (US).

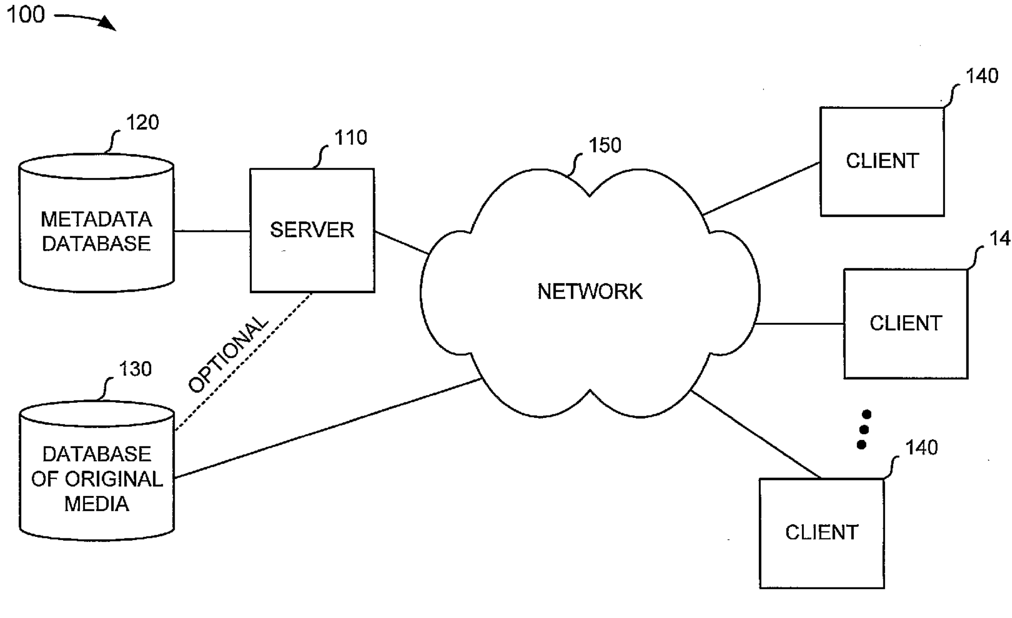
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) **Title:** SYSTEMS AND METHODS FOR FACILITATING PLAYBACK OF MEDIA



(57) **Abstract:** A system facilitates the browsing of information of interest. The system obtains a transcription of the information and provides the transcription to a user. The system also retrieves the information in its original format and presents the information to the user in the original format. The system visually synchronizes the presentation of the information in the original format with the transcription of the information.

SYSTEMS AND METHODS FOR FACILITATING PLAYBACK OF MEDIA

RELATED APPLICATIONS

This application claims priority under 35 U.S.C. § 119 based on U.S. Provisional Application Nos. 60/394,064 and 60/394,082, filed July 3, 2002, and Provisional Application No. 60/419,214, filed October 17, 2002, the disclosures of which are incorporated herein by reference.

This application is related to U.S. Patent Application, Serial No. 10/685,403 (Docket No. 02-4038), entitled, " MANAGER FOR INTEGRATING LANGUAGE TECHNOLOGY COMPONENTS " filed on October 16, 2003.

GOVERNMENT CONTRACT

The U.S. Government may have a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of Contract No. N66001-00-C-8008 awarded by the Defense Advanced Research Projects Agency (DARPA).

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to multimedia environments and, more particularly, to systems and methods for visually synchronizing the playback of any media (text, audio, video) with a textual representation of the media.

Description of Related Art

Much of the archived multimedia information that exists today is not easily manageable. For example, while mechanisms exist for searching and retrieving text, similar mechanisms do not exist for other types of media, such as audio or video. Audio and video from sources, such as television, radio, telephone, meetings, and presentations, have not been valued as archival sources due to the difficulty of locating information in large audio or video archives.

Recently, automatic content-based indexing and retrieval tools have been developed that may make audio and video sources as valuable an archival resource as text. These tools have made it easier to find audio or video sources of interest. The tools do not, however, facilitate the perusal of these audio or video sources. To browse an audio source, for example, a user must listen to the audio source to determine if it was the one the user desired. A user cannot do this much faster than the rate at which the audio was recorded.

Accordingly, there is a need for mechanisms that facilitate the perusal of media sources.

SUMMARY OF THE INVENTION

Systems and methods consistent with the present invention address this and other needs by visually synchronizing the playback of any media with a textual version of the media, thereby permitting a user to quickly skim or browse the media.

In one aspect consistent with the principles of the invention, a system facilitates the browsing of information of interest. The system obtains a transcription of the information and provides the transcription to a user. The system also retrieves the information in its original format and presents the information to the user in the original format. The system visually synchronizes the presentation of the information in the original format with the transcription of the information.

In another aspect consistent with the principles of the invention, a graphical user interface includes a transcription section, a speaker section, a topic section, and a request media button. The transcription section includes a transcription of non-text information. The speaker section identifies boundaries between speakers in the transcription section. The topic section includes one or more topics relating to the transcription. The request media button, when selected, causes retrieval of the non-text information to be initiated and the retrieved non-text information to be played. The request media button also causes the playing of the non-text information to be visually synchronized with the transcription in the transcription section.

BRIEF DESCRIPTION OF THE DRAWINGS

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

Fig. 1 is a diagram of a system in which systems and methods consistent with the present invention may be implemented;

Fig. 2 is an exemplary diagram of the server of Fig. 1 according to an implementation consistent with the principles of the invention;

Fig. 3 is an exemplary diagram of the metadata database of Fig. 1 according to an implementation consistent with the present invention;

Fig. 4 is an exemplary diagram of a metadata media file of Fig. 3 according to an implementation consistent with the principles of the invention;

Fig. 5 is an exemplary diagram of the database of original media of Fig. 1 according to an implementation consistent with the principles of the invention;

Fig. 6 is an exemplary diagram of the client of Fig. 1 according to an implementation consistent with the principles of the invention;

Fig. 7 is an exemplary diagram of a graphical user interface that may be presented via the client of Fig. 6 according to an implementation consistent with the principles of the invention;

Fig. 8 is a flowchart of exemplary processing for visually synchronizing the playback of an original media with a textual representation of the media;

Fig. 9 is a diagram of a graphical user interface that illustrates a user's request to play back an original media; and

Fig. 10 is a diagram of a graphical user interface that illustrates the synchronization of a HyperText Markup Language document to the playback of the original media.

DETAILED DESCRIPTION

The following detailed description of the invention refers to the accompanying drawings. The same reference numbers in different drawings may identify the same or similar elements. Also, the following detailed description does not limit the invention. Instead, the scope of the invention is defined by the appended claims and equivalents.

Systems and methods consistent with the present invention visually synchronize the playing back of a type of media, such as text, audio, and/or video, with a textual representation of the media. Such systems and methods permit a user to quickly browse the media in any language.

EXEMPLARY SYSTEM

Fig. 1 is a diagram of an exemplary system 100 in which systems and methods consistent with the present invention may be implemented. System 100 may include server 110, metadata database 120, database of original media 130, and clients 140 interconnected via a network 150. Network 150 may include any type of network, such as a local area network (LAN), a wide area network (WAN), a public telephone network (e.g., the Public Switched Telephone Network (PSTN)), a virtual private network (VPN), or a combination of networks. Server 110, database 130, and clients 140 may connect to network 150 via wired, wireless, and/or optical connections.

Generally, clients 140 may interact with server 110 to obtain information of interest from metadata database 120. A user of one of clients 140 may peruse the information and obtain the original media from database of original media 130 either directly or via server 110. Client 140 may present the information and original media to the user in such a manner that facilitates the user's perusal of the information.

Each of the components of system 100 will now be described in more detail.

Server 110

Server 110 may include a computer or another device that is capable of servicing client requests for information and providing such information to a client 140, possibly in the form of a HyperText Markup Language (HTML) document or web page. Fig. 2 is an exemplary diagram of server 110 according to an implementation consistent with the principles of the invention. Server 110 may include bus 210, processor 220, main memory 230, read only memory (ROM) 240, storage device 250, input device 260, output device 270, and communication interface 280. Bus 210 permits communication among the components of server 110.

Processor 220 may include any type of conventional processor or microprocessor that interprets and executes instructions. Main memory 230 may include a random access memory (RAM) or another type of dynamic storage device that stores information and instructions for execution by processor 220. ROM 240 may include a conventional ROM device or another type of static storage device that stores static information and instructions for use by processor 220. Storage device 250 may include a magnetic and/or optical recording medium and its corresponding drive.

Input device 260 may include one or more conventional mechanisms that permit an operator to input information to server 110, such as a keyboard, a mouse, a pen, voice recognition and/or biometric mechanisms, etc. Output device 270 may include one or more conventional mechanisms that output information to the operator, including a display, a printer, a pair of speakers, etc. Communication interface 280 may include any transceiver-like mechanism that enables server 110 to communicate with other devices and/or systems. For example, communication interface 280 may include mechanisms for communicating with another device or system via a network, such as network 150.

As will be described in detail below, server 110, consistent with the present invention, services requests for information and manages access to metadata database 120. Server 110 may perform these tasks in response to processor 220 executing sequences of instructions contained in, for example, memory 230. These instructions may be read into memory 230 from another computer-readable medium, such as storage device 250, or from another device via communication interface 280.

Execution of the sequences of instructions contained in memory 230 causes processor 220 to perform processes that will be described later. Alternatively, hardwired circuitry may be used in place of or in combination with software instructions to implement processes consistent with the present invention. Thus, processes performed by server 110 are not limited to any specific combination of hardware circuitry and software.

Metadata Database 120

Metadata database 120 may include a conventional database that stores metadata relating to any type of media in any language. A media processing system (not shown), such as the one described in John Makhoul et al., "Speech and Language Technologies for Audio Indexing and Retrieval," Proceedings of the IEEE, Vol. 88, No. 8, August 2000, pp. 1338-1353, may collect media from various sources, process the media, and create metadata relating to the original media.

In the case of audio or video, the media processing system may segment an input stream by speaker, cluster audio segments from the same speaker, identify speakers known to the system, and transcribe the spoken words. The media processing system may also segment the input stream into stories, based on their topic content, and locate the names of people, places, and organizations. The media processing system may further analyze the input stream to identify when each word is spoken. The media processing system may include any or all of this information in the metadata relating to the input stream.

Metadata database 120 may store metadata in files or tables. Fig. 3 is an exemplary diagram of metadata database 120 according to an implementation consistent with the principles of the invention. Metadata database 120 may include multiple metadata media files 310. Each of media files 310 may store metadata relating to a story or an episode (i.e., a collection of stories within an input stream). The metadata may differ depending on the type of media to which it corresponds. For a text input stream, for example, the metadata may include information relating to an author or publisher of the text. For an audio input stream, the metadata may include information regarding a speaker, or speakers, or a source of the audio. For a video input stream, the metadata may include information regarding one or more persons in the video (speaking or non-speaking) or a source of the video.

Fig. 4 is a diagram of an exemplary metadata media file 310 according to an implementation consistent with the principles of the invention. Media file 310 in Fig. 4 relates to an audio input stream from National Public Radio (NPR) Morning Edition on February 11, 2002, that began at 6:00 a.m. The metadata in media file 310 may include information 410 regarding the type of media involved (audio) and information 420 that identifies the source of the input stream (NPR Morning Edition). The metadata may also include data 430 that identifies relevant topics, data 440 that identifies speaker gender, and data 450 that identifies names of people, places, or organizations. The metadata may further include time data 460 that identifies the start and duration of each word spoken.

Database of Original Media 130

Database of original media 130 may include a conventional database that stores any type of media in any language. The media stored in database 130 may correspond to the metadata in metadata database 120. In other words, the original media may include the data from which the metadata was created. In other implementations, database 130 may contain additional media for which there is no corresponding metadata in metadata database 120.

Fig. 5 is an exemplary diagram of database of original media 130 according to an implementation consistent with the principles of the invention. Database 130 may include multiple original media files 510. Each of media files 510 may store data from an original input stream. For example, a media file 510 may correspond to an audio stream. In this case, the audio stream may be processed by a known audio compression technique, such as MP3 compression, and stored in media file 510. Another media file 510 may correspond to a video stream. In this case, the video stream may be processed by a known video compression technique, such as MPEG compression, and stored in media file 510. Yet another media file 510 may correspond to a text stream, such as news wire. In this case, the text stream may be processed by a known text compression technique and stored in media file 510. Where storage space is not limited, the media may be stored uncompressed.

The original media may be stored in such a way that it is easily retrievable as a whole and in portions. For example, a portion of an audio file may be retrieved by specifying that the portion of the file that occurred between 8:05 a.m. and 8:08 a.m. is desired. The database 130 may then provide the desired audio as streaming audio to client 140, for example.

Client 140

Client 140 may include a personal computer, a laptop, a personal digital assistant, or another type of device that is capable of interacting with server 110 and database of original media 130 to obtain information of interest. Client 140 may present the information to a user via a graphical user interface (GUI), possibly within a web browser window.

Fig. 6 is an exemplary diagram of client 140 according to an implementation consistent with the principles of the invention. Client 140 may include a bus 610, a processor 620, a memory 630, one or more input devices 640, one or more output devices 650, and a communication interface 660. Bus 610 may permit communication among the components of client 140.

Processor 620 may include any type of conventional processor or microprocessor that interprets and executes instructions. Memory 630 may include a RAM or another type of dynamic storage device that stores information and instructions for execution by processor 620; a ROM or another type of static storage device that stores static information and instructions for use by processor 620; and/or some other type of magnetic or optical recording medium and its corresponding drive. For example, memory 630 may include both long term and short term memory devices.

Input devices 640 may include one or more conventional mechanisms that permit a user to input information into client 140, such as a keyboard, mouse, pen, etc. Output devices 650 may include one or more conventional mechanisms that output information to the user, including a display, a printer, a pair of speakers, etc. Communication interface 660 may include any transceiver-like mechanism that enables client 140 to communicate with other devices and systems via a network, such as network 150.

As will be described in detail below, client 140, consistent with the present invention, visually synchronizes the playing back of a type of media, such as text, audio, and/or video, with a textual representation of the media. Client 140 may perform these operations in response to processor 620 executing software instructions contained in a computer-readable medium, such as memory 630. The software instructions may be read into memory 630 from another computer-readable medium or from another device via communication interface 660. The software instructions contained in memory 630 causes processor 620 to perform processes that will be described later. Alternatively, hardwired circuitry may be used in place of or in combination with software instructions to implement processes consistent with the present invention. Thus, processes performed by client 140 are not limited to any specific combination of hardware circuitry and software.

In an implementation consistent with the principles of the invention, client 140 provides a textual representation of a desired media in any language via a graphical user interface (GUI). Fig. 7 is a diagram of an exemplary GUI 700 that client 140 may present to a user according to an implementation consistent with the principles of the invention. GUI 700 may be part of an interface of a standard Internet browser, such as Internet Explorer or Netscape Navigator, or any browser that follows World Wide Web Consortium (W3C) specifications for HTML. The information presented by GUI 700 in this example relates to an episode of a television news program (i.e., ABC's World News Tonight from January 31, 1998).

GUI 700 may include a speaker section 710, a transcription section 720, and a topics section 730. Speaker section 710 may identify boundaries between speakers, the gender of a speaker, and the name of a speaker (when known). In this way, speaker segments are clustered together over the entire episode to group together segments from the same speaker under the same label. In the example of Fig. 7, one speaker, Elizabeth Vargas, has been identified by name.

Transcription section 720 may include a transcription of the desired media. Transcription section 720 may identify the names of people, places, and organizations by highlighting them in some manner. For example, people, places, organizations may be identified using different colors. Topic section 730 may include topics relating to the transcription in transcription section 720. Each of the topics may describe the main themes of the episode and may constitute a very high-level summary of the content of the transcription, even though the exact words in the topic may not be included in the transcription.

GUI 700 may also include a request media (RM) icon 740 corresponding to an embedded media player, such as the RealPlayer media player available from RealNetworks, that permits the original media corresponding to the transcription in transcription section 720 to be played back. When instructed to do so, such as when a user selects icon 740, the media player may access database of original media 130 to retrieve the original media and present the original media to the user. For example, if the original media is an audio stream, the media player may permit the original audio to be played. Similarly, if the original media is a video stream, the media player may permit the original video to be played. If the original media is a text stream, the media player may present the original text document.

EXEMPLARY PROCESSING

Fig. 8 is a flowchart of exemplary processing for visually synchronizing the playback of an original media with a textual representation of the media. Processing may begin with a user inputting, into client 140, a request for desired information. The information desired by the user may have originated in any form (e.g., text, audio, or video) and in any language (e.g., English, Chinese, or Arabic). A typical request may be as specific as "give me ABC's World News Tonight for January 3, 1998," or as general as "show me everything where Bill Clinton was the topic." Other requests may include data regarding the date, time, and source of the desired information, or relevant words next to each other or within a certain distance of each other (similar to a typical database query).

Client 140 may process (e.g., convert) the request, if necessary, and issue the request to server 110 (act 805). For example, client 140 may establish communication with server 110 via network 150, using conventional techniques. Once communication has been established, client 140 may transmit the request to server 110.

Server 110 may formulate a query based on the request from client 140 and use the query to access metadata database 120. Server 110 may retrieve metadata relating to the desired information from metadata database 120 (act 810). Server 110 may then convert the metadata to an appropriate form, such as an HTML document, and transmit the HTML document to client 140 for display in a standard web browser (acts 815 and 820). The HTML document may contain the original metadata information, such as speaker identifiers, topics, and word time codes. In other implementations, server 110 may convert the metadata to another form or transmit the metadata unconverted to client 140.

Client 140 may present the HTML document to the user via a GUI, such as GUI 700 (act 825). The user may read, skim, or browse the HTML document. At some point, the user may express a desire to play back the information in the HTML document in its original form (act 830). In this case, the user may highlight or otherwise identify a portion of the HTML document for which the user desires to obtain the original media and select request media icon 740. For example, the user may use a computer mouse to highlight the desired portion. Alternatively, the user may simply identify a starting point from which the original media is desired.

Fig. 9 is a diagram of GUI 700 that illustrates a user's request to play back an original media. The user highlights a portion of the HTML document at highlighted block 910. The user selects the request media icon 920 to initiate the playback process.

Returning to Fig. 8, when the user selects request media icon 740 (Fig. 7) client 140 initiates the embedded media player. The media player may determine the portion identified by the user, such as highlighted portion 910 (act 835). In particular, the media player may identify the time codes, corresponding to the beginning and ending (if applicable) of the identified portion, using the time codes in the HTML document.

The media player may then retrieve the desired portion of the original media (act 840). The media player may use conventional techniques to pull that portion of the original media from database of original media 130. For example, the media player may use the beginning and ending time codes (e.g., 7:03 p.m. to 7:05 p.m.) when accessing database 130. The original media from database 130 streams back to the media player. The media player then plays the original media for the user (act 845).

As the media player plays back the original media, GUI 700 visually synchronizes the playback with the transcription in the HTML document (act 850). To facilitate this, the media player lets client 140 know as time passes in the playback of the original media. Because the metadata of the HTML document includes time codes that identify exactly when each word in the transcription of the HTML document was spoken, client 140 knows precisely (possibly down to the millisecond) when to highlight (or otherwise visually distinguish) a word. Client 140 compares the times emitted by the media player with the time codes and highlights the appropriate words.

Fig. 10 is a diagram of GUI 700 that illustrates the synchronization of the HTML document to the playback of the original media. Client 140 visually distinguishes the word "american" in synchronism with the playback of the original media (audio, video) by the media player, as shown at the highlighted block 1010.

The user may be permitted to stop the playback at any time. The user may also be permitted to control the playback by, for example, fast forwarding, speeding it up, slowing it down, or backing it up so many seconds or so many words. The media player or the graphical user interface may present the user with a set of controls to permit the user to perform these functions.

The user may also be permitted to alter the HTML document in some manner and save the altered document back in metadata database 120. For example, the user may be permitted to highlight or comment on the document. Client 140, in this case, may send the altered document back to server 110 for storage in metadata database 120.

CONCLUSION

Systems and methods consistent with the present invention visually synchronize the playing back of a type of media, such as text, audio, and/or video, with a textual representation of the media. The systems and methods may highlight or otherwise visually distinguish words in the textual representation in synchronization with the playing back of the media. Such systems and methods permit a user to quickly browse the media in any language.

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

For example, it has been disclosed that a media player retrieves the original media once initiated by the client. In other implementations, the original media may be transmitted to the client along with the HTML document containing the metadata. In yet other implementations, more than the requested portion of the original media may be transmitted to the client in anticipation of its later request by the user.

It may also be possible to send the HTML document to the client without time codes. In this case, the client would need to request the time codes of the selected portion so that the playback of the original media can be synchronized with the textual representation of the media.

No element, act, or instruction used in the description of the present application should be construed as critical or essential to the invention unless explicitly described as such. Also, as used herein, the article "a" is intended to include one or more items. Where only one item is intended, the term "one" or similar language is used. The scope of the invention is defined by the claims and their equivalents.

WHAT IS CLAIMED IS:

1. A method for facilitating perusal of an item of interest, comprising:

retrieving a textual representation of the item;

presenting the textual representation to a user;

obtaining an original form of the item;

providing the item to the user in the original form; and

visually synchronizing the providing of the item in the original form with the textual representation of the item.

2. The method of claim 1, wherein the retrieving a textual representation includes:

generating a request concerning the item of interest,

sending the request to a server, and

obtaining, from the server, the textual representation of the item.

3. The method of claim 2, wherein the obtaining the textual representation includes:

using the request, by the server, to retrieve metadata relating to the item from a metadata database,

generating the textual representation of the item from the metadata, and

receiving, from the server, the generated textual representation.

4. The method of claim 3, wherein the generating the textual representation includes:

creating a HyperText Markup Language document from the metadata.

5. The method of claim 1, wherein the presenting the textual representation includes:

providing the textual representation within a graphical user interface of a web browser.

6. The method of claim 1, wherein the obtaining an original form of the item includes:

accessing a database of original media to retrieve the item in the original form.

7. The method of claim 1, wherein the obtaining an original form of the item includes:

receiving input, from the user, regarding a desire for the item in the original form,

initiating a media player, and

using the media player to obtain the item in the original form.

8. The method of claim 7, wherein the receiving input from the user includes:

receiving selection of a portion of the textual representation.

9. The method of claim 8, wherein the using the media player includes:

determining, by the media player, the portion selected by the user, and

retrieving the item in the original form corresponding to the determined portion.

10. The method of claim 9, wherein the determining the portion includes:

identifying time codes associated with a beginning and an ending of the selected portion.

11. The method of claim 9, wherein the portion selected by the user includes a starting position in the textual representation; and

wherein the determining the portion includes:

identifying time codes associated with the starting position in the textual representation.

12. The method of claim 1, wherein the textual representation includes time codes corresponding to when words in the textual representation were spoken.

13. The method of claim 12, wherein the visually synchronizing the providing of the item includes:

comparing times corresponding to the providing of the item in the original form to the time codes from the textual representation, and

visually distinguishing words in the textual representation when the words are spoken during the providing of the item in the original form.

14. The method of claim 1, wherein the providing the item to the user includes:

permitting the user to control the providing of the item in the original form.

15. The method of claim 14, wherein the permitting the user to control the providing includes:

allowing the user to at least one of fast forward, speed up, slow down, and back up the providing of the item in the original form.

16. The method of claim 1, wherein the item is an audio file and the textual representation of the item includes a transcription of the audio file and at least one of a speaker identifier, a topic, and one or more word time codes.

17. The method of claim 1, wherein the item is a video file and the textual representation of the item includes a transcription of the video file and at least one of a speaker identifier, a topic, and one or more word time codes.

18. The method of claim 1, wherein the original form of the item includes a format in which the item was originally created.

19. A system for facilitating browsing of an item of interest, comprising:

means for obtaining a transcription of the item;

means for providing the transcription to a user;

means for retrieving the item in an original form;

means for presenting the item to the user in the original form; and

means for visually synchronizing the presenting of the item in the original form with the transcription of the item.

20. A system for aiding a user in browsing information of interest, comprising:

a memory configured to store instructions; and

a processor configured to execute the instructions in memory to:

obtain a transcription of the information,

provide the transcription to a user,

retrieve the information in an original format,

present the information to the user in the original format, and

visually synchronize the presentation of the information in the original format with the transcription of the information.

21. The system of claim 20, wherein when obtaining a transcription, the processor is configured to:

generate a request concerning the information of interest,

send the request to a server, and

obtain, from the server, the transcription of the information.

22. The system of claim 20, wherein when providing the transcription, the processor is configured to:

present the transcription within a graphical user interface of a web browser.

23. The system of claim 20, wherein when retrieving the information in an original format, the processor is configured to:

obtain the information from a database of original media.

24. The system of claim 20, wherein when retrieving the information in an original format, the processor is configured to:

receive input, from the user, regarding a desire for the information in the original format, and

initiate a media player to obtain the information in the original format.

25. The system of claim 20, wherein when retrieving the information in an original format, the processor is configured to:

receive input, from the user, regarding a desire for the information in the original format,

receive selection of a portion of the transcription by the user, and

obtain the information in the original format corresponding to the selected portion.

26. The system of claim 25, wherein when obtaining the information in the original format, the processor is configured to:

identify time codes associated with a beginning and an ending of the selected portion.

27. The system of claim 25, wherein the portion selected by the user includes a starting position in the transcription; and

wherein when obtaining the information in the original format, the processor is configured to:

identify time codes associated with the starting position in the transcription.

28. The system of claim 20, wherein the transcription includes time codes corresponding to when words in the transcription were spoken.

29. The system of claim 28, wherein when visually synchronizing the presentation of the information, the processor is configured to:

compare times corresponding to the presentation of the information in the original format to the time codes from the transcription, and

visually distinguish words in the transcription when the words are played back during the presentation of the information in the original format.

30. The system of claim 20, wherein when presenting the information to the user, the processor is configured to:

permit the user to control the presentation of the information in the original format.

31. The system of claim 30, wherein when permitting the user to control the presentation of the information, the processor is configured to:

provide controls to the user to allow the user to at least one of fast forward, speed up, slow down, and back up the presentation of the information in the original format.

32. The system of claim 20, wherein the information is an audio file and the transcription of the information includes a transcription of the audio file and at least one of a speaker identifier, a topic, and one or more word time codes.

33. The system of claim 20, wherein the information is a video file and the transcription of the information includes a transcription of the video file and at least one of a speaker identifier, a topic, and one or more word time codes.

34. The system of claim 20, wherein the original format of the information includes a form in which the information was originally created.

35. A computer-readable medium that contains instructions for causing at least one processor to perform a method for facilitating browsing of audio and video information, comprising:

instructions for retrieving a textual representation of the information;

instructions for presenting the textual representation to a user;

instructions for obtaining the information in an original format;

instructions for providing the information to the user in the original format; and

instructions for visually synchronizing the providing of the information in the original format with the textual representation of the information.

36. A graphical user interface, comprising:

a transcription section that includes a transcription of non-text information;

a speaker section that identifies boundaries between speakers in the transcription section;

a topic section that includes one or more topics relating to the transcription; and

a request media button that, when selected, causes:

retrieval of the non-text information to be initiated,

playing of the non-text information, and the playing of the non-text information to be visually synchronized with the transcription in the transcription section.

37. The graphical user interface of claim 36, wherein the transcription visually distinguishes names of people, places, and organizations.

38. The graphical user interface of claim 36, wherein the speaker section further includes at least one of gender and names of the speakers.

39. The graphical user interface of claim 36, wherein the one or more topics relate to one or more main themes of the transcription.

40. The graphical user interface of claim 36, wherein the transcription includes time codes that identify when words in the transcription were spoken with regard to the non-text information.

41. The graphical user interface of claim 40, wherein the request media button causes words in the transcription to be visually distinguished in synchronism with the words in the non-text information being played.

42. The graphical user interface of claim 36, wherein the non-text information includes at least one of audio and video.

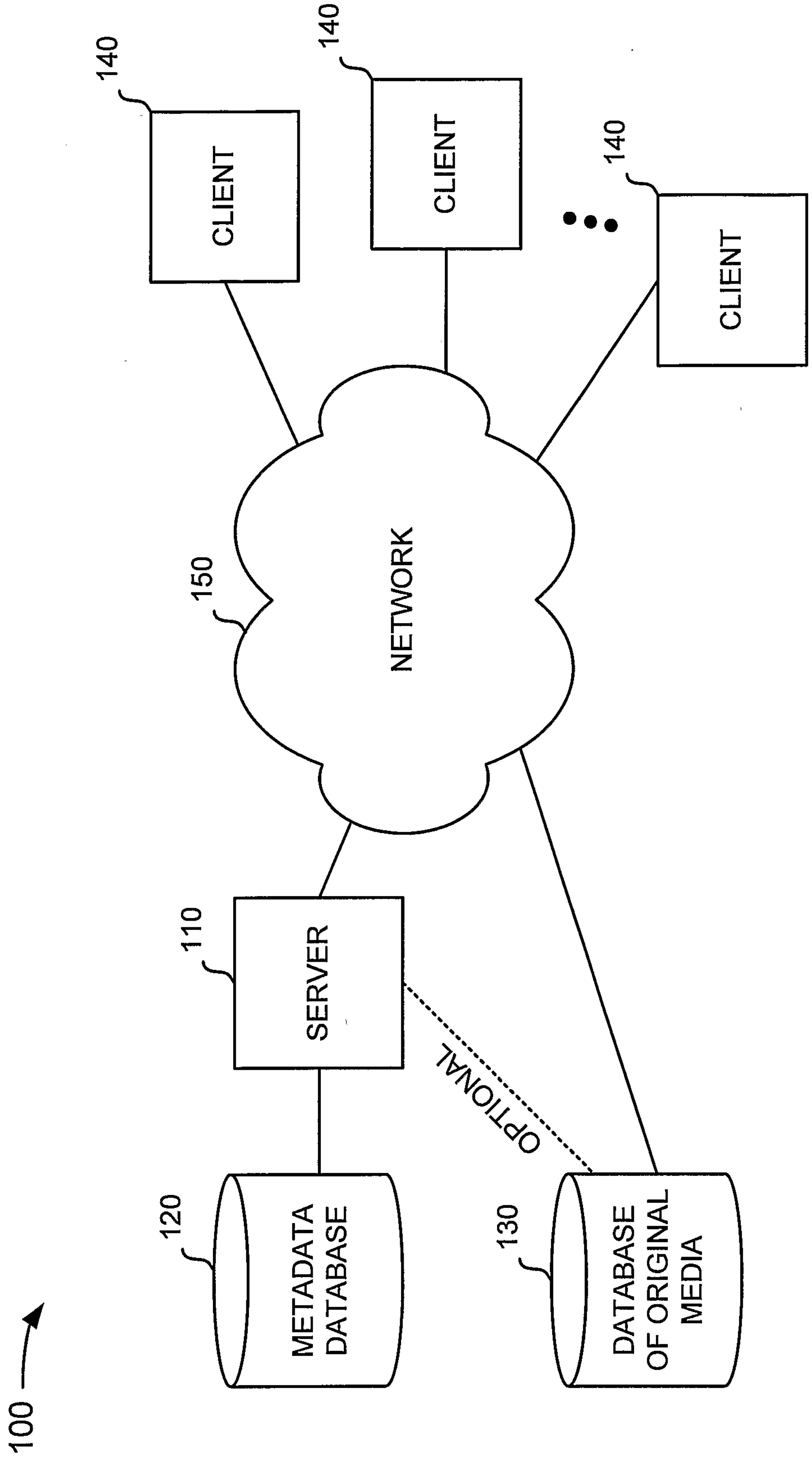

100
120
METADATA DATABASE
130
DATABASE OF ORIGINAL MEDIA
110
SERVER
OPTIONAL
150
NETWORK
140
CLIENT
140
CLIENT
140
CLIENT

FIG. 1

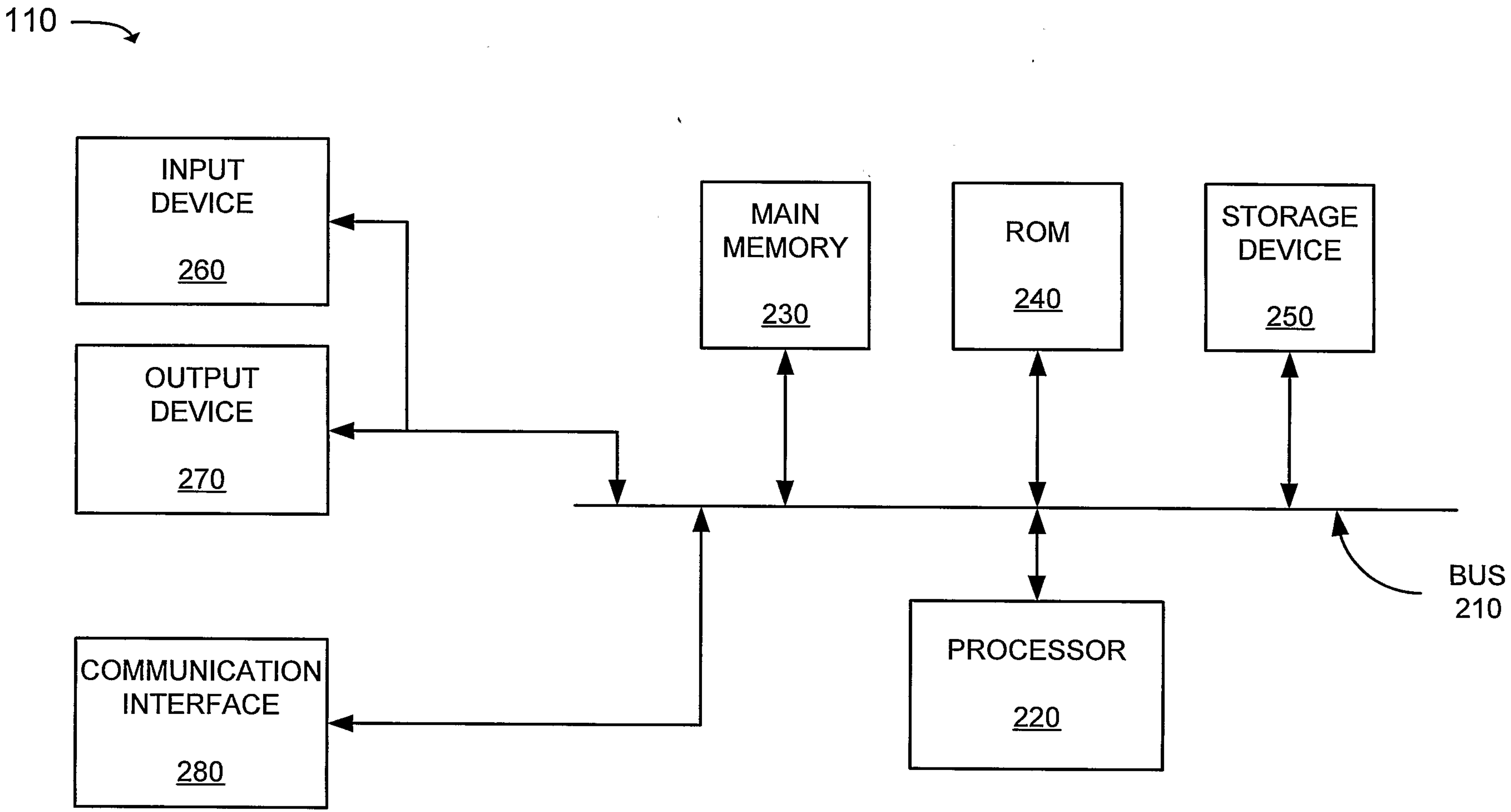

110
INPUT DEVICE 260
OUTPUT DEVICE 270
COMMUNICATION INTERFACE 280
MAIN MEMORY 230
ROM 240
STORAGE DEVICE 250
PROCESSOR 220
BUS 210

FIG. 2

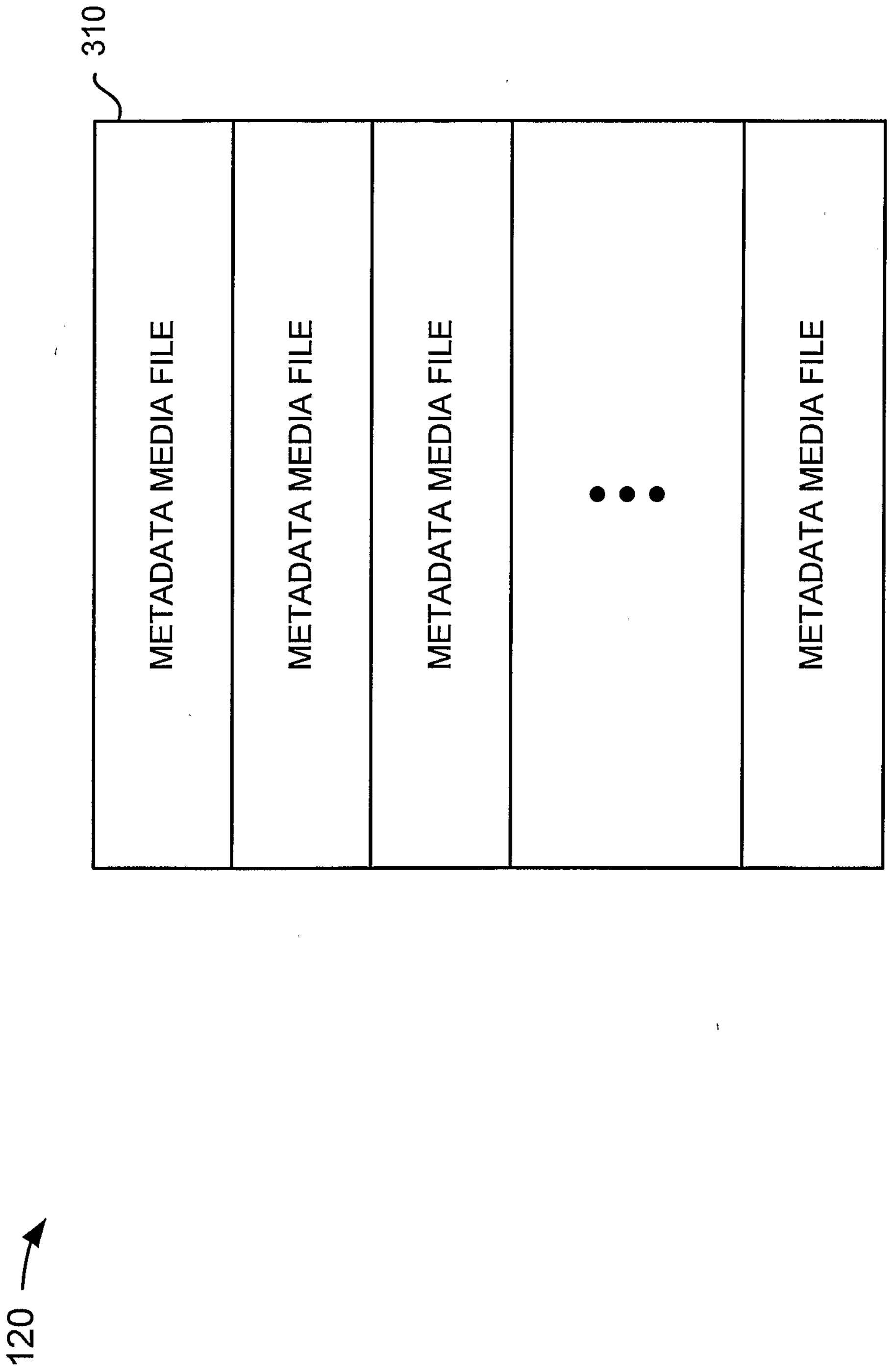
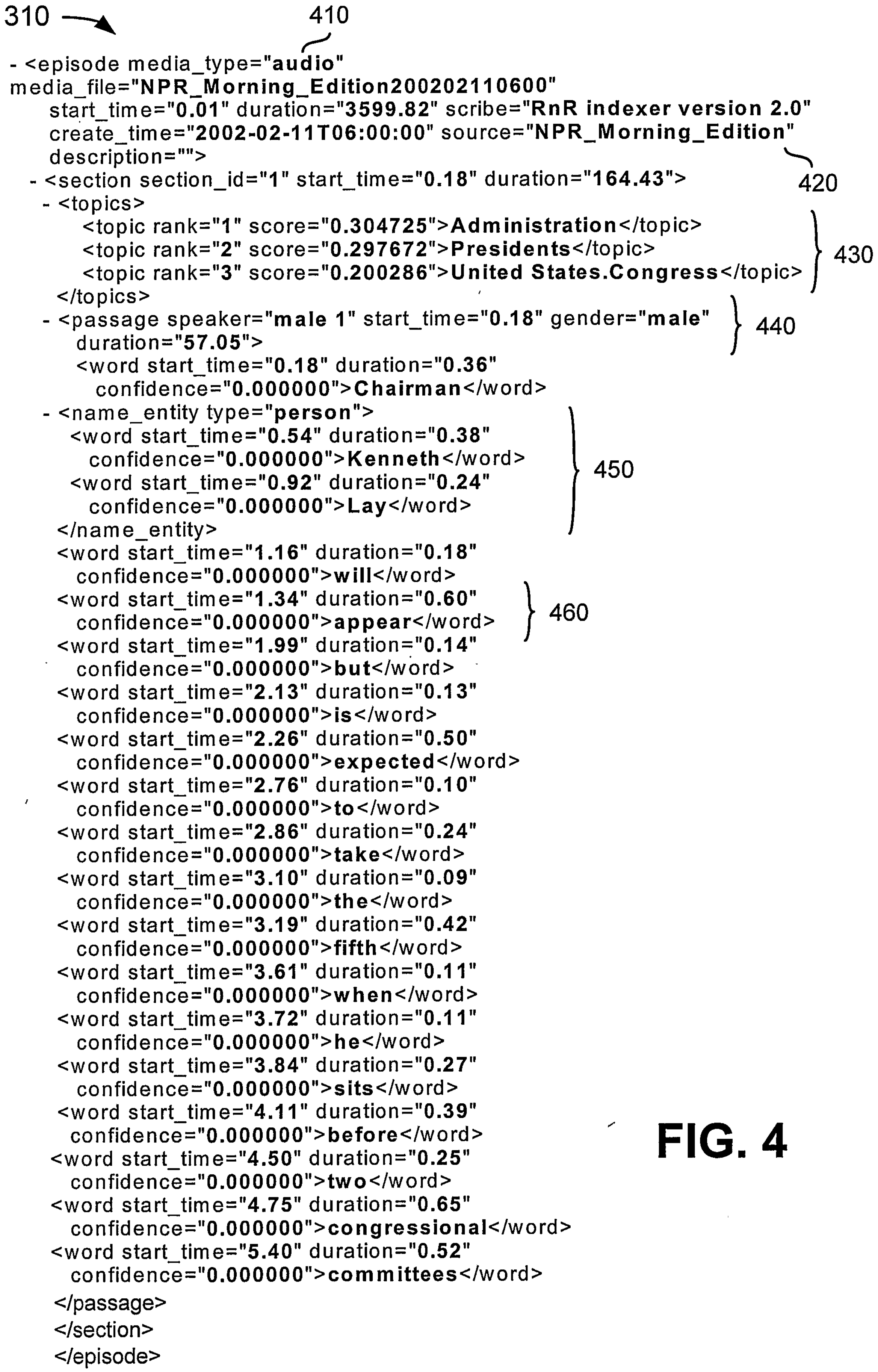

310
METADATA MEDIA FILE
METADATA MEDIA FILE
METADATA MEDIA FILE
• • •
METADATA MEDIA FILE
120

FIG. 3



```
- <episode media_type="audio"
media_file="NPR_Morning_Edition200202110600"
    start_time="0.01" duration="3599.82" scribe="RnR indexer version 2.0"
    create_time="2002-02-11T06:00:00" source="NPR_Morning_Edition"
    description="">
  - <section section_id="1" start_time="0.18" duration="164.43">
    - <topics>
        <topic rank="1" score="0.304725">Administration</topic>
        <topic rank="2" score="0.297672">Presidents</topic>
        <topic rank="3" score="0.200286">United States.Congress</topic>
      </topics>
    - <passage speaker="male 1" start_time="0.18" gender="male"
        duration="57.05">
        <word start_time="0.18" duration="0.36"
          confidence="0.000000">Chairman</word>
    - <name_entity type="person">
        <word start_time="0.54" duration="0.38"
          confidence="0.000000">Kenneth</word>
        <word start_time="0.92" duration="0.24"
          confidence="0.000000">Lay</word>
      </name_entity>
      <word start_time="1.16" duration="0.18"
        confidence="0.000000">will</word>
      <word start_time="1.34" duration="0.60"
        confidence="0.000000">appear</word>
      <word start_time="1.99" duration="0.14"
        confidence="0.000000">but</word>
      <word start_time="2.13" duration="0.13"
        confidence="0.000000">is</word>
      <word start_time="2.26" duration="0.50"
        confidence="0.000000">expected</word>
      <word start_time="2.76" duration="0.10"
        confidence="0.000000">to</word>
      <word start_time="2.86" duration="0.24"
        confidence="0.000000">take</word>
      <word start_time="3.10" duration="0.09"
        confidence="0.000000">the</word>
      <word start_time="3.19" duration="0.42"
        confidence="0.000000">fifth</word>
      <word start_time="3.61" duration="0.11"
        confidence="0.000000">when</word>
      <word start_time="3.72" duration="0.11"
        confidence="0.000000">he</word>
      <word start_time="3.84" duration="0.27"
        confidence="0.000000">sits</word>
      <word start_time="4.11" duration="0.39"
       confidence="0.000000">before</word>
     <word start_time="4.50" duration="0.25"
       confidence="0.000000">two</word>
     <word start_time="4.75" duration="0.65"
       confidence="0.000000">congressional</word>
     <word start_time="5.40" duration="0.52"
       confidence="0.000000">committees</word>
     </passage>
     </section>
     </episode>
```

FIG. 4

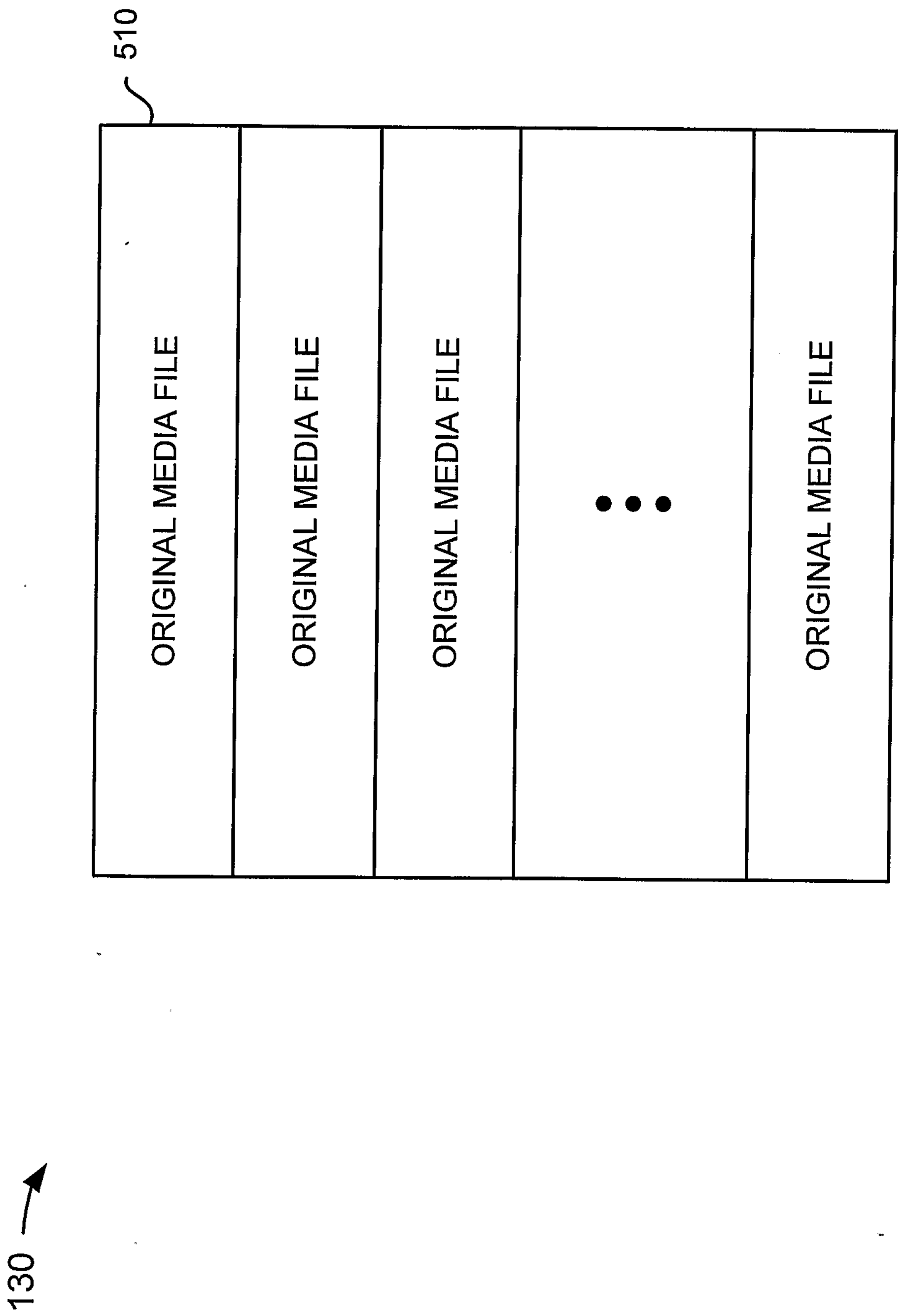

510
ORIGINAL MEDIA FILE
ORIGINAL MEDIA FILE
ORIGINAL MEDIA FILE
• • •
ORIGINAL MEDIA FILE
130

FIG. 5

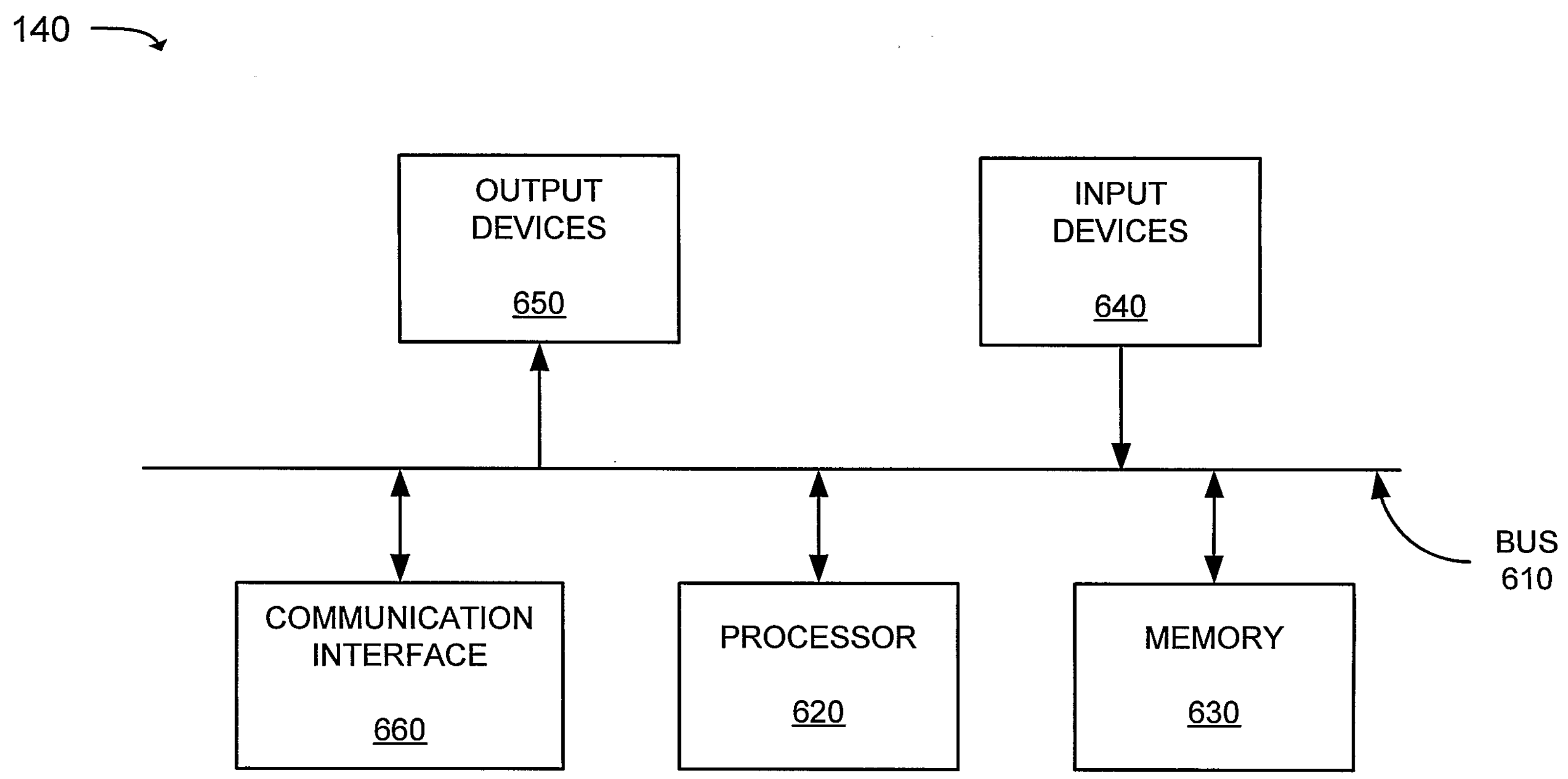
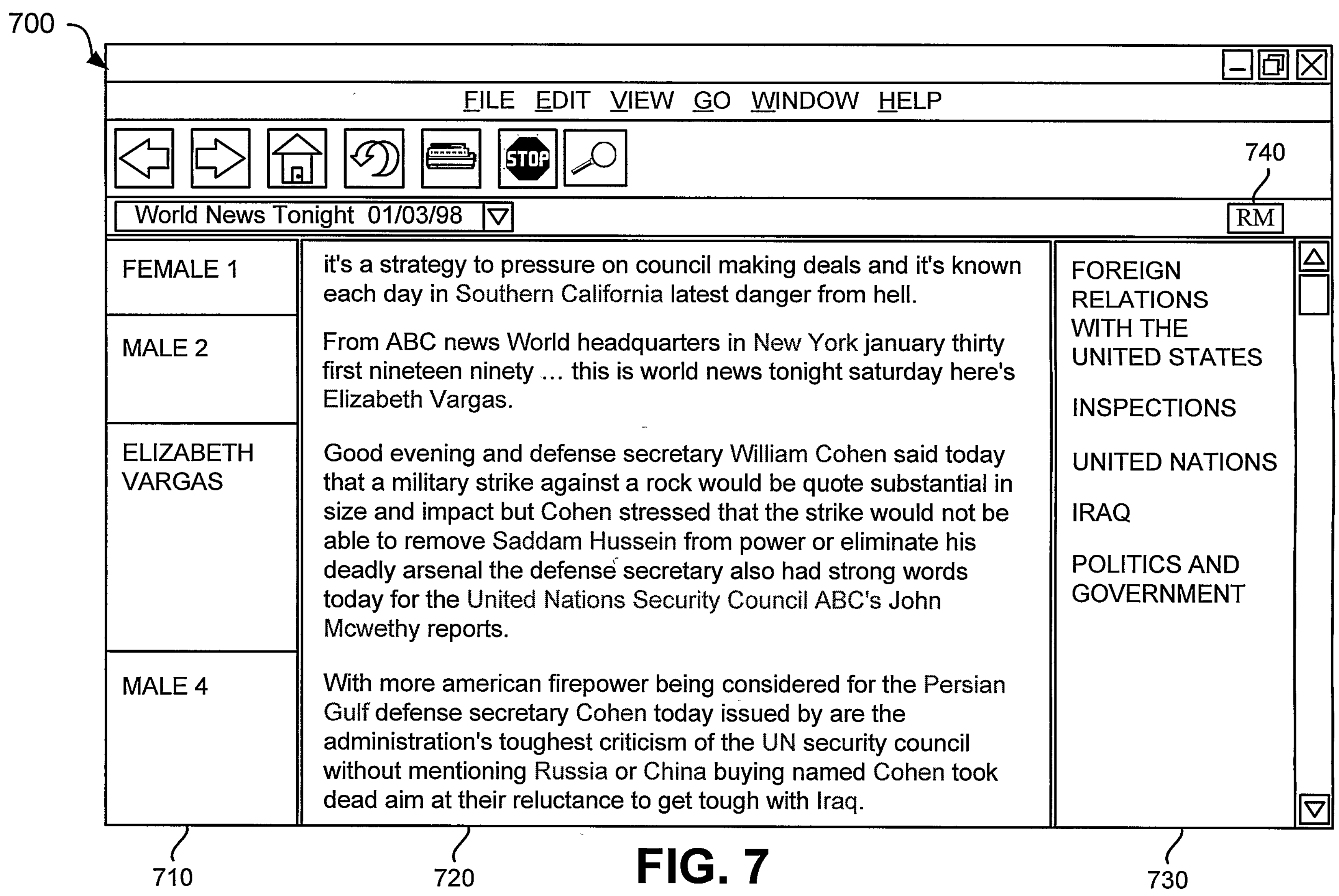

140
OUTPUT DEVICES 650
INPUT DEVICES 640
BUS 610
COMMUNICATION INTERFACE 660
PROCESSOR 620
MEMORY 630

FIG. 6



FEMALE 1	it's a strategy to pressure on council making deals and it's known each day in Southern California latest danger from hell.
MALE 2	From ABC news World headquarters in New York january thirty first nineteen ninety … this is world news tonight saturday here's Elizabeth Vargas.
ELIZABETH VARGAS	Good evening and defense secretary William Cohen said today that a military strike against a rock would be quote substantial in size and impact but Cohen stressed that the strike would not be able to remove Saddam Hussein from power or eliminate his deadly arsenal the defense secretary also had strong words today for the United Nations Security Council ABC's John Mcwethy reports.
MALE 4	With more american firepower being considered for the Persian Gulf defense secretary Cohen today issued by are the administration's toughest criticism of the UN security council without mentioning Russia or China buying named Cohen took dead aim at their reluctance to get tough with Iraq.

FOREIGN RELATIONS WITH THE UNITED STATES

INSPECTIONS

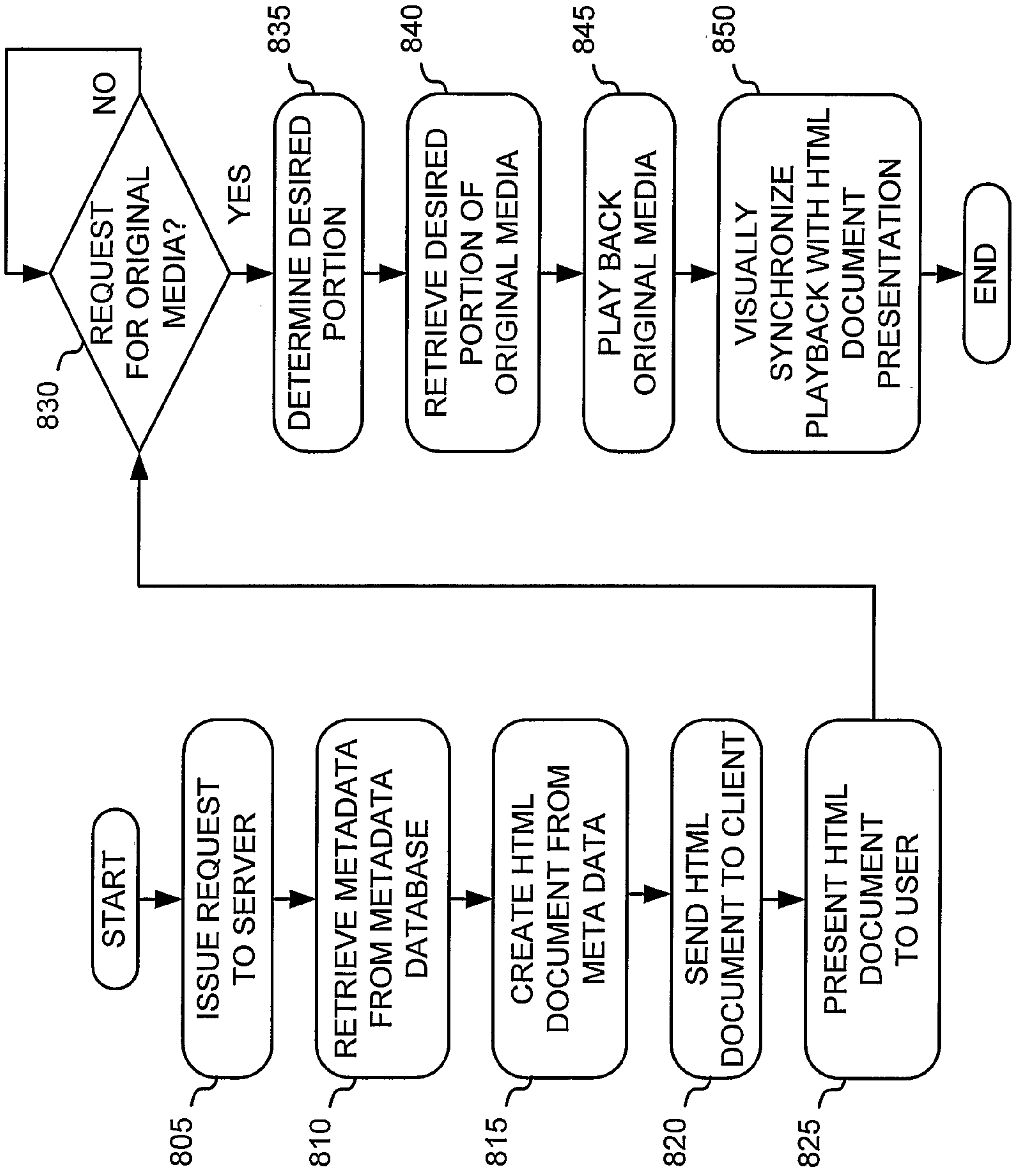
UNITED NATIONS

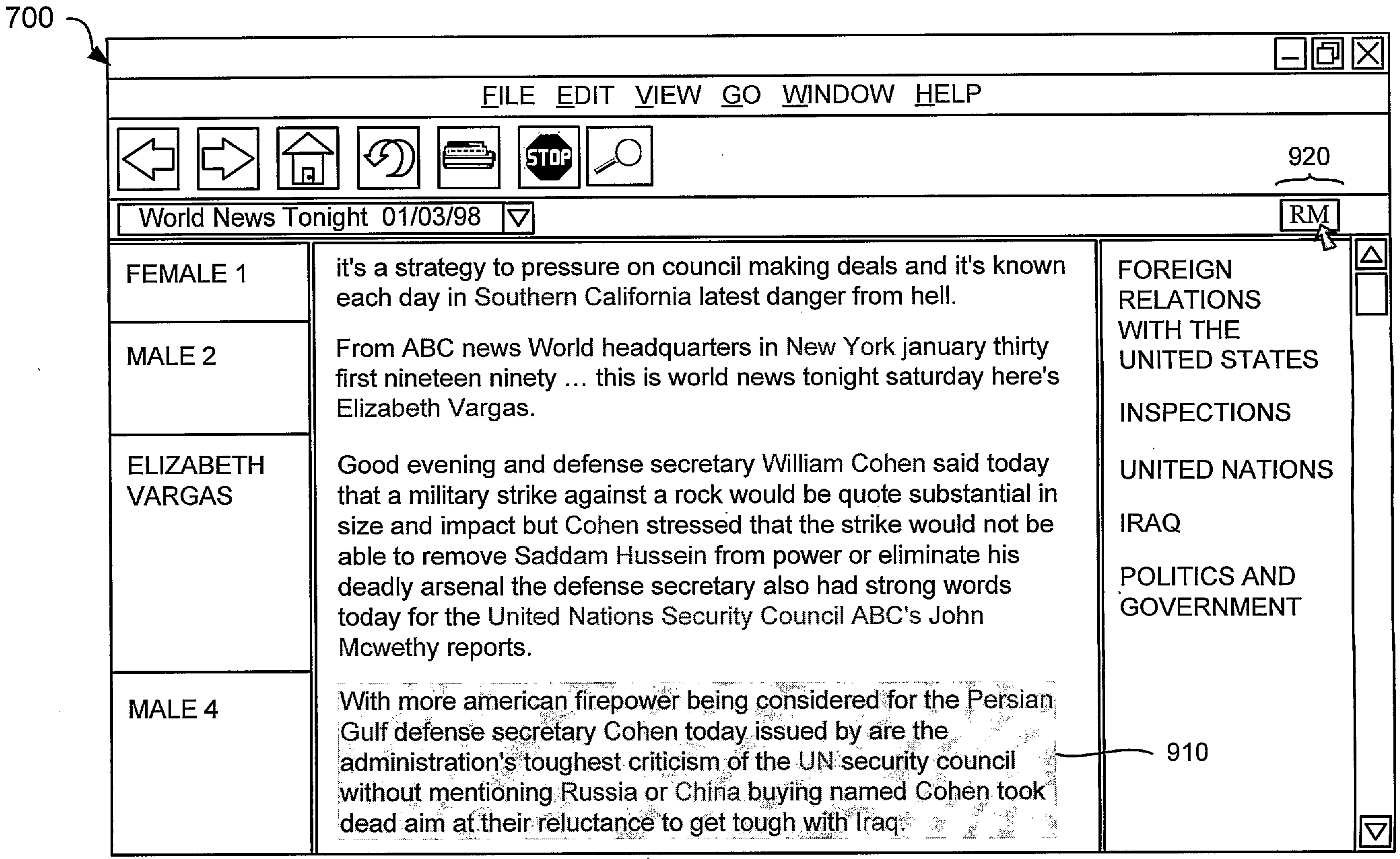
IRAQ

POLITICS AND GOVERNMENT

FIG. 7

FIG. 8


START
ISSUE REQUEST TO SERVER
805
RETRIEVE METADATA FROM METADATA DATABASE
810
CREATE HTML DOCUMENT FROM META DATA
815
SEND HTML DOCUMENT TO CLIENT
820
PRESENT HTML DOCUMENT TO USER
825
830
REQUEST FOR ORIGINAL MEDIA?
NO
YES
DETERMINE DESIRED PORTION
835
RETRIEVE DESIRED PORTION OF ORIGINAL MEDIA
840
PLAY BACK ORIGINAL MEDIA
845
VISUALLY SYNCHRONIZE PLAYBACK WITH HTML DOCUMENT PRESENTATION
850
END



FEMALE 1	it's a strategy to pressure on council making deals and it's known each day in Southern California latest danger from hell.
MALE 2	From ABC news World headquarters in New York january thirty first nineteen ninety ... this is world news tonight saturday here's Elizabeth Vargas.
ELIZABETH VARGAS	Good evening and defense secretary William Cohen said today that a military strike against a rock would be quote substantial in size and impact but Cohen stressed that the strike would not be able to remove Saddam Hussein from power or eliminate his deadly arsenal the defense secretary also had strong words today for the United Nations Security Council ABC's John Mcwethy reports.
MALE 4	With more american firepower being considered for the Persian Gulf defense secretary Cohen today issued by are the administration's toughest criticism of the UN security council without mentioning Russia or China buying named Cohen took dead aim at their reluctance to get tough with Iraq.

FOREIGN RELATIONS WITH THE UNITED STATES

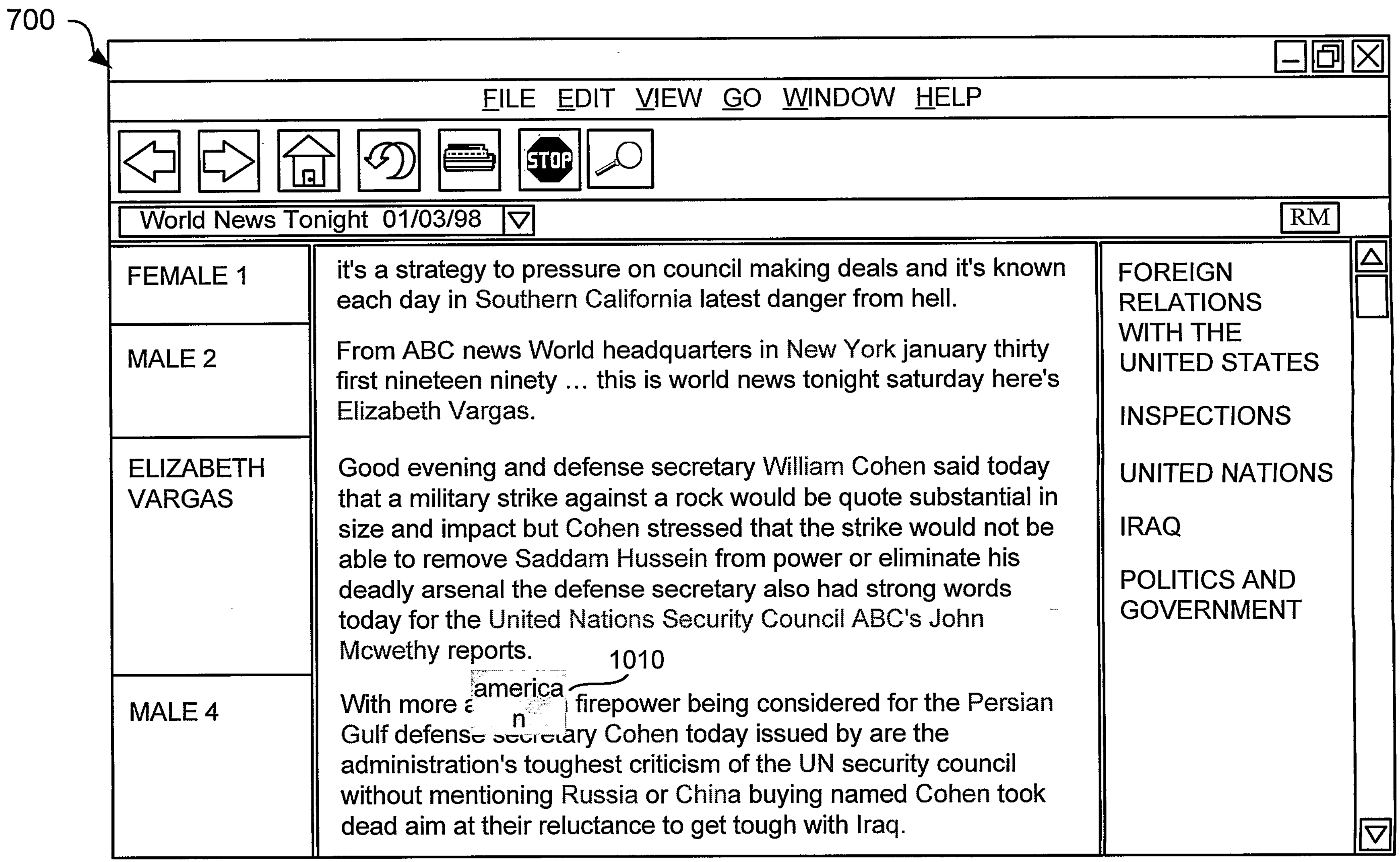
INSPECTIONS

UNITED NATIONS

IRAQ

POLITICS AND GOVERNMENT

FIG. 9



FEMALE 1	it's a strategy to pressure on council making deals and it's known each day in Southern California latest danger from hell.
MALE 2	From ABC news World headquarters in New York january thirty first nineteen ninety ... this is world news tonight saturday here's Elizabeth Vargas.
ELIZABETH VARGAS	Good evening and defense secretary William Cohen said today that a military strike against a rock would be quote substantial in size and impact but Cohen stressed that the strike would not be able to remove Saddam Hussein from power or eliminate his deadly arsenal the defense secretary also had strong words today for the United Nations Security Council ABC's John Mcwethy reports.
MALE 4	With more american firepower being considered for the Persian Gulf defense secretary Cohen today issued by are the administration's toughest criticism of the UN security council without mentioning Russia or China buying named Cohen took dead aim at their reluctance to get tough with Iraq.

FOREIGN RELATIONS WITH THE UNITED STATES

INSPECTIONS

UNITED NATIONS

IRAQ

POLITICS AND GOVERNMENT

FIG. 10