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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2008 (24.01.2008)

PCT

(10) International Publication Number
WO 2008/011142 A2

(51) International Patent Classification:
G10L 21/00 (2006.01) *G10L 15/26* (2006.01)

(21) International Application Number: PCT/US2007/016451

(22) International Filing Date: 19 July 2007 (19.07.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/490,798 20 July 2006 (20.07.2006) US

(71) Applicant *(for all designated States except US)*: **MSPOT, INC.** [US/US]; 459 Hamilton Avenue, Suite 307, Palo Alto, CA 94301 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **HO, Edwin** [US/US]; 905 Amarillo Avenue, Palo Alto, CA 94303 (US).

(74) Agent: **LOHSE, Timothy, W.**; DLA Piper US LLP, 2000 University Avenue, East Palo Alto, CA 94303 (US).

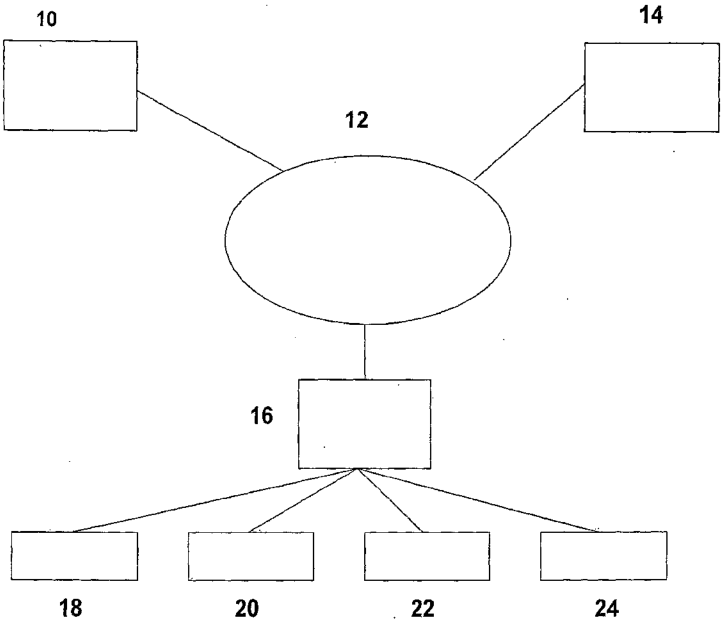
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, 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, MT, 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*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHOD AND APPARATUS FOR PROVIDING SEARCH CAPABILITY AND TARGETED ADVERTISING FOR AUDIO, IMAGE, AND VIDEO CONTENT OVER THE INTERNET



(57) **Abstract:** The present invention provides an apparatus and method for extracting the content of a video, image, and/or audio file or podcast, analyzing the content, and then providing a targeted advertisement, search capability and/or other functionality based on the content of the file or podcast.

METHOD AND APPARATUS FOR PROVIDING SEARCH CAPABILITY AND TARGETED ADVERTISING FOR AUDIO, IMAGE, AND VIDEO CONTENT OVER THE INTERNET

Field of the Invention

The present invention relates to extracting and analyzing the content of audio, image, and/or video data associated with Internet downloads, Internet podcasts or other digital distribution channels, producing descriptive data concerning that content, and then performing an action that utilizes the descriptive data, such as providing targeted advertising or a search capability.

Background of the Invention

The Internet is increasingly being populated with audio, image, and video content. Server storage capacity and user bandwidth continues to increase and many websites now contain a wealth of audio, image, and video content, including music, photographs, and movies. Audio and video content can be streamed over the Internet or downloaded by a user. Another popular means of obtaining audio and video content is through the use of "podcasts." This term was coined after Apple Computer, Inc. introduced its iPod™ product. The iPod™ devices, certain cellphone handsets, and other handheld devices are capable of connecting to a server over the Internet to receive podcasts. A podcast is an automatic downloading of audio and/or video content over the Internet, sometimes as part of a subscription to the content. For example, a user can subscribe to a television or radio program through a website and have the television or radio program downloaded automatically when the user connects his or her unit to the network.

In another field, text searching and intelligent search engines for the Internet are widespread. Many search engines (e.g., Google™) allow the user to enter keywords (e.g., "lawnmowers") and the search engine then searches for content based on those keywords by searching for the search terms themselves and/or or by searching for concepts that are related to the search terms. However, in the past, these searches were performed on databases created only from textual data available on the Internet. Audio, image, and video content were not included within these searches, unless they were associated with text that was created by a human being for that content (e.g., a textual title for a photograph).

In another field, advertising over the Internet is widespread. Advertising can be targeted to certain topics in which a user is likely to be interested. For instance, many websites that provide a search engine will send ads to the user based on the content of the search entered by the user. As an example, a user who searches for "lawnmowers" through a search engine might be provided with links to websites offering lawnmowers for sale.

To date, there has been no means for performing searches within audio, image, or video content or providing targeted advertising for audio, image, or video content.

It would be desirable to be able to provide a search capability, indexing capability, and other functionality using a database that includes content extracted from audio, image, and video data. This would have the practical effect of making the content of audio, image, and video content searchable and able to be indexed and categorized for future use.

It would be desirable to be able to provide ads automatically along with particular audio, image, or video content that is related to the subject matter of the content. For example, if a user subscribes to a television program discussing stocks and bonds through a podcast service, it would be useful to be able to automatically provide ads for stock brokers along with the podcast. It further would be desirable to be able to provide such ads at particular times within the podcasts such that the ads are relevant to the content of the podcast at a particular time, such as immediately after a certain word is spoken in the podcast.

Summary of the Invention

An apparatus and method for extracting the content of the audio, image, and video data, analyzing the content, and then providing a targeted advertisement, search capability, and/or other functionality based on the content is provided. One application of this invention is to provide targeted advertising that is provided in conjunction with audio, image, or video content over the Internet. Another application is to provide a search capability for audio, image, or video content.

One embodiment of the invention involves using a server to receive the audio content of a video or audio file or podcast. The server then performs a speech-to-text conversion on the audio and stores the extracted text in a database, in raw form and/or in various database fields. The server then receives a search inquiry from a network user. A search engine will run a search and will search within a database that includes the extracted audio data, and will provide a link to the audio data if relevant to the search entered. The server also may provide targeted advertising to the user based on the content that was extracted from the audio data, if relevant to the search entered.

Another embodiment of the invention involves using a server to download images or the video content of a video file or podcast. The server then performs image recognition to identify known images (e.g., a photograph of Abraham Lincoln), and stores those images and associated descriptive data in a database. The server then receives a search inquiry from a network user. A search engine will run a search and will search within a database that includes the podcast descriptive data and will provide a link to the image or video podcast if relevant to the search entered. The server also may provide targeted advertising to the user based on the descriptive data associated with the video podcast.

Brief Description of the Drawings

Fig. 1 is a diagram of the basic hardware system used in the preferred embodiment.

Fig. 2 is a flowchart of the basic method used in one embodiment for extracting text from and creating descriptive data for audio data.

Fig. 3 is a flowchart of the basic method used in another embodiment for creating descriptive data for image data or video data.

Detailed Description of the Preferred Embodiments

Embodiments implementing the present invention are described with reference to Figures 1-3. Figure 1 shows the basic components of the hardware of one embodiment. Typically, a user will operate a computing device **10** to connect to a network **12**, such as the Internet. Computing device **10** can be any device with a processor and memory, and includes PCs, laptops, mobile phones, PDAs, servers, etc. Computing device **10** preferably includes a display device and a media player. The connection to network **12** can be through any type of network connection, cellular network, mobile phone network, etc. The network **12** will connect a plurality of users and a plurality of servers and communicate data/content between the servers and the users. In one embodiment, Server A **14** will provide video, image, and/or audio content over the network **12**, such as through a podcast or download to other computing devices connected to the network. Server B **16** will be able to access that content through the network **12**. Server B **16** can include (or can be coupled to other devices containing) a database **18**, storage device **20**, search engine **22**, and advertising engine **24**. The database **18** typically comprises a database software program running on a server or other computer. The storage device **20** typically comprises magnetic or optical storage devices such as hard disk drives, RAID devices, DVD drives, or other storage devices. The storage device **20** typically stores the software run by the server and other associated computers as well as the underlying data and database structures for database **18**. The search engine **22** typically comprises a software program running on a server or other computer that is capable of identifying relevant data records in database **18** based on a search request entered on computing device **10** by a user. The advertising engine **24** typically comprises a software program running on a server or other computer that is capable of identifying advertising data that is relevant to the search request entered on computing device **10**. Database **18**, storage device **20**, search engine **22**, and advertising engine **24** are well-known in the art and may all be contained on a single server (such as Server B **16**) or on multiple servers.

Figure 2 illustrates an embodiment relating to audio content or to the audio portion of a file or podcast that includes both video and audio. The method illustrated in Figure 2 is preferably implemented on a server or other computing device. Server B **16** will first download the audio data offered by Server A **14** over the network **12** (step **30**). Server B

16 will then automatically process the data, including the step of performing speech-to-text conversion on that audio data and/or creating descriptive data. (step **32**). Speech-to-text conversion is well-known in the art. Creating descriptive data involves processing the text data to determine descriptive data that falls within certain predetermined database fields (e.g., a field indicating the general realm of the audio content, such as stock market information or movie news). Such processing essentially creates metadata that describes the content of the audio podcast. For instance, the database could include a field called "genre" that describes the general realm of the content. The entry that is placed into that field would be based on the content itself. As an example, if the extracted textual data includes the words "foreign policy" and "President," then an entry of "politics" could be placed in the genre field. That metadata would then be associated with that particular audio content. In this manner, audio content can be indexed (and later searched). The text, the descriptive data, and/or the audio data are imported into a database. (step **34**).

Referring still to Figure 2, a user will then input a search request (e.g., "lawnmowers") on computing device **10**, such as through an Internet search engine run by Server B **16**. That request will be received by Server B **16** over network **12** (step **36**). Server B **16** and/or search engine **22** will then execute the search within the database **18** that includes the extracted textual data and/or descriptive metadata that previously was generated for the audio data (step **38**). If the search implicates the extracted textual data or descriptive metadata, then server B **16** and/or advertising engine **24** optionally: (i) will identify a relevant advertisement based on the descriptive metadata, and that advertisement will be sent to computing device **10** for display (step **40**), and/or (ii) will provide the audio data (which it previously obtained from server A **14** and stored) or a link to the audio data stored on server A **14** to the user (step **42**). Server B **16** and/or advertising engine **24** optionally can format the advertisement to fit the display and graphics parameters of the display device of computing device **10** prior to transmitting the advertisement to computing device **10**.

Figure 3 illustrates an embodiment that relates to images, video content, or to the video portion of a file or podcast that includes both video and audio. The method illustrated in Figure 3 is preferably implemented on a server or other computing device. Server B **16** will first download the image data or video data offered by Server A **14** over the network **12** (step **50**). Server B **16** will then automatically process the image data or video data, including the step of performing image recognition on that image or video data. (step **52**). Image recognition involves comparing one or more frames of the video data to a set of previously stored, known images, such as images of famous politicians, pop icons, etc. Image recognition is well-known in the art. The step of image recognition will generate recognition data (e.g., the name of a famous politician that shows up in Frame X of the video data) (step **52**). Server B will then import the image data, video data and/or the recognition data into database **18**. The recognition data can be further processed and the resulting descriptive data and/or the recognition data itself stored in certain database fields (e.g., a field indicating the names of persons who appear in the video) (step **54**). Steps **52** and **54** essentially create metadata that describes the content of the image or video data. For instance, the database could include a field called "genre" that describes the general realm of the content. The entry that is placed into that field would be based on the recognition data. As an example, if the recognition data includes "Abraham Lincoln" (because the prior step of image recognition had created that data based on an image in the video data) then an entry of "politics" could be placed in the genre field. The underlying image or video content will then be associated with the recognition data ("Abraham Lincoln") generated as a result of the image recognition step as well as descriptive data ("politics") generated through processing the recognition data. In this manner, video content can be indexed.

Referring again to Figure 3, a user will then input a search request on computing device **10**. That request will be received by Server B **16** over network **12** (step **56**). Server B **16** and/or search engine **22** will then execute the search within the database **18** that includes the recognition data and/or descriptive data that previously was created for the video data (step **58**). If the search implicates the recognition data and/or descriptive data, then server B **16** and/or advertising engine **24** optionally: (i) will generate an advertisement based on the recognition data and/or descriptive data, and that advertisement will be sent to computing device **10** for display (step **60**), and/or (ii) will provide the image or video data (which it previously obtained from server A **14** and stored) or a link to the image or video data stored on server A **14** to the user (step **62**).

With both audio and video downloads and podcasts, the timing of the advertisements can be synchronized with the audio and video content after the text data, descriptive data and/or recognition data has been created as discussed above. For example, if it has been determined that a certain video podcast contains a news segment on lawnmowers, an advertisement on lawnmowers can be integrated into the podcast to appear at the very moment when the news segment on lawnmowers begins, or even when the word "lawnmower" is spoken. Thus, after the user downloads the podcast and watches the news segment, the advertisement will appear on his or her screen at precisely the right moment. This is yet another benefit of converting audio, image, and video content into a text form that can be indexed, searched, and analyzed.

While the foregoing has been with reference to particular embodiments of the invention, it will be appreciated by those skilled in the art that changes in these embodiments may be made without departing from the principles and spirit of the invention, the scope of which is defined by the appended claims.

What is claimed is:

1. A method of converting audio data into a searchable form, comprising the steps of:

receiving audio content;

performing speech-to-text conversion on said audio content;

importing the text created by said conversion into a database;

receiving a search request from a user over a network; and

executing the search request, wherein the search is performed in a database that includes said text or portions thereof.

2. The method of claim 1, wherein the step of receiving audio content comprises the step of downloading said content over a network.

3. The method of claim 2, wherein the method further comprises the step of providing said content or a link to said content to said user over said network.

4. The method of claim 1, wherein the step of receiving audio content comprises receiving said content over a network through a podcast.

5. The method of claim 4, wherein the method further comprises the step of providing said content or a link to said content to said user over said network.

6. A method of converting audio data into a searchable form, comprising the steps of:

receiving audio content;

performing speech-to-text conversion on said audio content;

processing the text created by said conversion to create descriptive data that describes the content of at least some of said text;

importing one or both of at least some of said text and at least some of said descriptive data into a database;

receiving a search request from a user over a network; and

executing the search request, wherein the search is performed in a database that includes one or both of at least some of said text and at least some of said descriptive data.

7. The method of claim 6 wherein the method further comprises the step of: generating advertisements based upon the search request.

8. The method of claim 7 wherein the method further comprises the step of: transmitting the advertisements to said user over said network.

9. A method of converting image or video data into a searchable form, comprising the steps of:

receiving image content or video content;

performing image recognition on said image or video content;

importing the result of the image recognition step into a database;

receiving a search request from a user over a network; and

executing the search request, wherein the search is performed in a database that includes said result of the image recognition step.

10. The method of claim 9, wherein the step of receiving image content or video content comprises the step of downloading said content over a network.

11. The method of claim 10, wherein the method further comprises the step of providing said content or a link to said content to said user over said network.

12. The method of claim 9, wherein the step of receiving image content or video content comprises receiving said content over a network through a podcast.

13. The method of claim 12, wherein the method further comprises the step of providing said content or a link to said content to said user over said network.

14. A method of converting image or video data into a searchable form, comprising the steps of:

receiving image content or video content;

performing image recognition on said image or video content;

processing the result of the image recognition step to create descriptive data that describes the content of said result;

importing said result and at least some of said descriptive data into a database;

receiving a search request from a user over a network; and

executing the search request, wherein the search is performed in a database that includes one or both of said result and at least some of said descriptive data.

15. The method of claim 14 wherein the method further comprises the step of: generating advertisements based upon the search request.

16. The method of claim 15 wherein the method further comprises the step of: transmitting the advertisements to said user over said network.

17. A system for converting audio data into a searchable form, comprising:

a first computing device for receiving audio content over a network and performing speech-to-text conversion on said audio content;

a database for storing the resulting text; and

a second computing device for receiving a search request from a user over the network wherein the second computing device is capable of executing the search in a database that includes said resulting text or portions thereof.

18. The system of claim 17 wherein the first computing device and the second computing device are the same device.

19. The system of claim 17 wherein the system further comprises: a third computing device for issuing the search request over said network.

20. A system for converting image or video data into a searchable form, comprising:

a first computing device for receiving image or video content over a network and performing image recognition on said image or video content;

a database for storing the result of the image recognition; and

a second computing device for receiving a search request from a user over the network wherein the second computing device is capable of executing the search in a database that includes said result.

21. The system of claim 20 wherein the first computing device and the second computing device are the same device.

22. The system of claim 20 wherein the system further comprises: a third computing device for issuing the search request over said network.

23. A computing system, comprising:

means for receiving audio content;

means for performing speech-to-text conversion on said audio content;

means for importing the text created by said conversion into a database;

means for receiving a search request from a user over a network; and

means for executing the search request, wherein the search is performed in a database that includes said text or portions thereof.

24. A computing system, comprising:

means for receiving image content or video content;

means for performing image recognition on said image or video content;

means for importing the result of the image recognition step into a database;

means for receiving a search request from a user over a network; and

means for executing the search request, wherein the search is performed in a database that includes said result of the image recognition step.

25. A computing system for executing a set of instructions, wherein the instructions comprise:

instructions for receiving audio content;

instructions for performing speech-to-text conversion on said audio content;

instructions for importing the text created by said conversion into a database;

instructions for receiving a search request from a user over a network; and

instructions for executing the search request, wherein the search is performed in a database that includes said text or portions thereof.

26. A computing system for executing a set of instructions, wherein the instructions comprise:

instructions for receiving image content or video content;

instructions for performing image recognition on said image or video content;

instructions for importing the result of the image recognition step into a database;

instructions for receiving a search request from a user over a network; and

instructions for executing the search request, wherein the search is performed in a database that includes said result of the image recognition step.

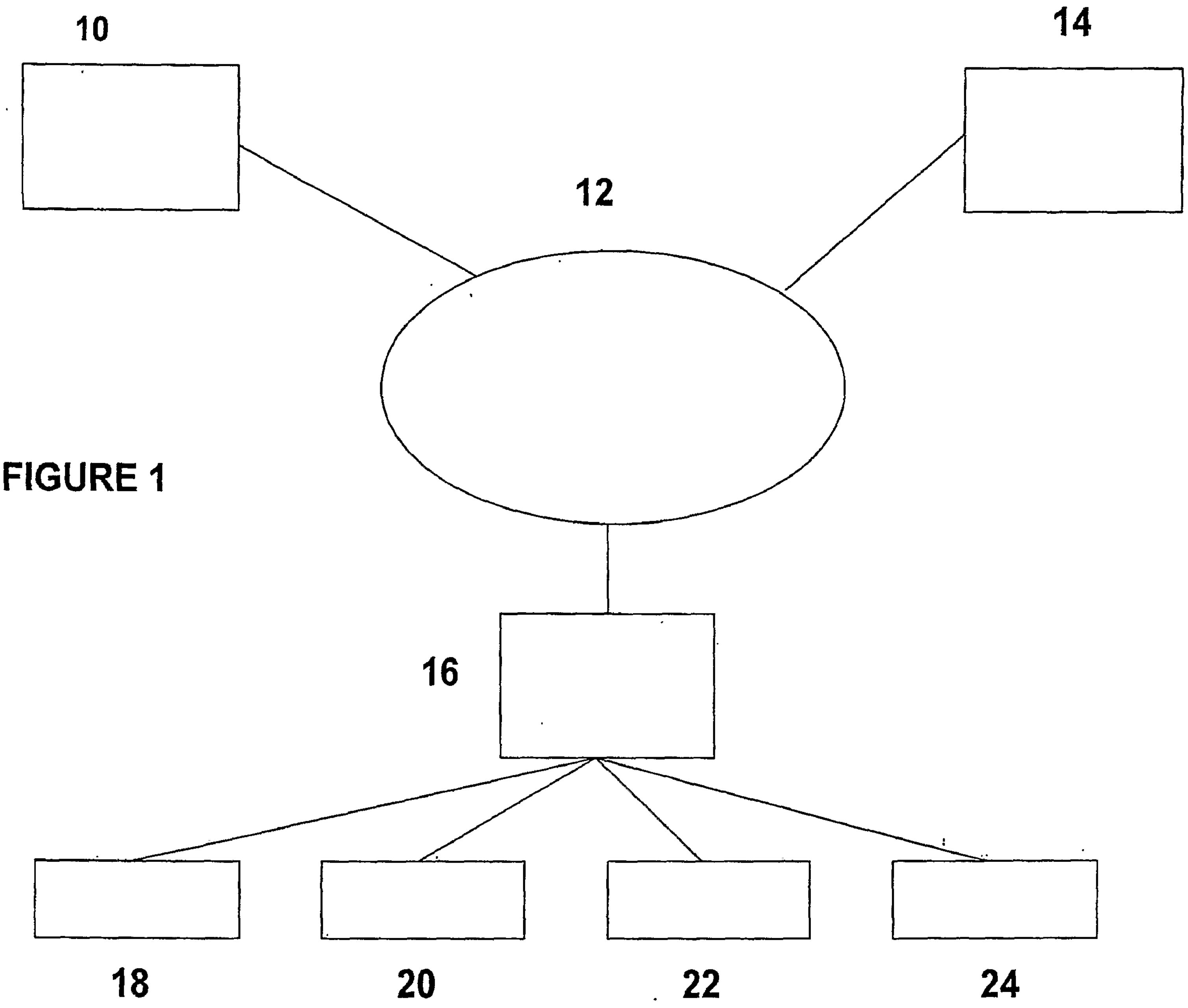


FIGURE 1

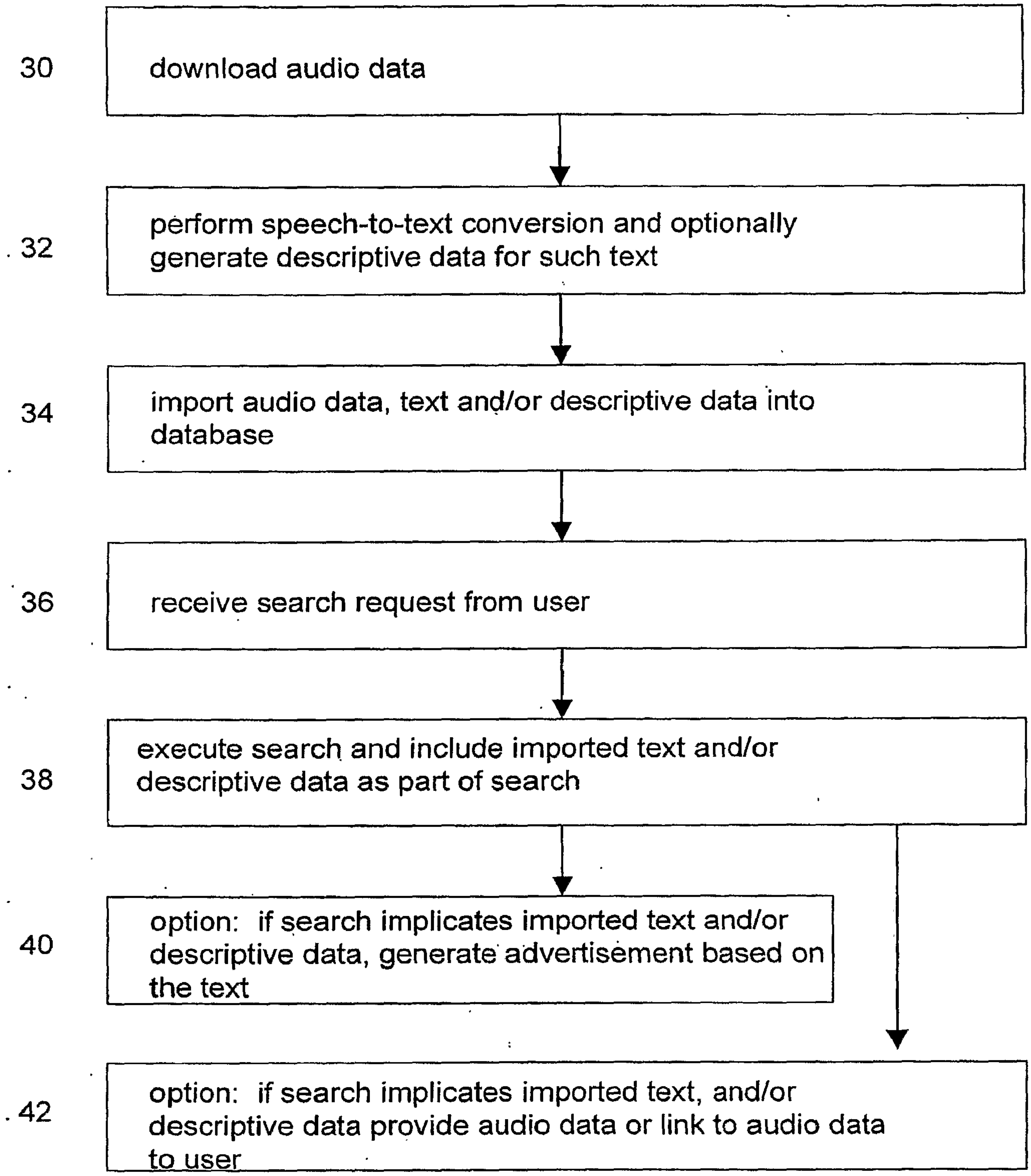

30
download audio data
32
perform speech-to-text conversion and optionally generate descriptive data for such text
34
import audio data, text and/or descriptive data into database
36
receive search request from user
38
execute search and include imported text and/or descriptive data as part of search
40
option: if search implicates imported text and/or descriptive data, generate advertisement based on the text
42
option: if search implicates imported text, and/or descriptive data provide audio data or link to audio data to user

FIGURE 2

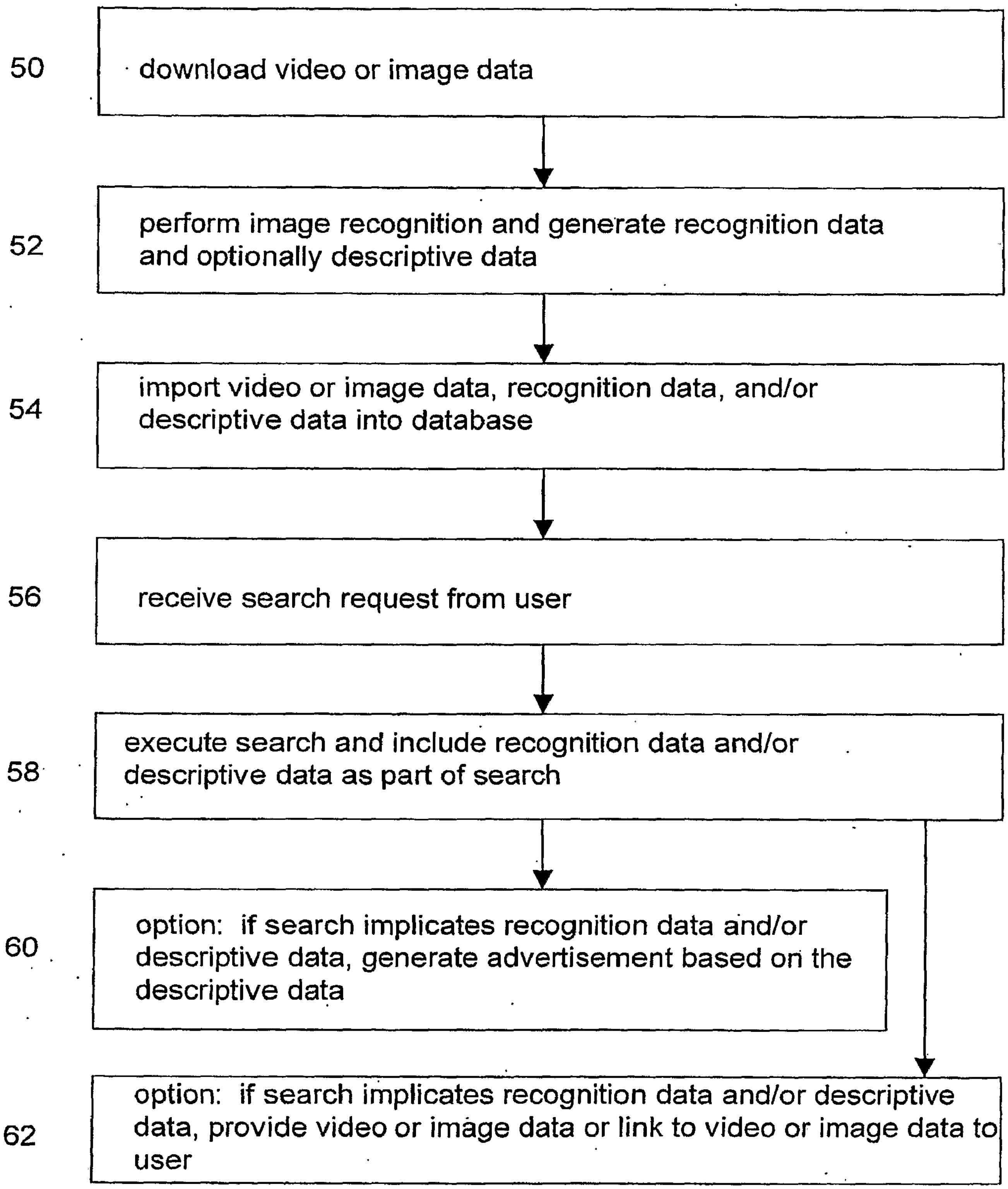

50
download video or image data
52
perform image recognition and generate recognition data and optionally descriptive data
54
import video or image data, recognition data, and/or descriptive data into database
56
receive search request from user
58
execute search and include recognition data and/or descriptive data as part of search
60
option: if search implicates recognition data and/or descriptive data, generate advertisement based on the descriptive data
62
option: if search implicates recognition data and/or descriptive data, provide video or image data or link to video or image data to user

FIGURE 3