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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 August 2004 (26.08.2004)

PCT

(10) International Publication Number
WO 2004/072780 A2

(51) International Patent Classification⁷: G06F

(21) International Application Number: PCT/US2004/003198

(22) International Filing Date: 5 January 2004 (05.01.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/444,982 5 February 2003 (05.02.2003) US

(71) Applicant *(for all designated States except US)*: **VERINT SYSTEMS, INC.** [US/US]; Worldwide Headquarters, 330 South Service Road, Melville, NY 11747 (US).

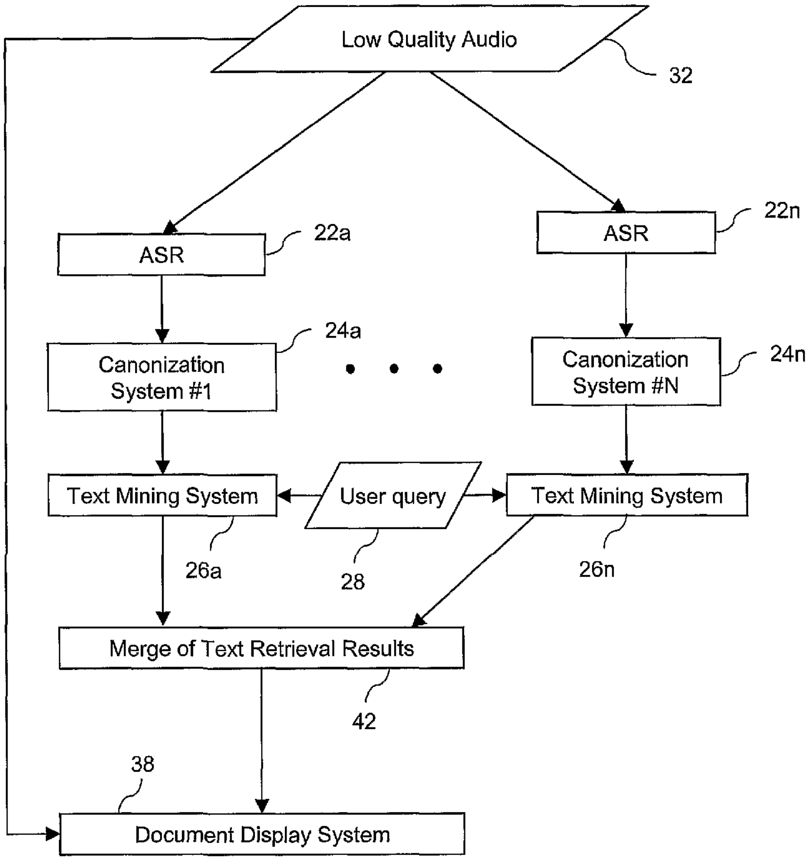
(72) Inventors: **ARIEL, Assaf**; Refidim 14, Tel-Aviv (IL). **BRAND, Michael**; 14/21 Hayim Barlev St., Newe Savyion, Or Yehuda 60408 (IL). **HOROWITZ, Itsik**; Heil-Hayam 5, Rishon-Le-Zion (IL). **SHOCHET, Ofer**; 50 Brodetsky St., Tel-Aviv (IL). **STAUBER, Itzik**; Bilu 33, Raanana (IL). **ZIV, Dror, Daniel**; 8 Blackberry Lane, Huntington, NY 11743 (US).

(74) Agent: **GOLLHOFER, Richard, A.**; Staas & Halsey LLP, 1201 New York Avenue, NW, Suite 700, Washington, DC 20005 (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,

[Continued on next page]

(54) Title: METHOD FOR AUTOMATIC AND SEMI-AUTOMATIC CLASSIFICATION AND CLUSTERING OF NON-DETERMINISTIC TEXTS



(57) Abstract: Non-deterministic text with average word recognition precision below 50 % is processed utilizing non-textual differences between words or sequences of words in the text to provide more useful information to users by resolving more than two decision options. One or more indexes that indicate non-textual differences between n-word sequences, where n is a positive integer, may be generated for use in data mining that considers the non-textual differences. Alternatively, multiple indexes may be generated using different data mining techniques that may or may not utilize non-textual differences and then the results produced by the different data mining techniques may be merged to identify non-textual differences. These techniques may be used in classifying, labeling, categorizing, filtering, clustering, or retrieving documents, or in discovering salient terms in a set of documents.

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, 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, 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.

METHOD FOR AUTOMATIC AND SEMI-AUTOMATIC CLASSIFICATION AND CLUSTERING OF NON-DETERMINISTIC TEXTS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application is related to and claims priority to U.S. provisional application entitled METHOD FOR AUTOMATIC AND SEMI-AUTOMATIC CLASSIFICATION AND CLUSTERING OF NON-DETERMINISTIC TEXTS having serial number 60/444,982, by Assaf ARIEL, Itsik HOROWITZ, Itzik STAUBER, Michael BRAND, Ofer SHOCHET and Dror ZIV, filed February 5, 2003 and incorporated by reference herein. This application is also related to the application entitled AUGMENTATION AND CALIBRATION OF OUTPUT FROM NON-DETERMINISTIC TEXT GENERATORS BY MODELING ITS CHARACTERISTICS IN SPECIFIC ENVIRONMENTS by Michael BRAND, filed concurrently and incorporated by reference herein.

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0002] The present invention is directed to processing of information in non-deterministic texts to increase the usefulness of the texts and, more particularly, to using non-textual information to indicate the importance or recognition accuracy of individual words or sequences of words.

2. Description of the Related Art

[0003] In general, spoken document retrieval (SDR) is composed of two stages: transcription of speech and information retrieval (IR). Transcription of the speech is often referred to as speech-to-text (STT) or automatic speech recognition (ASR), and is often performed using a large vocabulary continuous speech recognizer (LVCSR). Information retrieval (IR) is a general term referring to all forms of data mining. One common form of data mining, for example, is query-based retrieval, where, based on a user's query, documents are retrieved and presented to the user, ordered by an estimated measure of their relevance to the query. Traditionally, this stage is performed on the text output of the first stage.

[0004] There are many known techniques for extracting useful information from texts, commonly referred to as text mining or text data mining which is a sub-discipline of data mining.

Many of these techniques have been used on text output by speech-to-text algorithms or automatic character recognition systems. However, in systems that use text that has been converted from digitized speech or is based on character recognition, there has been little success when the original source is of low quality, such as telephone conversations or handwritten text, due to the low precision of accuracy of the resulting texts. As a result, most commentators in the field have discouraged application of techniques developed for easily recognized source material to source material that is difficult to recognize. Examples of such techniques can be found in U.S. Patents 5,625,748; 6,397,181 and 6,598,054, all incorporated by reference herein.

[0005] Therefore, there are no known systems that provide easy access to poor quality audio, except when it is in a predictable format, such as the rules that conversations between air traffic controllers and persons in the cockpit of an aircraft follow.

SUMMARY OF THE INVENTION

[0006] It is an aspect of the present invention to improve access to text by using non-textual information.

[0007] It is another aspect of the present invention to use conventional text mining techniques in previously developed text mining software in a way that utilizes non-textual information in data mining.

[0008] It is a further aspect of the present invention to improve access to documents produced by speech recognizers using recognition confidence measurement.

[0009] The above aspects can be attained by a method for processing documents derived from at least one of spontaneous and conversational expression and containing non-deterministic text with average word recognition precision below 50 percent, the processing utilizing non-textual differences between n-word sequences in the documents to resolve more than two decision options, where n is a positive integer. Such text may be obtained by automatic character recognition or automatic speech recognition of audio signals received via a telephone system. In the preferred embodiment, the non-textual differences between the n-word sequences relate to recognition confidence of the n-word sequences

[0010] When the processing requires fast access to the information stored in a large corpus of documents, e.g. for the purpose of data mining, the data is preferrably pre-processed to index the n-word sequences in a method that utilizes the non-textual differences between them. Such a procedure can speed up many forms of data access, and in particular many forms of data mining, including query based retrieval, as would be apparent to a person skilled in the art.

[0011] The data mining may include extracting parameters from the documents utilizing the non-textual differences between the n-word sequences and establishing relations between the parameters extracted from the documents. The parameters extracted from the documents may be fully known, such as parameters available in document metadata or may be hidden variables that cannot be fully determined from information existing in the document. Examples of extracted parameters include an assessment of relevance to a query based on the non-textual differences between the n-word sequences and an assessment of the document's relevance to a category.

[0012] As an alternative to creating index(es) indicating non-textual differences between n-word sequences, algorithm(s) can be used to convert text containing non-textual differences between the n-word sequences into different standard text documents. Many different algorithms may be used to transform non-deterministic text into standard text documents usable in text mining. For example, the algorithm to extract standard text documents from text with non-textual differences may apply a thresholding algorithm with varying thresholds. Then, one or more data mining techniques, each of which does not utilize non-textual differences, can be applied to these standard text documents and the outputs of the different data mining techniques can be merged to obtain information that is equivalent to that obtained by data mining that utilizes the non-textual differences.

[0013] Whether or not the index(es) include an indication of non-textual differences, the documents may be categorized, clustered, classified, filtered or labeled, e.g., by using an algorithm to detect salient terms in the documents based on non-linguistic differences between the n-word sequences.

[0014] In response to a query using any type(s) of index(es), information related to at least one of the documents may be displayed, including at least some non-textual differences between n-word sequences. Portions of the document(s) may be selectively displayed based on confidence of the accuracy of the displayed words. For example, salient terms in the document(s) may be displayed based on processing of confidence levels of recognition of the salient terms that resolves more than two decision options. In addition, parameters extracted from the documents and indications of the relations between these parameters may be displayed graphically.

[0015] In response to the display of such information, a user may indicate errors in recognition. In this case at least one word in the document is preferably replaced with a corrected word supplied by the user and the confidence level(s) of the corrected word(s) are reset to indicate high recognition accuracy.

[0016] These, together with other aspects and advantages which will be subsequently apparent, reside in the details of construction and operation as more fully hereinafter described and claimed, reference being had to the accompanying drawings forming a part hereof, wherein like numerals refer to like parts throughout.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a flowchart of a conventional spoken document retrieval system.

Figure 2 is a flowchart of one method of spoken document retrieval according to the invention.

Figure 3 is a flowchart of another method of spoken document retrieval according to the invention.

Figures 4 and 5 are block diagrams of spoken document retrieval systems according to the invention.

Figure 6 is a block diagram of one confidence sensitive inverted index and one regular index containing confidence information.

Figure 7 is a flowchart of text processing according to the invention.

Figures 8 and 9 are examples of displays generated by telephone call processing applications according to the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0017] Following are several terms used herein that are in common use in automatic speech recognition or data mining.

labeling	a form of data processing where documents are analyzed, and the analysis results (referred to as "labels") are made available for later processing stages. For example, a topical analysis of documents is a labeling of the documents by subject.

retrieving a form of data mining where a subset of a document corpus is returned in response to a query. Preferably, the documents are each given a rank pertaining to their relevance to the query, and are sorted by decreasing relevance.

categorizing a form of data mining where several "categories" are defined, and the documents of a corpus are labeled according the category to which they fit best. A common variation is multilabel categorizing, where each document may fit zero or more categories. Preferably, information is given regarding the quality of the fit.

clustering a form of data mining similar to categorization, with the difference that the "categories" are not predefined, and the data mining must reveal them automatically.

classifying a process performed on a stream of incoming documents, where each is labeled and then forwarded for relevant additional processing (manual or automatic) based on the labels that have been discovered.

filtering a process performed on a stream of incoming documents, where each is labeled and then forwarded or discarded based on the labels that have been discovered.

salient terms terms whose appearance in a document provides information relevant to its correct labeling, and consequently to all forms of data mining subsequent to labeling.

[0018] First, processing performed by a typical spoken document retrieval system will be described with reference to Fig. 1. High quality audio 20 is input into an ASR system 22 using LVCSR. ASR system 22 converts spoken words into textually represented words, but often has other outputs as well. These outputs may include timing information, an indication of the confidence of recognition of particular words and phrases, alternative likely transcriptions, and more.

[0019] The LVCSR output cannot be piped directly into a traditional text mining system. It has to be converted into searchable text. For this reason, canonization 24 is performed to produce canonized text, also referred to below as standard text documents, used by conventional text mining software. Most commonly, canonization simply involves taking the textual words out of the LVCSR output and concatenating them. More sophisticated canonization schemes involve usage of both textual and non-textual information to convert the LVCSR output into a format more easily handled by text mining system 26. Usage of textual information may include capitalization and punctuation based on grammatical rules. Usage of non-textual information may include capitalization and punctuation using timing information and word(s) omitted based on low confidence levels.

[0020] Text mining system 26 receives input from many different audio segments and stores the information in some format that will be convenient for later processing, a process known as "indexing". When asked to produce output, typically, though not exclusively, by user query 28, text mining system 26 searches its index and produces output. For example the output may be the identities of the audio segments that were requested for retrieval, ranked and scored by some relevance metric. The output may also include other information, such as the phrases in the retrieved segments that have proved to be salient terms, because of which the document was given the score that it was given.

[0021] All this information is finally piped into document display system 30 which can use all of it, and add to it the original audio segment(s) 20, to give the user feedback to user query 28 that is as informative and audio-visually appealing as possible.

[0022] A simplified embodiment of a method according to the present invention for spoken document retrieval of low grade audio 32 is illustrated in Fig. 2. A similar system could be used for text generated by a handwritten text recognition system. In embodiment illustrated in Fig. 2, the traditional text mining system has been replaced with data mining system 34 that is designed especially for speech in low grade audio 32. Data mining system 34 doesn't require that the output of ASR 36 be canonized into text before it is handled, and can therefore utilize all information available in the output of ASR 36. No information is lost in a canonization process, and words that are indexed can receive different and appropriate handling based on non-textual information, such as their confidence scores. The canonization stage 24 of the process illustrated in Fig. 1 is therefore entirely omitted, and the output of ASR 36 is available with more of its information in subsequent processing stages, including indexing, retrieval and display. Since non-textual information is available, document display system 38 in Fig. 2 displays different information than document display system 30 in Fig. 1.

[0023] Traditionally, speech data mining was confined to high quality audio 20 such as broadcast quality audio. Broadcast audio is of high quality, typically achieving 60-80% word figurations illustrated in Figs. 4 and 5 and other configurations are possible and will be apparent to a person of ordinary skill in the art. For example, the functions performed by separate servers in Figs. 4 and 5 may be performed by separate modules in a single computing system.

[0029] In the configuration illustrated in Fig. 4, a system according to the invention is used interactively offline. Voice data are supplied from voice acquisition module(s) 50 by network 52 and stored in data storage 54. Network 52 may be any known type of network, such as a local area network (LAN), wide area network (WAN), the Internet, etc. The voice data may be in the form of WAV files or any other audio file format.

[0030] In either of the configurations illustrated in Figs. 4 and 5, the system is accessed by one or more user terminals 56, such as personal computers or other devices that include a user interface which may include a display. In the interactive offline system illustrated in Fig. 4, users log into the system at various times to submit queries to voice oriented information retrieval (VOIR) indexing server 58. Voice data from voice acquisition module(s) 50 are supplied to speech categorization server 60 which, if necessary, converts the data before supplying the voice data to LVCSR(s) 22 and performs load balancing when more than one LVCSR 22 is used.

[0031] LVCSR(s) 22 output words and additional data, such as speaker-change, timing information, confidence scores, etc. In addition, call metadata, such as the time that a call was made and the number dialed, is obtained from voice acquisition module(s) 50 together with the voice data. All these types of data are combined, e.g., by speech categorization server 60 and forwarded, in online mode to speech analysis server 62 and in offline mode to VOIR indexing server 58. Regardless of whether the method illustrated in Fig. 2 or Fig. 3 is implemented, results of a query in offline mode can be displayed on user terminals 56 with at least some of the non-textual differences between n-word sequences indicated. Examples of how the non-textual differences are conveyed to the user will be described below with reference to Figs. 7 and 8.

[0032] The online configuration illustrated in Fig. 5, may be used when the volume of voice data is too large to allow effective offline processing, or it is desired to use push-technology alerts to people who may want the data. For example, a police inspector may want to be paged when the system detects a phone conversation relevant to her case. In the online configuration illustrated in Fig. 5, the output of LVCSR(s) 22 is supplied via network 52 to speech analysis server 62 which labels the voice data. For example, the voice data may be labeled according to importance, subject matter, person or group that needs to respond, etc. The labeling of the transcribed voice data is combined with the output of LVCSRs 22 and call metadata, and forwarded to categorization queue and workflow manager 64. The users at user terminals 56 are provided this information by categorization queue and workflow manager 64. Using the labeling provided by speech analysis server 62, categorization queue and workflow manager 64 supplies text, voice data and call metadata appropriate for that user, depending on importance, topic, identity of the user, etc.

[0033] Training of speech analysis server 62 may be accomplished by offline processing using VOIR indexing server 58 in an implementation that includes both servers 58 and 62. One or more users label calls by importance, subject matter, relevant person or group, etc. The labels assigned by users can be provided to speech analysis server 62 as training data to recognize similar calls during online processing of calls in a call center, for example. In addition, training may continue during online processing as users correct the labeling provided by speech analysis server 62. When all processing is offline, VOIR indexing server 58 is trained in a similar manner.

[0034] In a typical implementation of the invention, a single LVCSR 22 pass is sufficient for each call. If the method described above with reference to Fig. 2 is implemented, LVCSR 22 supplies metadata, including confidence scores, associated with recognized words to VOIR indexing server 58 which generates an index that indicates at least some of the non-textual differences between n-word sequences. If the method described above with reference to Fig. 3 is implemented, VOIR indexing server 58 maintains an index for each canonization system 24. In either case, the index(es) and the voice data (preferably compressed to minimize space requirements) or other data from which indexed text is obtained (such as handwritten documents) are preferably stored in data storage 54.

[0035] In the preferred embodiment, if the method illustrated in Fig. 2 is implemented, data in data storage 54 is indexed by use of at least one confidence sensitive inverted index. A confidence sensitive inverted index maps from terms to a sorted linked list identifying all documents where each term occurs and from each appearance of a document in this list to a sorted linked list identifying all positions in which the term appears and the confidence level of its recognition. In addition (or alternatively), indexed data may include aggregated information relating to confidence.

[0036] An example is illustrated in Fig. 6 of a confidence sensitive inverted index 65 and a regular (forward) index 66 containing confidence information with the two indexes 65, 66 referencing each other. In the mapping 67 from terms in the documents to a sorted linked list of documents 68, each appearance 69 of a term in the document can carry additional data, such as its position in the document, its timing information, recognition score of that appearance of the term, etc. Also, an expected number of real occurrences of the term (e.g., term i which points to mapping 67i) in the indexed document (e.g., 68a) can be calculated based on the individual recognition scores of the occurrences.

[0037] Another example of aggregated information relating to confidence information that can be saved is the strength of association between every document and each category. This information can be saved either in a regular (forward) index, like index 66, another inverted index (not shown), or both. Information not relating to confidence, such as call metadata, can also be indexed, either in another inverted index, in a forward index like index 66, or both. In both cases, if an inverted index is used, confidence sensitive inverted index 65 or a separate index can be used. Furthermore, additional mapping technologies, in addition to or instead of a mapping into a sorted linked list, can also be used. Data storage 54 can also store information other than the indexes, such as the data that is being indexed. This data may include, among others, call audio, voice data and call metadata, and may include additional indexes used to refer to the same data.

[0038] A more detailed flow of processing through the system illustrated in Figs. 4 and 5 is provided in Fig. 8. Online processing flow corresponding to the configuration illustrated in Fig. 5 is illustrated in Fig. 8 by solid lines, while offline processing flow corresponding to the configuration illustrated in Fig. 4 is illustrated by dash-dot lines. Low quality source data 32, such as recorded telephone conversations, supplied by voice acquisition module(s) 50, undergo text extraction 74 in LVCSR(s) 22 controlled by speech categorization server 60. In offline mode, the results 76, which may include text, confidence scores, timing information and text alternative lattice information (potentially, other information, as well), undergo indexing 78 in VOIR indexing server 58 and are stored in data storage 54. In the online mode, results 76 are supplied to speech analysis server 62 which may perform labeling 80 of the calls, as described above. Data from VOIR indexing server 58 are used for category training 82, so that the categorization 84 can later be used in either online or offline mode.

[0039] One embodiment of the system illustrated in Fig. 4 is used to process recorded telephone conversations at a call center by automatically generating transcriptions of the conversations. In this embodiment, offline ad-hoc querying 86 (Fig. 8) utilizes categorization 84 or rule-based keyword spotting 88 to obtain information 90 related to at least one of the documents, including at least some of the non-textual differences between n-word sequences that may be displayed on user terminal(s) 56 in the format illustrated in Fig. 8 or 9. The display illustrated in Fig. 8 provides an example of user input keywords 102 "call OR meeting" that have been found in 170 documents, eight of which are displayed on screen in Fig. 8. Preferably, the documents may be listed in a table 104 in an order based in part on the confidence of accuracy of the keywords displayed in the list. In the example illustrated in Figs. 8 and 9, table 104 includes call metadata, such as start time.

[0040] In the example illustrated in Fig. 8, a waveform 106 of a portion of the seventh document (indicated as selected by shading in table 104) is displayed in the lower portion of the screen with indications of when the keywords were detected. Below the waveform is the text 108 recognized by LVCSR(s) 22. Preferably, text 108 indicates the recognition confidence of the words and the salient terms listed in the query using one or more of highlighting, underlining, color or shade, size and style of fonts. Also shown in the example illustrated in Figs. 8 and 9 are labels of the conversation, such as "Technical" and "Incomplete" which follow the "Categories" 116 and appear in the column under "Contact Related To" in table 104, along with similar category information. Confidence of these labels is also indicated.

[0041] In one embodiment of the invention, a user may listen to the entire recording by using a pointing device, such as a computer mouse, to select a row in table 104 corresponding to the recording or can hear just the segments of audio corresponding to transcribed salient terms by selecting the speaker icon under the word "Play" on the row. Once a row has been selected, a user may select one of the words, such as "call" in a user-selectable speech bubble 110 associated with the waveform, or in the adjoining text, to skip directly to the point in a conversation where the word was said. A pointer 112 below the waveform 106 indicates what sound is being played back to the user and a vertical cursor 114 indicates what word was recognized for the associated sound.

[0042] Preferably, user terminal(s) 56 can also be used to graphically display results, e.g., content information 90, indicating parameters extracted from the documents. Examples illustrated in Fig. 9 are bar graphs 118, 120. In Fig 9, left bar graph 118 shows the number of calls matching a query based on call date, while right graph 120 shows the relations of several categories to the user query.

[0043] The present invention has been described with respect to embodiments using text documents generated from telephone calls. However, as noted above, the invention is not limited to texts generated in this manner and can also be applied to text obtained in other ways, such as from fact extraction systems. Furthermore, the present invention can be used with any system for processing documents that derive from at least one of spontaneous and conversational expression which outputs non-deterministic text with average word recognition below 50 percent.

[0044] The many features and advantages of the invention are apparent from the detailed specification and, thus, it is intended by the appended claims to cover all such features and advantages of the invention that fall within the true spirit and scope of the invention. Further, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and operation illustrated and described, and accordingly all suitable modifications and equivalents may be resorted to, falling within the scope of the invention.

CLAIMS

What is claimed is:

1. A document processing method, comprising:

processing documents derived from at least one of spontaneous and conversational expression and containing non-deterministic text with average word recognition precision below 50 percent, said processing utilizing non-textual differences between n-word sequences in the documents to resolve more than two decision options, where n is a positive integer.

2. A method as recited in claim 1, wherein said processing includes data mining of the documents.

3. A method as recited in claim 2, wherein said data mining includes retrieving at least one of the documents utilizing the non-textual differences between the n-word sequences in the documents.

4. A method as recited in claim 2, wherein said data mining includes extracting parameters from the documents, utilizing the non-textual differences between said n-word sequences.

5. A method as recited in claim 4, wherein said data mining further includes producing graphic results indicating the relations between the parameters extracted from the documents.

6. A method as recited in claim 4, wherein at least one of the parameters extracted from the documents is an assessment of relevance to a query based on the non-textual differences between the n-word sequences.

7. A method as recited in claim 4, wherein at least one of the extracted parameters is an assessment of a hidden variable that cannot be fully determined from information existing in the document.

8. A method as recited in claim 4, wherein at least one of the extracted parameters is the assessment of the document's relevance to a category.

9. A method as recited in claim 2, wherein said processing includes categorizing the documents.

10. A method as recited in claim 9, wherein said categorizing includes use of at least one algorithm to detect salient terms in the documents based on non-linguistic differences between the n-word sequences.

11. A method as recited in claim 2, further comprising clustering the documents.

12. A method as recited in claim 11, wherein said clustering includes discovering salient terms in the documents based on non-linguistic differences between the n-word sequences.

13. A method as recited in claim 11, wherein said clustering includes assessing a relation between the n-word sequences based on non-textual differences.

14. A method as recited in claim 4, wherein said data mining includes establishing relations between the parameters extracted from the documents.

15. A method as recited in claim 1, wherein the non-textual differences between the n-word sequences relate to recognition confidence of the n-word sequences.

16. A method as recited in claim 1, further comprising at least one of classifying and filtering the documents as the documents are received.

17. A method as recited in claim 1, further comprising labeling the documents as the documents are received.

18. A method as recited in claim 1, further comprising displaying information related to at least one of the documents, including at least some of the non-textual differences between the n-word sequences.

19. A method as recited in claim 18, wherein said displaying uses at least one of gray scaling, color, font-size and font style to indicate at least some of the non-textual differences between the n-word sequences.

20. A method as recited in claim 18, wherein said displaying selectively displays portions of the at least one of the documents based on confidence of accuracy of words displayed.

21. A method as recited in claim 18, wherein said displaying further displays salient terms in the at least one of the documents based on said processing of confidence levels of the salient terms that resolves more than two decision options.

22. A method as recited in claim 21, wherein a number of the salient terms are available for display and said displaying is further based on the number of the salient terms available for display and available space for display of the salient terms.

23. A method as recited in claim 1, further comprising:

receiving user input indicating errors in recognition; and

replacing at least one word in the document with a corrected word based on the user input and setting the confidence levels of the corrected word to indicate high recognition accuracy.

24. A method as recited in claim 1, further comprising generating the documents by automatic speech recognition of audio signals received via a telephone system.

25. A method as recited in claim 1, further comprising generating the documents by automatic character recognition.

26. A method as recited in claim 1, further comprising generating the documents by a fact extraction system.

27. A method as recited in claim 1, wherein said processing includes

applying different data mining techniques, each of which does not indicate non-textual differences; and

merging results of the different data mining techniques to obtain results that are dependent on the non-textual differences between the n-word sequences.

28. A method as recited in claim 27, wherein the different data mining techniques include at least one of retrieving, categorizing, filtering, classifying, labeling and clustering documents without utilization of any non-textual differences between the n-word sequences.

29. A method as recited in claim 27,

wherein said applying uses a plurality of different algorithms to transform non-deterministic text into standard text documents usable in text mining, and

wherein the data mining techniques operate on the standard text documents.

30. A method as recited in claim 29,

wherein said processing further includes generating a plurality of indexes of the standard text documents, and

wherein the data mining techniques operate on the indexes to obtain the results.

31. A method as recited in claim 30, wherein the data mining techniques include

receiving a query; and

retrieving the results relevant to the query:

32. A method as recited in claim 30, wherein the data mining of at least some of the different indexes is performed by data mining software that does not output non-textual differences.

33. A method as recited in claim 29, wherein the different algorithms are thresholding algorithms using different confidence thresholds to determine omitted words that fall below the confidence thresholds.

34. A method as recited in claim 1, further comprising:

receiving user input indicating a change in labeling of at least one document; and

replacing at least part of information provided by at least one label for the at least one document based on the user input.

35. A document processing method, comprising:

producing at least one index of n-word sequences in documents derived from at least one of spontaneous and conversational expression and containing non-deterministic text with average word recognition precision below 50 percent, utilizing non-textual differences between the n-word sequences, where n is a positive integer; and

processing the documents based on the non-textual differences between the n-word sequences in the at least one index, where said processing resolves more than two decision options.

36. A method as recited in claim 35, wherein the non-textual differences between the n-word sequences relate to recognition confidence of the n-word sequences.

37. At least one computer readable medium storing instructions for controlling at least one computer system to perform a document processing method comprising:

processing documents derived from at least one of spontaneous and conversational expression and containing non-deterministic text with average word recognition precision below 50 percent, said processing utilizing non-textual differences between n-word sequences in the documents, where n is a positive integer and said processing resolves more than two decision options.

38. At least one computer readable medium as recited in claim 37, wherein said processing includes data mining of the documents.

39. At least one computer readable medium as recited in claim 38, wherein said data mining includes retrieving at least one of the documents utilizing the non-textual differences between the n-word sequences in the documents.

40. At least one computer readable medium as recited in claim 38, wherein said data mining includes

extracting parameters from the documents, utilizing the non-textual differences between said n-word sequences; and

establishing relations between the parameters extracted from the documents.

41. At least one computer readable medium as recited in claim 40, wherein said data mining further includes producing graphic results indicating the relations between the parameters extracted from the documents.

42. At least one computer readable medium as recited in claim 40, wherein at least one of the parameters extracted from the documents is an assessment of relevance to a query based on the non-textual differences between the n-word sequences.

43. At least one computer readable medium as recited in claim 40, wherein at least one of the extracted parameters is an assessment of a hidden variable that cannot be fully determined from information existing in the document.

44. At least one computer readable medium as recited in claim 40, wherein at least one of the extracted parameters is the assessment of the document's relevance to a category.

45. At least one computer readable medium as recited in claim 38, wherein said processing includes categorizing the documents.

46. At least one computer readable medium as recited in claim 45, wherein said categorizing includes use of at least one algorithm to detect salient terms in the documents based on non-linguistic differences between the n-word sequences.

47. At least one computer readable medium as recited in claim 38, further comprising clustering the documents.

48. At least one computer readable medium as recited in claim 47, wherein said clustering includes discovering salient terms in the documents based on non-linguistic differences between the n-word sequences.

49. At least one computer readable medium as recited in claim 47, wherein said clustering includes assessing a relation between the n-word sequences based on non-textual differences.

50. At least one computer readable medium as recited in claim 37, wherein the non-textual differences between the n-word sequences relate to recognition confidence of the n-word sequences.

51. At least one computer readable medium as recited in claim 37, further comprising at least one of classifying and filtering the documents as the documents are received.

52. At least one computer readable medium as recited in claim 37, further comprising labeling the documents as the documents are received.

53. At least one computer readable medium as recited in claim 37, further comprising displaying information related to at least one of the documents, including at least some of the non-textual differences between the n-word sequences.

54. At least one computer readable medium as recited in claim 53, wherein said displaying uses at least one of gray scaling, color, font-size and font style to indicate at least some of the non-textual differences between the n-word sequences.

55. At least one computer readable medium as recited in claim 53, wherein said displaying selectively displays portions of the at least one of the documents based on confidence of accuracy of words displayed.

56. At least one computer readable medium as recited in claim 53, wherein said displaying further displays salient terms in the at least one of the documents based on said

processing of confidence levels of the salient terms that resolves more than two decision options.

57. At least one computer readable medium as recited in claim 56, wherein a number of the salient terms are available for display and said displaying is further based on the number of the salient terms available for display and available space for display of the salient terms.

58. At least one computer readable medium as recited in claim 37, further comprising:

receiving user input indicating errors in recognition; and

replacing at least one word in the document with a corrected word based on the user input and setting the confidence levels of the corrected word to indicate high recognition accuracy.

59. At least one computer readable medium as recited in claim 37, further comprising generating the documents by automatic speech recognition of audio signals received via a telephone system.

60. At least one computer readable medium as recited in claim 37, further comprising generating the documents by automatic character recognition.

61. At least one computer readable medium as recited in claim 37, further comprising generating the documents by a fact extraction system.

62. At least one computer readable medium as recited in claim 37, wherein said processing includes

applying different data mining techniques, each of which does not indicate non-textual differences; and

merging results of the different data mining techniques to obtain the non-textual differences between the n-word sequences.

63. At least one computer readable medium as recited in claim 62, wherein the different data mining techniques include at least one of retrieving, categorizing, filtering, classifying,

labeling and clustering documents without utilization of any non-textual differences between the n-word sequences.

64. At least one computer readable medium as recited in claim 62,

wherein said applying uses a plurality of different algorithms to transform non-deterministic text into standard text documents usable in text mining, and

wherein the data mining techniques operate on the standard text documents.

65. At least one computer readable medium as recited in claim 64,

wherein said processing further includes generating a plurality of indexes of the standard text documents, and

wherein the data mining techniques operate on the indexes to obtain the results.

66. At least one computer readable medium as recited in claim 65, wherein the data mining techniques include

receiving a query; and

retrieving the results relevant to the query.

67. At least one computer readable medium as recited in claim 66, wherein the data mining of at least some of the different indexes is performed by data mining software that does not output non-textual differences.

68. At least one computer readable medium as recited in claim 64, wherein the different algorithms are thresholding algorithms using different confidence thresholds to determine omitted words that fall below the confidence thresholds.

69. At least one computer readable medium as recited in claim 37, further comprising:

receiving user input indicating a change in labeling of at least one document; and

replacing at least part of information provided by at least one label for the at least one document based on the user input.

70. At least one computer readable medium for controlling at least one computer system to perform document processing method, comprising:

producing at least one index of n-word sequences in documents derived from at least one of spontaneous and conversational expression and containing non-deterministic text with average word recognition precision below 50 percent, utilizing non-textual differences between the n-word sequences, where n is a positive integer; and

processing the documents based on the non-textual differences between the n-word sequences in the at least one index, where said processing resolves more than two decision options.

71. At least one computer readable medium as recited in claim 70, wherein the non-textual differences between the n-word sequences relate to recognition confidence of the n-word sequences.

72. An apparatus for processing documents, comprising:

processing means for processing documents derived from at least one of spontaneous and conversational expression and containing non-deterministic text with average word recognition precision below 50 percent, said processing utilizing non-textual differences between n-word sequences in the documents, where n is a positive integer and said processing resolves more than two decision options.

73. An apparatus as recited in claim 72, wherein said processing means comprises index means for producing at least one index of the n-word sequences utilizing the non-textual differences between the n-word sequences.

74. An apparatus as recited in claim 73, wherein said processing means comprises data mining means for retrieving at least one of the documents utilizing the at least one index.

75. An apparatus as recited in claim 74,

wherein said data mining means comprises:

parameter extraction means for extracting parameters from the documents, utilizing the non-textual differences between said n-word sequences; and

relations establishment means for establishing relations between the parameters extracted from the documents, and wherein said apparatus further comprises display means for producing graphic results indicating the relations between the parameters extracted from the documents.

76. An apparatus as recited in claim 75, wherein at least one of the extracted parameters is an assessment of a hidden variable that cannot be fully determined from information existing in the at least one of the documents.

77. An apparatus as recited in claim 72, wherein the non-textual differences between the n-word sequences relate to recognition confidence of the n-word sequences.

78. An apparatus as recited in claim 72, wherein said processing means comprises categorizing means for categorizing the documents utilizing at least one algorithm based on non-linguistic differences between the n-word sequences.

79. An apparatus as recited in claim 72, wherein said processing means comprises clustering means for clustering the documents by assessing a relation between the n-word sequences based on non-textual differences.

80. An apparatus as recited in claim 72, wherein said processing means comprises means for at least one of classifying and filtering the documents as the documents are received.

81. An apparatus as recited in claim 72, further comprising display means for displaying information related to at least one of the documents, including at least some of the non-textual differences between the n-word sequences.

82. An apparatus as recited in claim 81, wherein said display means selectively displays portions of the at least one of the documents based on confidence of accuracy of words displayed.

83. An apparatus as recited in claim 72,

further comprising input means for receiving user input indicating errors in recognition, and wherein said processing means comprises means for replacing at least one word in the at least one of the documents with a corrected word based on the user input and setting the confidence levels of the corrected word to indicate high recognition accuracy.

84. An apparatus as recited in claim 72, coupled to a telephone system and further comprising automatic speech recognition means for generating the documents by automatic speech recognition of audio signals received via the telephone system.

85. An apparatus as recited in claim 72, further comprising automatic character recognition means for generating the documents by automatic character recognition.

86. An apparatus as recited in claim 72, wherein said processing means comprises:

data mining means for applying different data mining techniques, each of which does not indicate non-textual differences; and

merge means for merging results of the different data mining techniques to obtain the non-textual differences between the n-word sequences.

87. An apparatus as recited in claim 86, wherein said data mining means includes means for at least one of retrieving, categorizing, filtering, classifying, labeling and clustering documents without utilization of any non-textual differences between the n-word sequences.

88. An apparatus as recited in claim 87, wherein said data mining means uses a plurality of different algorithms to transform non-deterministic text into standard text documents usable in text mining and the data mining techniques operate on the standard text documents.

89. An apparatus as recited in claim 87,

further comprising indexing means for generating a plurality of indexes of the standard text documents, and

wherein said data mining means uses the different indexes in applying the different data mining techniques.

90. An apparatus as recited in claim 89,

further comprising input means for receiving a query; and wherein said data mining means further includes retrieving means for retrieving the results relevant to the query.

91. A data processing system, comprising:

at least one server to process documents, derived from at least one of spontaneous and conversational expression and containing non-deterministic text with word recognition precision of less than 50 percent, utilizing non-textual differences between n-word sequences, where n is a positive integer.

92. A data processing system as recited in claim 91, wherein said at least one server includes an indexing server producing at least one index of the n-word sequences utilizing the non-textual differences between the n-word sequences,

93. A data processing system as recited in claim 92, wherein said indexing server retrieves at least one of the documents utilizing data mining of the at least one index.

94. A data processing system as recited in claim 91,

wherein said at least one server extracts parameters from the documents, utilizing the non-textual differences between said n-word sequences, and establishes relations between the parameters extracted from the documents, and

wherein said data processing system further comprises at least one display device producing graphic results indicating the relations between the parameters extracted from the documents.

95. A data processing system as recited in claim 94, wherein at least one of the extracted parameters is an assessment of a hidden variable that cannot be fully determined from information existing in the at least one of the documents.

96. A data processing system as recited in claim 91, wherein the non-textual differences between the n-word sequences relate to recognition confidence of the n-word sequences.

97. A data processing system as recited in claim 91, further comprising at least one display device displaying information related to at least one of the documents, including at least some of the non-textual differences between the n-word sequences

98. A data processing system as recited in claim 97, wherein said at least one display device selectively displays portions of at least one of the documents based on confidence of accuracy of words displayed.

99. A data processing system as recited in claim 91, wherein said at least one server applies different data mining techniques, each of which does not indicate non-textual differences and merges results of the different data mining techniques to obtain the non-textual differences between the n-word sequences.

100. A data processing system as recited in claim 99, wherein said at least one server uses a plurality of different algorithms to transform non-deterministic text into standard text documents usable in text mining and the data mining techniques operate on the standard text documents.

101. A data processing system as recited in claim 100, wherein said at least one server generates a plurality of indexes of the standard text documents and uses the different indexes in applying the different data mining techniques.

102. A data processing system as recited in claim 99, wherein the different data mining techniques include at least one of retrieving, categorizing, filtering, classifying, labeling and clustering documents without utilization of any non-textual differences between the n-word sequences.

103. A data processing system as recited in claim 102, wherein said at least one server uses a plurality of different algorithms to transform non-deterministic text into standard text documents usable in text mining and the data mining techniques operate on the standard text documents.

104. A data processing system as recited in claim 91,

further comprising at least one user terminal providing user input indicating errors in recognition in a document, and

wherein said at least one server replaces at least one word in the document with a corrected word based on the user input and sets confidence levels of the corrected word to indicate high recognition accuracy.

105. A data processing system as recited in claim 91, further comprising at least one of an automatic speech recognition unit, an automatic character recognition unit and a fact extraction unit to generate the documents from data that on average produces word recognition precision of less than 50 percent.

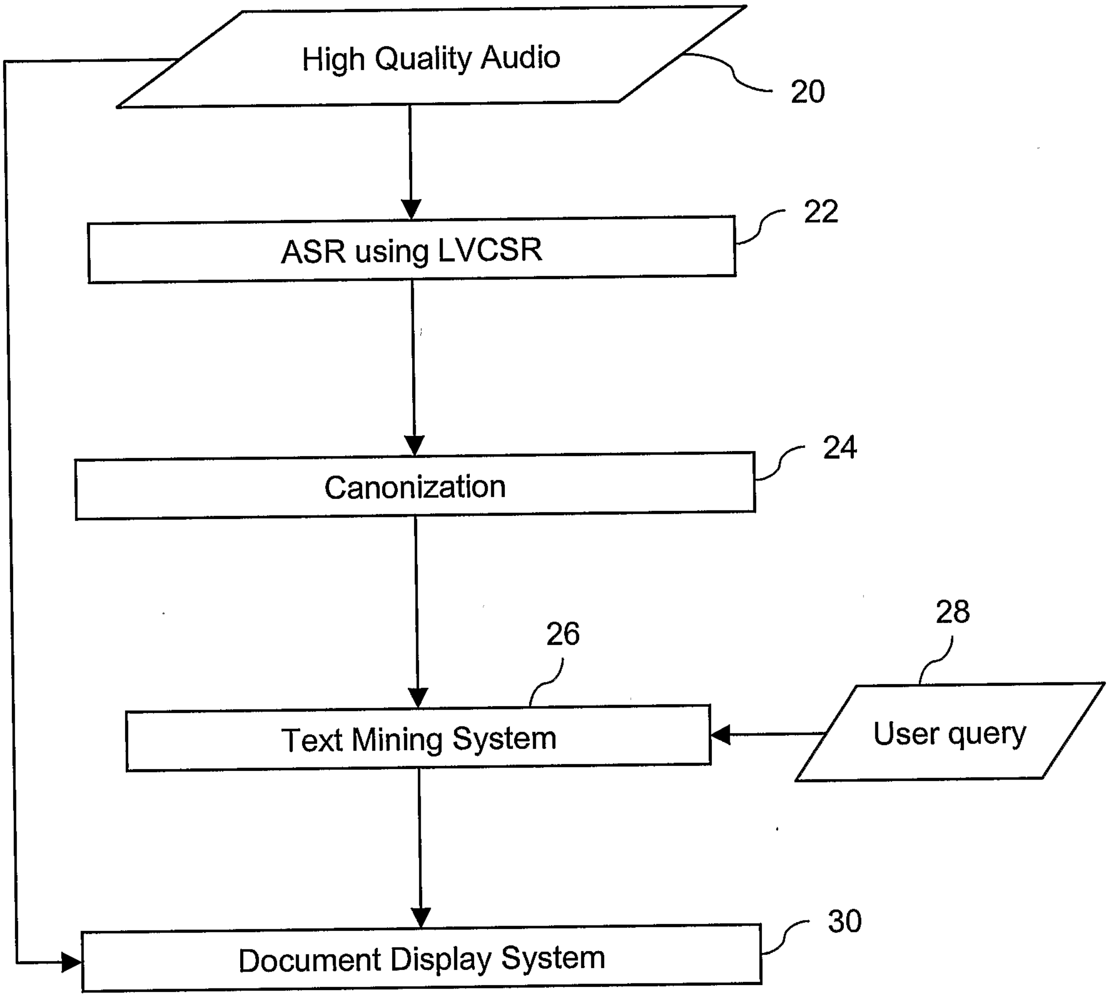


FIG. 1

PRIOR ART

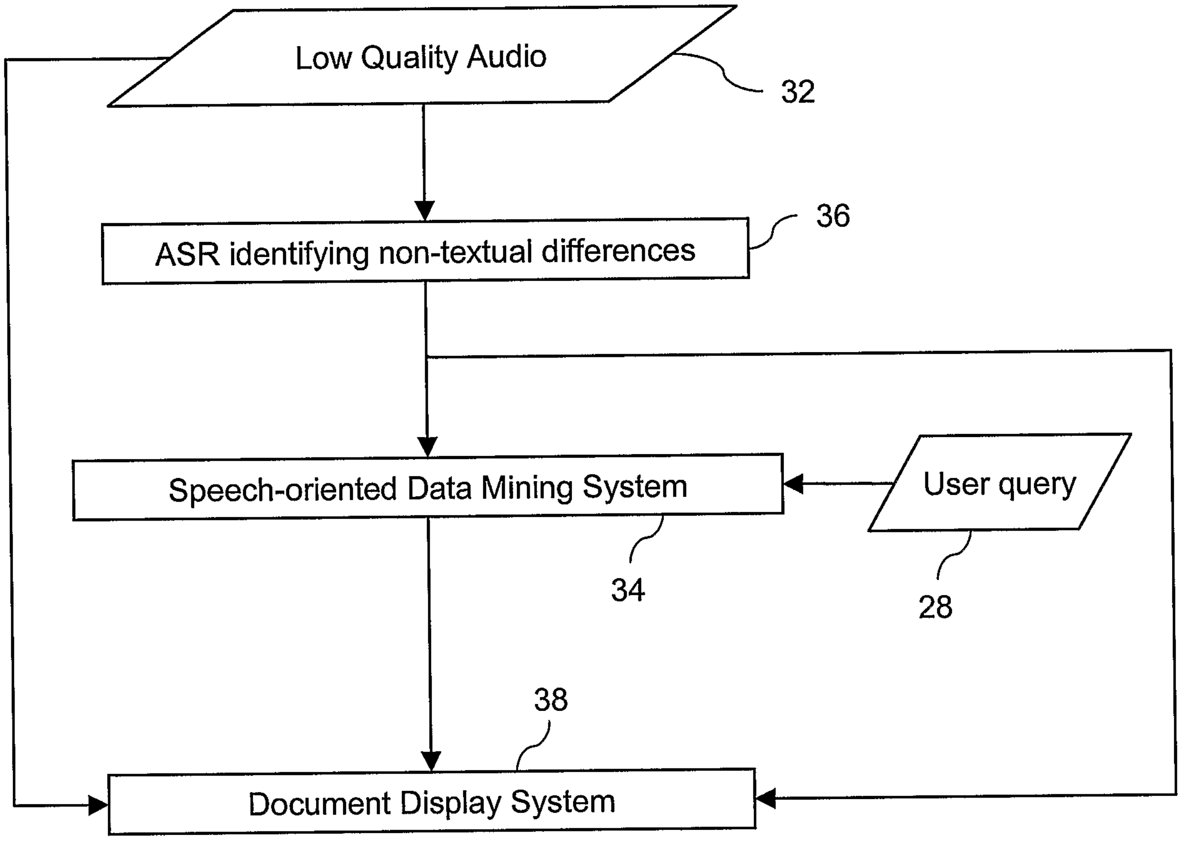

Low Quality Audio
32
ASR identifying non-textual differences
36
Speech-oriented Data Mining System
34
User query
28
38
Document Display System

FIG. 2

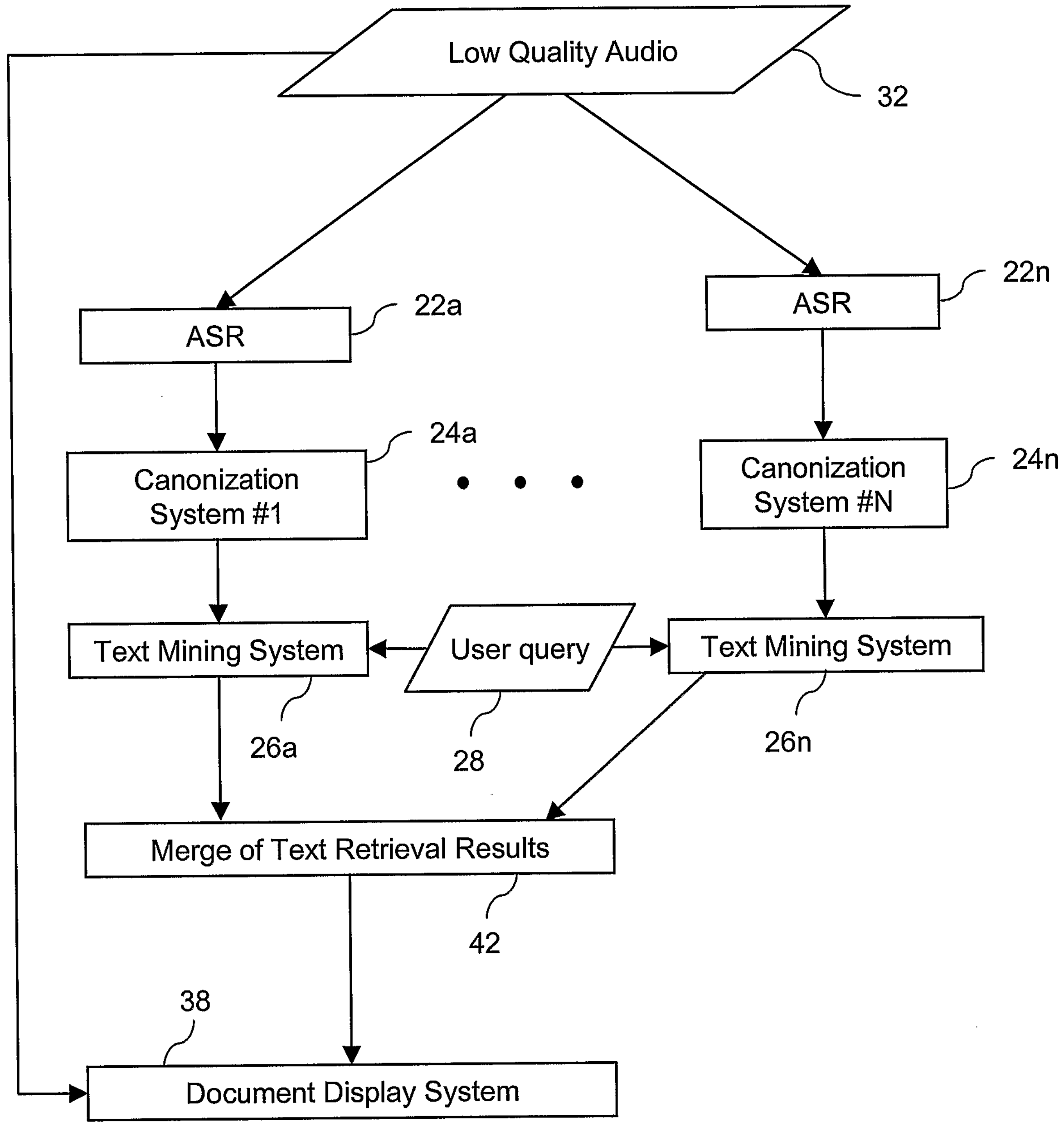

Low Quality Audio
32
ASR
22a
ASR
22n
Canonization System #1
24a
• • •
Canonization System #N
24n
Text Mining System
User query
Text Mining System
26a
28
26n
Merge of Text Retrieval Results
42
38
Document Display System

FIG. 3

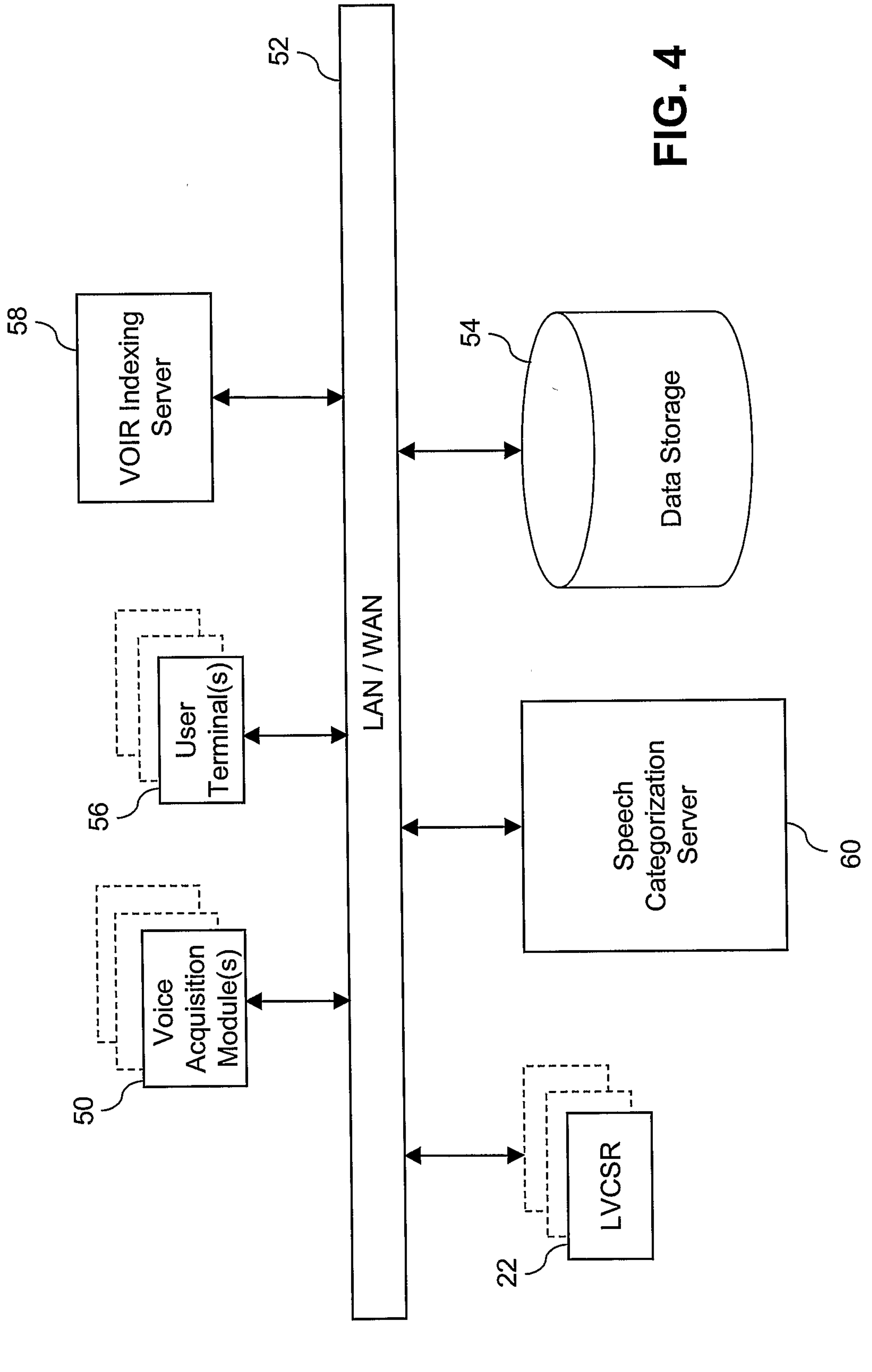

52
LAN / WAN
58
VOIR Indexing Server
56
User Terminal(s)
50
Voice Acquisition Module(s)
54
Data Storage
60
Speech Categorization Server
22
LVCSR

FIG. 4

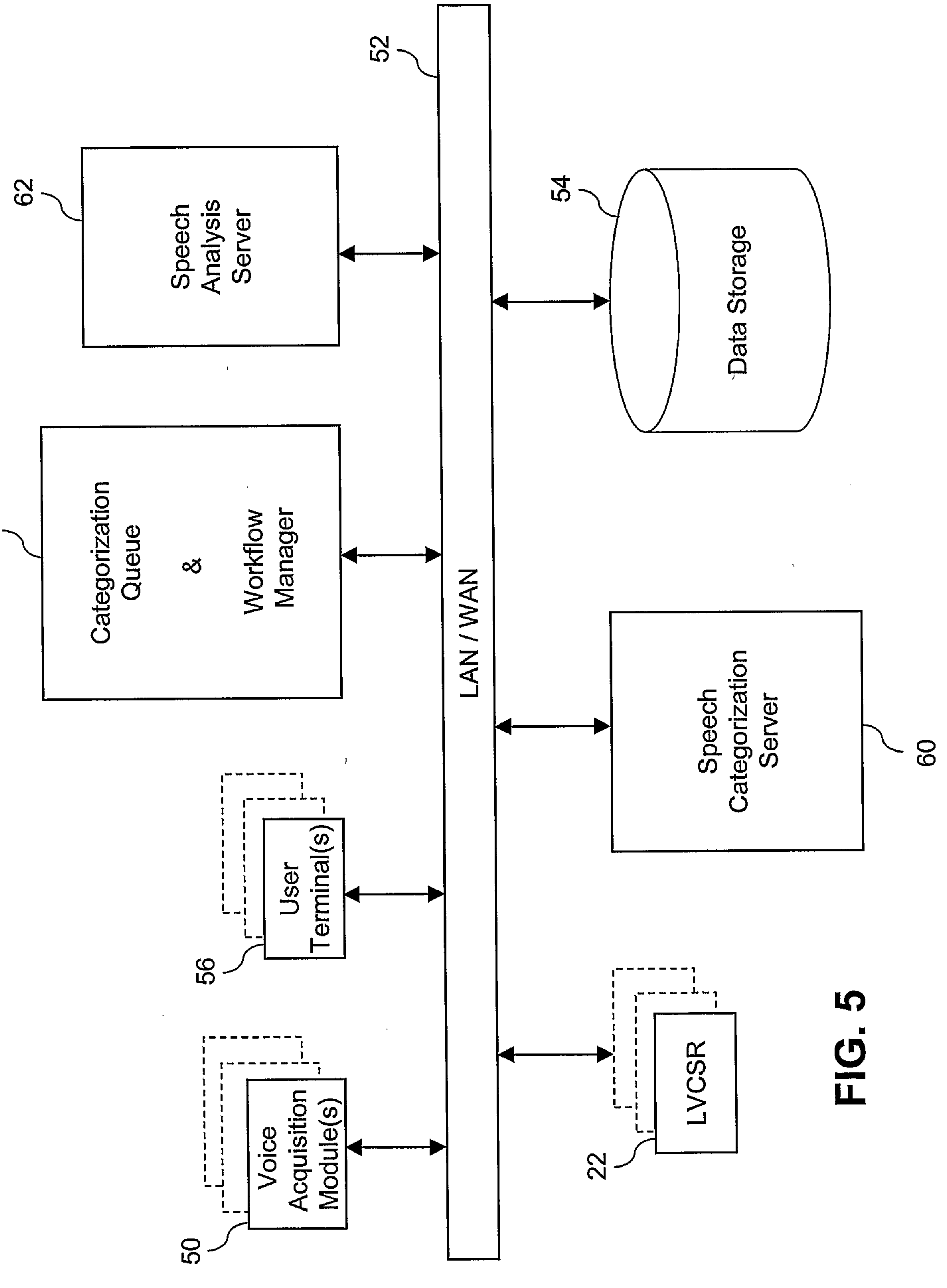

52
LAN / WAN
62
Speech Analysis Server
64
Categorization Queue & Workflow Manager
56
User Terminal(s)
50
Voice Acquisition Module(s)
54
Data Storage
60
Speech Categorization Server
22
LVCSR

FIG. 5

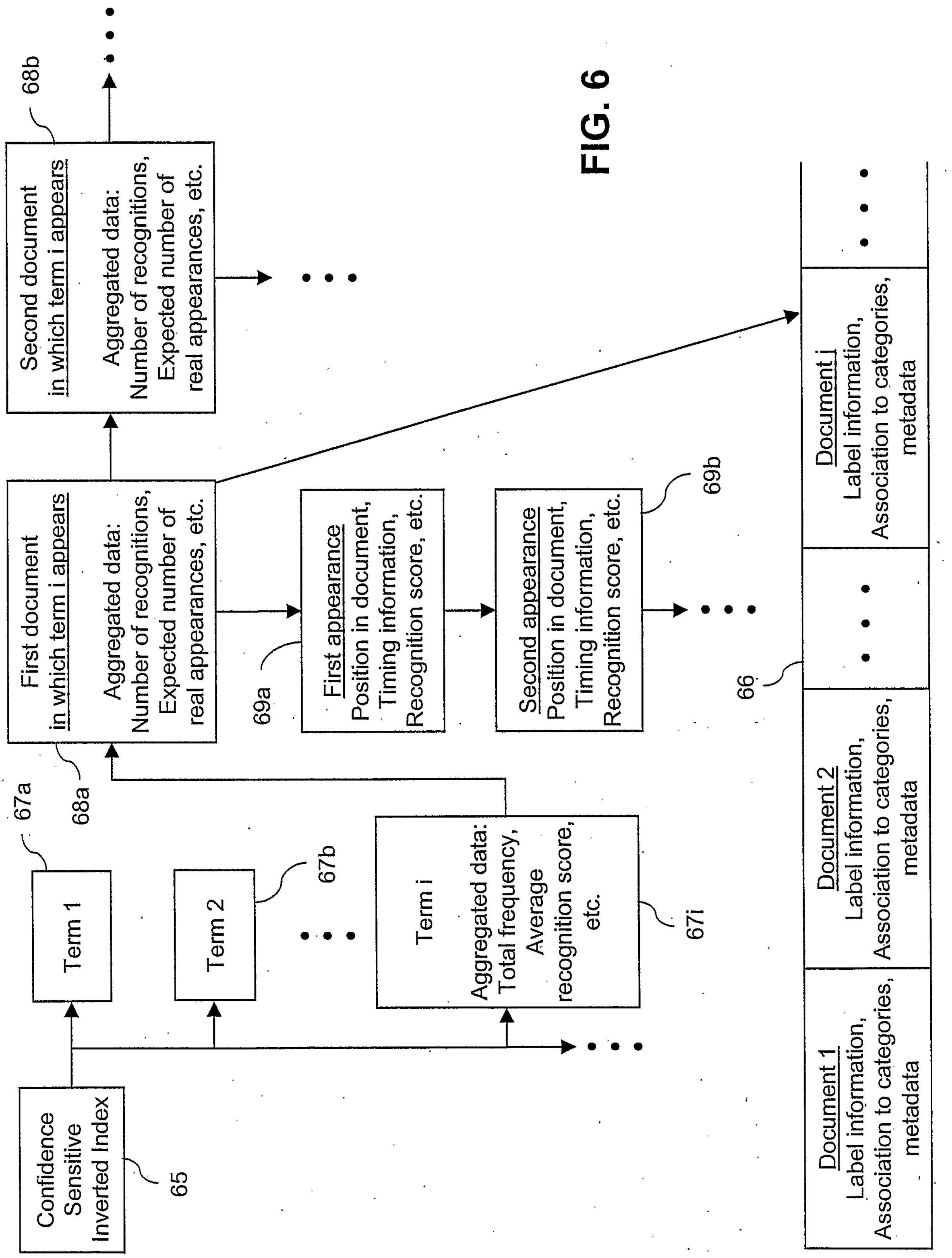

Confidence Sensitive Inverted Index
65
Term 1
67a
Term 2
67b
Term i
Aggregated data: Total frequency, Average recognition score, etc.
67i
First document in which term i appears
Aggregated data: Number of recognitions, Expected number of real appearances, etc.
68a
Second document in which term i appears
Aggregated data: Number of recognitions, Expected number of real appearances, etc.
68b
First appearance
Position in document, Timing information, Recognition score, etc.
69a
Second appearance
Position in document, Timing information, Recognition score, etc.
69b
Document 1
Label information, Association to categories, metadata
Document 2
Label information, Association to categories, metadata
Document j
Label information, Association to categories, metadata
66

FIG. 6

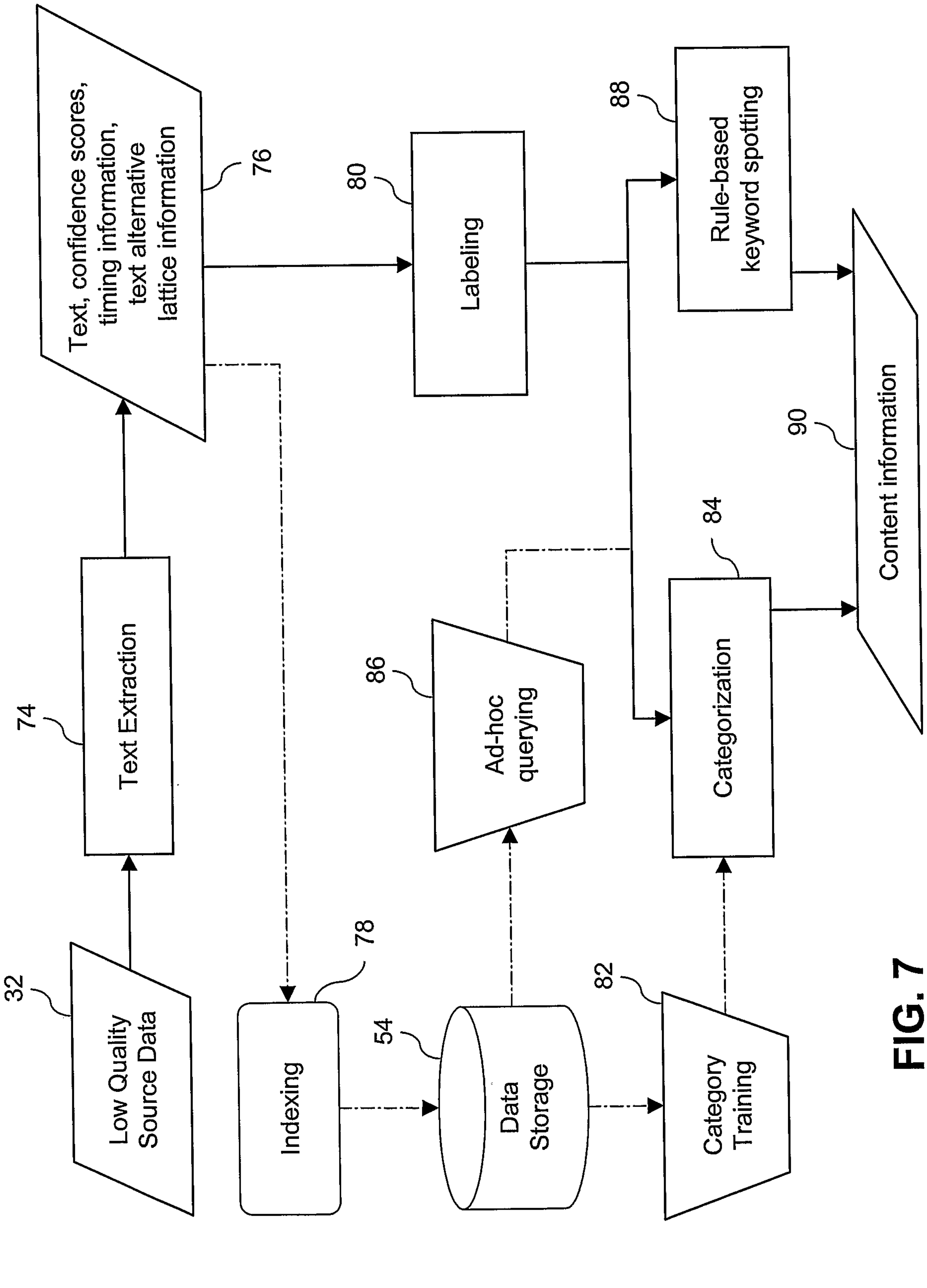
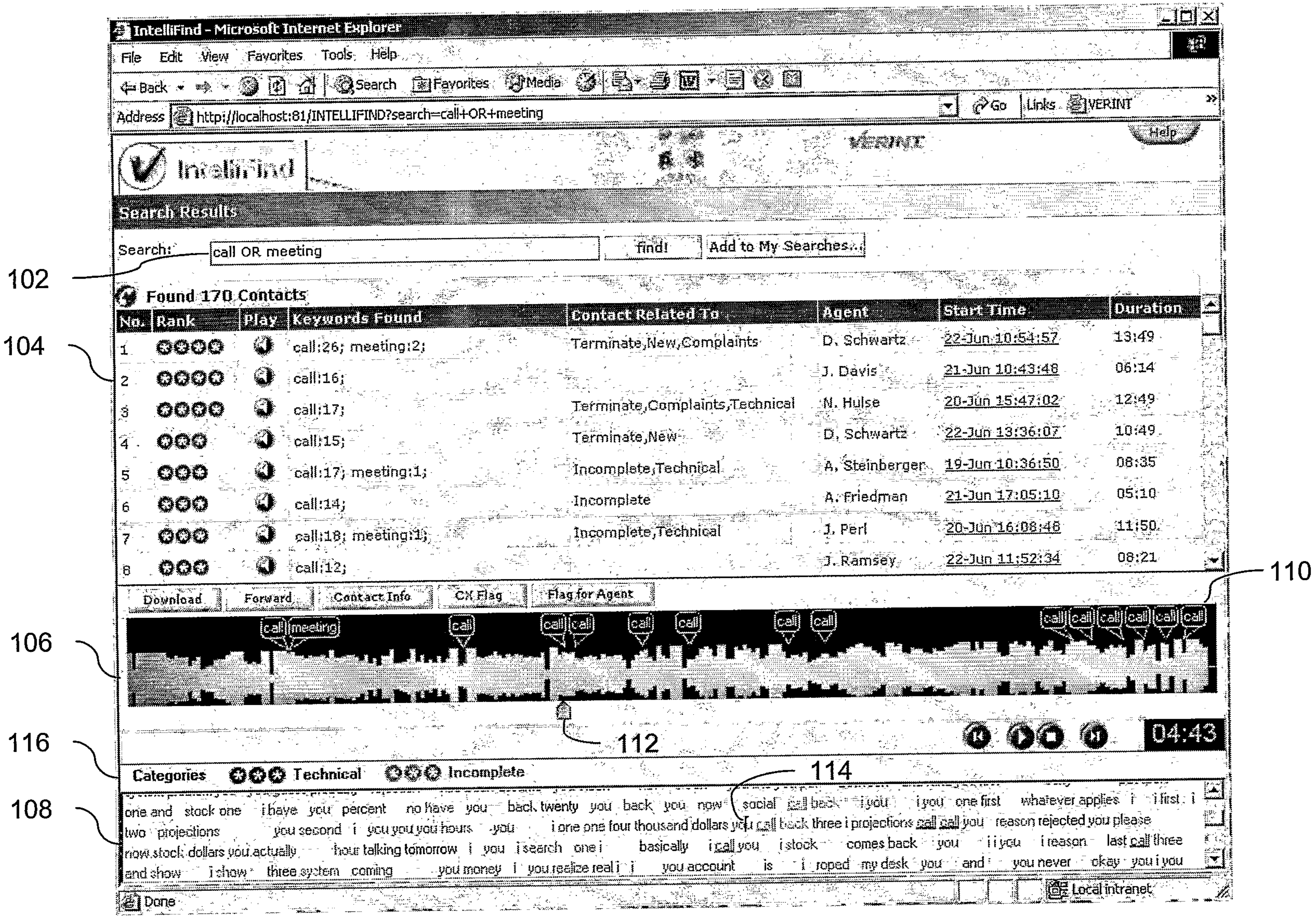

32
Low Quality Source Data
74
Text Extraction
76
Text, confidence scores, timing information, text alternative lattice information
78
Indexing
54
Data Storage
80
Labeling
86
Ad-hoc querying
82
Category Training
84
Categorization
88
Rule-based keyword spotting
90
Content information

FIG. 7



Address http://localhost:81/INTELLIFIND?search=call+OR+meeting

Search: call OR meeting [find!] [Add to My Searches...]

Found 170 Contacts

No.	Rank	Play	Keywords Found	Contact Related To	Agent	Start Time	Duration
1	●●●●	▶	call:26; meeting:2;	Terminate,New,Complaints	D. Schwartz	22-Jun 10:54:57	13:49
2	●●●●	▶	call:16;		J. Davis	21-Jun 10:43:48	06:14
3	●●●●	▶	call:17;	Terminate,Complaints,Technical	N. Hulse	20-Jun 15:47:02	12:49
4	●●●	▶	call:15;	Terminate,New	D. Schwartz	22-Jun 13:36:07	10:49
5	●●●	▶	call:17; meeting:1;	Incomplete,Technical	A. Steinberger	19-Jun 10:36:50	08:35
6	●●●	▶	call:14;	Incomplete	A. Friedman	21-Jun 17:05:10	05:10
7	●●●	▶	call:18; meeting:1;	Incomplete,Technical	J. Perl	20-Jun 16:08:48	11:50
8	●●●	▶	call:12;		J. Ramsey	22-Jun 11:52:34	08:21

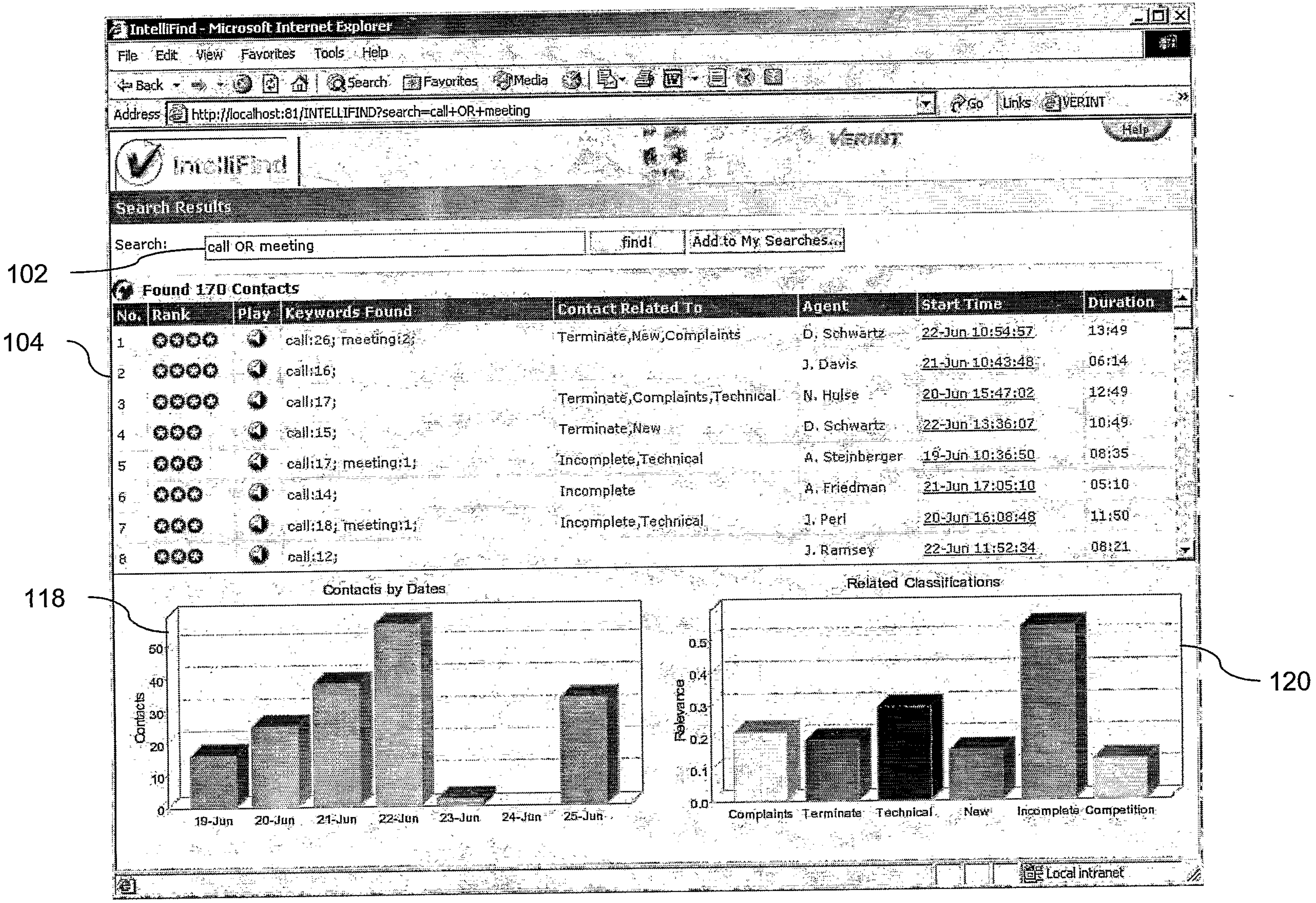
[Download] [Forward] [Contact Info] [CX Flag] [Flag for Agent]

04:43

Categories ●●● Technical ●●● Incomplete

one and stock one i have you percent no have you back twenty you back you now social call back i you i you one first whatever applies i i first i
two projections you second i you you you hours you i one one four thousand dollars you call back three i projections call call you reason rejected you please
now stock dollars you actually hour talking tomorrow i you i search one i basically i call you i stock comes back you i i you i reason last call three
and show i show three system coming you money i you realize real i i you account is i roped my desk you and you never okay you i you

FIG. 8


IntelliFind - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://localhost:81/INTELLIFIND?search=call+OR+meeting
Go Links VERINT
IntelliFind
VERINT
Help
Search Results
102
104
118
120

Search: call OR meeting find! Add to My Searches...

Found 170 Contacts

No.	Rank	Play	Keywords Found	Contact Related To	Agent	Start Time	Duration
1	✪✪✪✪	◀	call:26; meeting:2;	Terminate,New,Complaints	D. Schwartz	22-Jun 10:54:57	13:49
2	✪✪✪✪	◀	call:16;		J. Davis	21-Jun 10:43:48	06:14
3	✪✪✪✪	◀	call:17;	Terminate,Complaints,Technical	N. Hulse	20-Jun 15:47:02	12:49
4	✪✪✪	◀	call:15;	Terminate,New	D. Schwartz	22-Jun 13:36:07	10:49
5	✪✪✪	◀	call:17; meeting:1;	Incomplete,Technical	A. Steinberger	19-Jun 10:36:50	08:35
6	✪✪✪	◀	call:14;	Incomplete	A. Friedman	21-Jun 17:05:10	05:10
7	✪✪✪	◀	call:18; meeting:1;	Incomplete,Technical	J. Perl	20-Jun 16:08:48	11:50
8	✪✪✪	◀	call:12;		J. Ramsey	22-Jun 11:52:34	08:21

FIG. 9