HAL
archives-ouvertes.fr

The Cambridge University Multimedia Document Retrieval Demo System

A Tuerk, S Johnson, Pierre Jourlin, K Jones, P Woodland

▶ **To cite this version:**

A Tuerk, S Johnson, Pierre Jourlin, K Jones, P Woodland. The Cambridge University Multimedia Document Retrieval Demo System. International Journal of Speech Technology, Springer Verlag, 2001, 4, pp.241 - 250. 10.1023/A:1011360624662 . hal-02171704

HAL Id: hal-02171704

https://hal-univ-avignon.archives-ouvertes.fr/hal-02171704

Submitted on 8 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.


INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY 4, 241–250, 2001
© 2001 Kluwer Academic Publishers. Manufactured in The Netherlands.

The Cambridge University Multimedia Document Retrieval Demo System

A. TUERK AND S.E. JOHNSON
Cambridge University Engineering Department, Trumpington Street, Cambridge, CB2 1PZ, UK
at233@eng.cam.ac.uk
sej28@eng.cam.ac.uk

P. JOURLIN AND K. SPÄRCK JONES
Cambridge University Computer Laboratory, Pembroke Street, Cambridge, CB2 3QG, UK
pj207@cl.cam.ac.uk
ksj@cl.cam.ac.uk

P.C. WOODLAND
Cambridge University Engineering Department, Trumpington Street, Cambridge, CB2 1PZ, UK
pcw@eng.cam.ac.uk

Received August 11, 2000; Revised March 22, 2001

Abstract. The Cambridge University Multimedia Document Retrieval (CU-MDR) Demo System is a web-based application that allows the user to query a database of radio broadcasts that are available on the Internet. The audio from several radio stations is downloaded and transcribed automatically. This gives a collection of text and audio documents that can be searched by a user. The paper describes how speech recognition and information retrieval techniques are combined in the CU-MDR Demo System and shows how the user can interact with it.

Keywords: information retrieval, spoken document retrieval, speech recognition

1. Introduction

To provide content-specific access to the vast amount of text data that is available on the Internet, search engines have been developed that operate on text documents of various formats. Since there is an increasing amount of audio data containing speech on the Internet, a similar device is desirable that operates automatically on audio streams without the need for manual transcription. The Cambridge University Multimedia Document Retrieval (CU-MDR) demo system attempts to fill this need.

2. Related Work

In the recent past, several information retrieval systems have been developed that allow users to query a database of broadcast news audio. As well as the CU-MDR system, there are Informedia (Witbrock and Hauptmann, 1997), SpeechBot (Van Thong et al., 2000) and the THISL demo (Renals et al., 2000), to name but a few. While all these systems use a speech recognition engine as a central element in their architecture, the SpeechBot system is the one that is most closely related to the CU-MDR demo. Both the CU-MDR demo

and SpeechBot recognise compressed Internet audio whereas the other two systems recognise uncompressed data from FM radio and TV broadcasts. In addition, both these systems as well as the THISL demo have a web-based interface. In contrast to the other systems, high recognition accuracy has been a major goal in the development of the CU-MDR demo. While this is not crucial for the purpose of maximising a standard information retrieval performance metric (Garofolo et al., 2000), it is necessary if the user wants to browse the transcriptions of the returned documents interactively to find the documents that truly satisfy the user's information need. Browsing the automatic transcripts, however, is preferable to browsing the audio because downloading audio data can take a long time, especially over the Internet. Two other unique features of the CU-MDR demo system are interactive query expansion and the possibility to filter incoming news broadcasts with a set of fixed queries. These two features are part of the CU-MDR demo interface.

3. System Description

3.1. Overview

The CU-MDR demo system downloads the audio track of news broadcasts from the Internet once a day. Audio, which is in the compressed RealAudio format,[1] is converted into standard wav format from which a transcription is produced using our large vocabulary broadcast news recognition engine. This yields a collection of text and audio documents which can be searched by the user. A request by a user, which can be a natural language query or just a list of words, starts a search on the collection of text documents. The returned documents can then be browsed as both text and audio.

3.2. Data Sources

The CU-MDR demo currently downloads radio broadcasts from the Internet sites of the two American news broadcasters, NPR[2] and PBS,[3] and British news from the BBC.[4] Data are downloaded on a daily basis. From the NPR web site we download about two hours of data during the week and one hour at weekends; from the PBS web site it is between 45 minutes and one hour on weekdays only; and from the BBC web site one hour during the week and 30 minutes at weekends. In July 2000, the total size of the CU-MDR database was about 700 hours, with the earliest documents in the database dating back to May 1999. The downloaded data vary in compression and segmentation. While PBS and NPR data are segmented into different stories, the BBC data are available only as a continuous stream of either one hour or 30-minute broadcasts. As for compression, NPR data are compressed with the 8.5 kbps RealAudio codec, PBS data retain a relatively high quality with 32.1 kbps compression and BBC data are compressed with the 16 kbps RealAudio codec. The CU-MDR demo database covers most of the international and main national news in the U.S. and Great Britain. However, to extend the contents of the CU-MDR demo to other areas of interest we are planning to download data from a larger number of web sites.

4. The Speech Recognition Module

4.1. The MDR Recogniser

The CU-MDR system embeds a Hidden Markov Model (HMM) recogniser that uses two separate passes to transcribe the audio data. This recogniser has a vocabulary of approximately 60,000 words and is similar to the system described in Odell et al. (1999).

The audio stream is split first into acoustically homogeneous segments and labelled as wideband speech, narrowband speech and music. Music segments are discarded. For recognition, a 39-dimensional feature vector is used that consists of 13 PLP coefficients and their first and second order derivatives. Cepstral mean normalisation is performed for each segment.

For the first recognition pass, the recogniser uses one of two sets of gender-independent HMMs, which are aimed at either US English or British English broadcast news sources. For the U.S. NPR and PBS broadcast data, the acoustic models were trained on the 140 hours of the 1997 and 1998 U.S. Broadcast News training data (Pallett et al., 1998, 1999). The models for British English sources were trained on this data supplemented with an additional 50 hours of BBC news data, which was transcribed and made available to us by the THISL project (Renals et al., 2000). The HMMs are cross-word triphone Gaussian mixture models with 16 components per distribution. The CU-MDR demo recogniser uses a 4-gram language model that was trained on approximately 260 million words of broadcast news and newspaper articles. For PBS and NPR data the first pass transcriptions are subsequently used to determine the gender for each segment by performing a forced alignment with gender-dependent models and assigning the

gender of the model with the highest likelihood to the segment. This is not done for BBC data because no gender-dependent British English models have been trained, due to the lack of gender-specific information in the manual transcription of the BBC training data. For each bandwidth and, in the case of NPR and PBS, each gender, the segments are clustered using a covariance-based algorithm described in Johnson and Woodland (1998). The clusters together with the transcriptions of the first pass are used to perform unsupervised adaptation of the gender-dependent American English and the gender-independent British English second pass models using MLLR (Leggetter and Woodland, 1995). The acoustic models for the second pass have a higher resolution than the models for the first pass with 20 components per mixture Gaussian. The 4-gram language model is used to generate the final hypothesis.

4.2. Recognition Experiments

To evaluate the performance of the MDR demo recogniser on the Internet audio we compressed the 1997 Hub4 broadcast news test data (Pallett et al., 1998) with RealProducer Plus 6.0.3[5] to 16 kbps and 8.5 kbps. The size of the 16 kbps and 8.5 kbps data were 15 and 28 times smaller, respectively, than the size of the uncompressed data. The sampling rate of the 16 kbps data is 16 kHz, whereas it is 8 kHz for the 8.5 kbps data. Therefore, only narrow-band models were used to recognise the 8.5 kbps data. The recogniser in these experiments used the same wide-band and narrow-band models that were used in Odell et al. (1999). This means that the acoustic models were not trained specifically on the compressed data. The results of these recognition experiments can be found in Tables 1–3. The first row of these tables shows the percent word error rate (WER) for the first recognition pass. The second row gives the WER for the second pass. The columns show the error rates for the different F-conditions as defined in the 1997 Hub4 broadcast news task (Pallett et al., 1998). The meaning of the F-conditions is given in Table 4.

As can be seen from Tables 1 and 2 the overall increase in word error rate for 16 kbps compressed data is only 4% relative to the uncompressed rate. This is quite remarkable given that the models were not trained on the compressed data. For 8.5 kbps the word error rate increases by 32% relative to the uncompressed rate but is still within a reasonable range to allow efficient document retrieval. The differences in recognition performance between the different compression levels were all found to be statistically significant at the 0.1% level using the NIST matched-pair sentence segment word error rate test.

The recogniser, excluding segmentation, ran in 9.5 times real-time[6] on the uncompressed data of which the first pass took about 2 times real-time. On 16 kbps the run time was increased to 11.4 times real-time,

Table 1. Word error rates on the 1997 Broadcast News task for uncompressed data.

	Overall	F0	F1	F2	F3	F4	F5	FX
1st pass	20.6	12.6	19.6	25.3	31.5	25.3	25.3	41.2
2nd pass	16.4	9.8	16.1	19.5	24.9	20.1	19.9	33.6

Table 2. Word error rates on the 1997 Broadcast News task for 16 kbps compressed data.

	Overall	F0	F1	F2	F3	F4	F5	FX
1st pass	21.5	13.5	21.2	26.3	32.7	26.5	23.5	40.6
2nd pass	17.0	10.4	17.0	20.9	25.9	22.1	21.4	30.8

Table 3. Word error rates on the 1997 Broadcast News task for 8.5 kbps compressed data.

	Overall	F0	F1	F2	F3	F4	F5	FX
1st pass	27.2	18.5	26.9	28.1	39.5	35.5	32.4	51.7
2nd pass	21.7	14.2	21.6	23.0	33.2	27.0	26.6	42.0

Table 4. F-conditions for the 1997 Broadcast News task.

F0	planned, low noise broadcast speech
F1	spontaneous broadcast speech
F2	speech over telephone channels
F3	speech in the presence of background music
F4	speech under degraded acoustic conditions
F5	speech from non-native speakers
FX	all other speech

and on 8.5 kbps the run time was 15.9 times real-time. Because recognition accuracy was satisfactory for the purpose of document retrieval no further development to improve the performance of the recogniser on these types of compressed data has been done. To evaluate the recognition performance on some typical Internet audio broadcasts we manually transcribed a set of documents of about one hour's length which was made up of approximately one-half hour of NPR data and one-half hour of PBS data. These experiments showed a word error rate of 19.6% on the NPR data and a word error rate of 33.4% on the PBS data. The relatively large difference in recognition accuracy for NPR and PBS data is a result of the different nature of the data prior to compression. Whereas NPR data is predominantly composed of planned speech (F0-type data), PBS contains a considerable portion of spontaneous speech (F1-type data) and data with background noise (F4, FX-type data). For this reason, recognition performance on PBS data is worse than on NPR data even though PBS data are compressed to 32.1 kbps, whereas NPR are compressed only to 8.5 kbps.

5. The Information Retrieval Module

The basic information retrieval engine used in the CU-MDR demo is the Okapi-based benchmark system described in Johnson et al. (2000b). The IR system uses stopping and Porter stemming (Porter, 1980) to preprocess queries and documents. Following Robertson and Spärck Jones (1997) and Spärck Jones et al. (1998), the score for a given query and document is calculated by summing the tf-idf based combined weights for each query term. The final ranked list of documents for a query is produced by sorting these scores in descending order.

In the CU-MDR demo this basic system also can be used with query expansion using two rather different but complimentary mechanisms. Semantic posets, which are discussed in Jourlin et al. (2000), are a structure that allows the system to exploit semantic information and, here, has been realised with the help of a geographic database and WordNet1.6 (Fellbaum, 1998). This structure can be used either to add words that are semantically equivalent to the query or to add subcategories of places. An example of the first situation is, for instance, the addition of the word "flu" to a query containing the word "influenza", whereas an example for the latter case is the addition of "London" to a query containing "England".

Relevance feedback also is available for query expansion. This allows the user to mark the documents that contain relevant information and have the system suggest additional query words that distinguish those documents from the non-relevant ones. These words are chosen according to their Offer weight (Spärck Jones et al., 1998). When activated, both query expansion methods bring up a list of words from which the user can select the words that are perceived to be helpful in expanding the query.

The retriever can work not only with documents where the story boundaries are known, but also with documents where this is not the case. In the latter situation the complete document is split into windows of 30 seconds in length whose midpoints are 15 seconds apart. This results in a collection of pseudo-documents, which are searched by the retriever. A post-processing step identifies the windows from the ranked list returned by the retriever, which originated from within a time period of 30 seconds in the same show. It is assumed that these windows came from the same story source, and hence the corresponding windows are merged to produce a single quasi-story with a new start and end time which is presented to the user. The score for this new quasi-story is taken as that of the highest scoring window used when forming it. This process is similar to that described in more detail in Johnson et al. (2000a).

The retrieval system of the CU-MDR demo was formally evaluated in the TREC8 evaluation (Johnson et al., 2000b) and found to give state of the art performance (Garofolo et al., 2000).

6. The CU-MDR Demo Interface

6.1. Login and Search Page

The registered user accesses the MDR demo system via the login page. After successfully logging into the

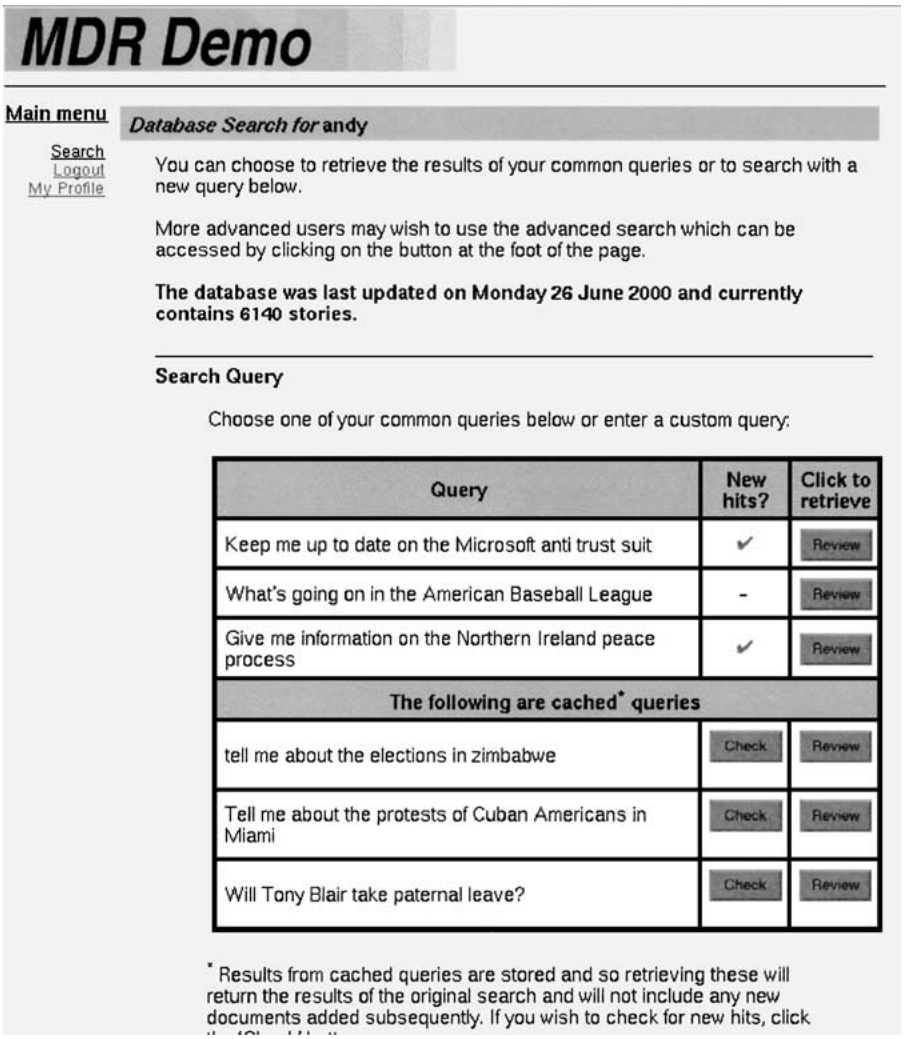

Figure 1. MDR demo search page. This page contains information about a user's fixed and most recent queries.

system the user is transferred to the search page that contains information about the user's queries. An example of such a page can be seen in Fig. 1. The upper half of the query table contains the user's fixed queries. This set of queries is looked up in the database whenever the user logs into the system. If new documents are found which have not been seen before, the user is notified by a green tick in the box adjacent to the query. In the example (see Fig. 1) the information retriever found new matches for the queries "Keep me up to date on the Microsoft anti-trust suit" and "Give me information on the Northern Ireland peace process". This feature effectively allows the user to filter the stories from incoming broadcasts. The lower half of the table records the three most recent queries that the user has entered. One can either choose to see the results of the last search by clicking on the "Review" button or start a new search with the query by clicking on the "Check" button. Alternatively, a new query can be typed into the search field which is below the query table.

6.2. *Presentation of Search Results*

Once the retriever has returned the results for a particular search, a list of extracts from the returned text documents is created. Each extract is designed to represent the section of approximately 100 words in the complete document that is most relevant to the query. A small subset of such a list can be seen in Fig. 2.

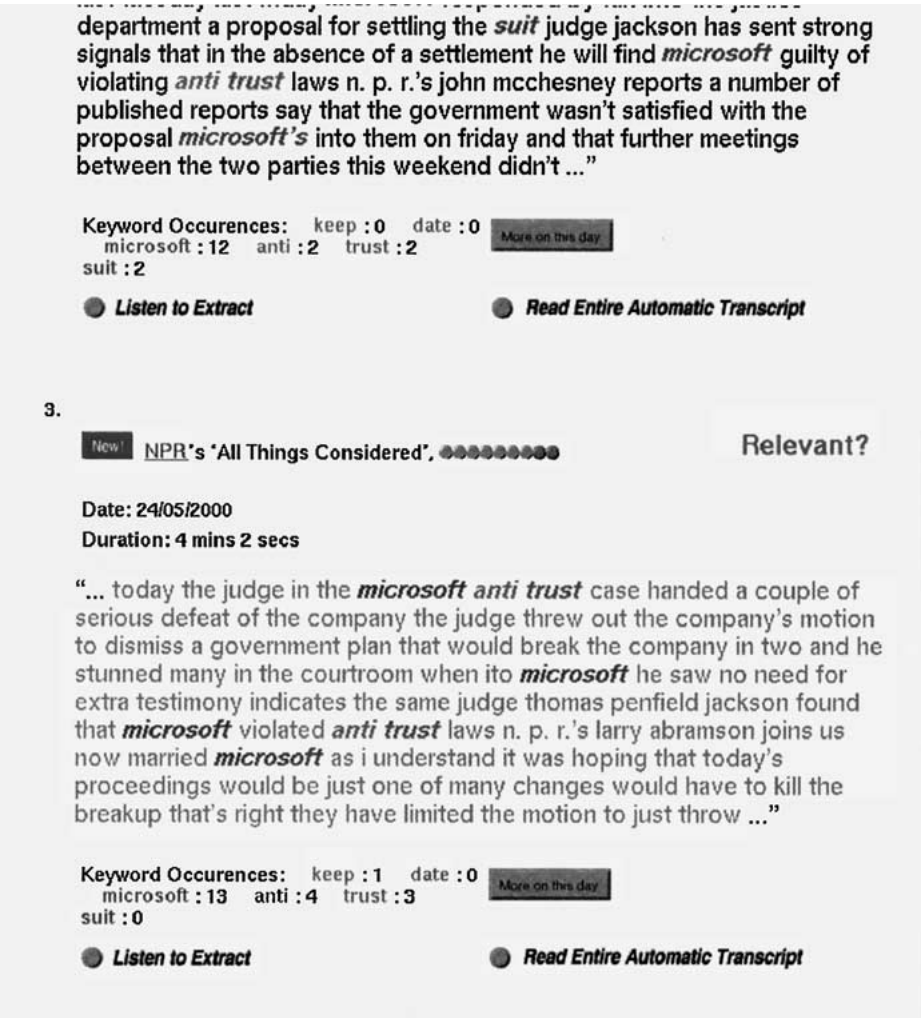

Figure 2. Part of the search results for the fixed query "Keep me up to date on the Microsoft anti-trust suit".

This figure shows the document with the third highest retriever score in the list of documents that were returned for the fixed query "Keep me up to date on the Microsoft anti-trust suit". The extract is highlighted because it was found for one of the user's fixed queries and has not been seen by the user in a previous session. In addition, each extract highlights the query words (each in a different colour) and shows how often the stemmed query words that are not in the stop-list of the retriever were found in the whole document. As can be seen in this example, the words "me, up, to, on, the" were removed from the original query because they belong to the stop-list and are therefore considered to be non-content-words. For each extract the results page also shows when the corresponding audio was broadcast and by which station. In Fig. 2 the document that is represented by the extract was broadcast during National Public Radio's "All Things Considered" on May 24, 2000. The total duration of the story is four minutes and seven seconds, whereas the extract is only around 30 seconds long. Next to the information about the origin of the broadcast is a line of green dots that indicate the retriever score of the document relative to the highest retriever score in the current search. If the documents are sorted according to the highest retriever score first criterion, all the dots of the first document are always green. The red "Relevant?" button at the end of this line is a toggle button that allows the user to mark the document as relevant if indeed it was found to answer the user's information need. Clicking on this button changes it to a green "Relevant!" button. The use of this button will be explained in

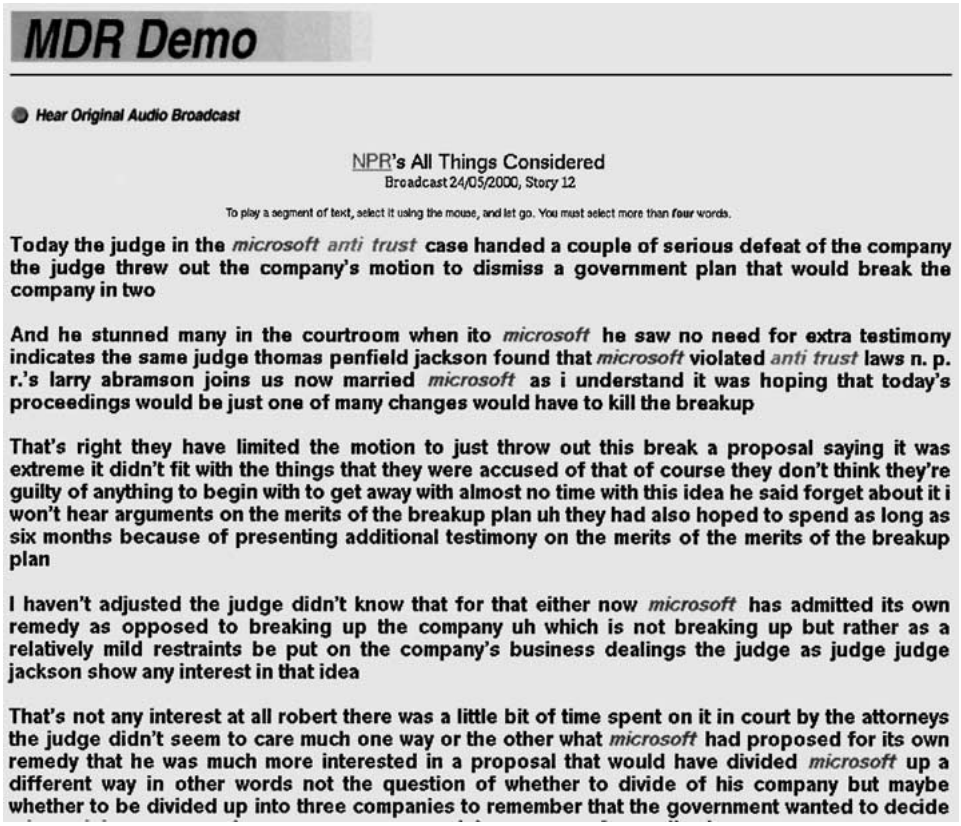

Figure 3. Web page showing the complete transcription of the third most relevant document for query "Keep me up to date on the Microsoft anti-trust suit".

Section 6.3. The user can listen to the part of the sound source that corresponds to the extract. The whole automatic transcript can be accessed on a separate web page (see Fig. 3) where the user can listen to selected parts of the transcription by highlighting them. The whole list of extracts can be sorted using different criteria, e.g., highest relevance score first, most recent first.

6.3. Query Expansion

The MDR demo system supports two forms of query expansion, i.e., relevance feedback and semantic posets. Both methods can be activated independently by clicking on the respective radio buttons next to the "Expand Query" button at the top of the results page.

6.3.1. Relevance Feedback. It was already pointed out in Section 6.2 that the user can mark documents that were found by the retriever as relevant by clicking on the "Relevant?/Relevant!" toggle button next to the document extract. This enables the user to state whether a document that had a high retriever score is indeed relevant from the user's point of view. If the user chooses to expand the query based on the documents that were marked as relevant, a list of terms from these documents is created that contains some terms that are specific to these documents and distinguishes them from the ones that were not marked as relevant. Here, the terms which are stemmed words from the recogniser vocabulary are selected according to their Offer weight (Spärck Jones et al., 1998). An example of such a list of terms can be found in Fig. 4. As one can see, most of the words that are associated with the terms in this list (judge, Jackson, browser, . . .) are related to the Microsoft anti-trust case. On this page the terms that best correspond to the user's interests can be selected and the search repeated with the selected terms appended to the original query. This feature also can be helpful in finding documents in which certain

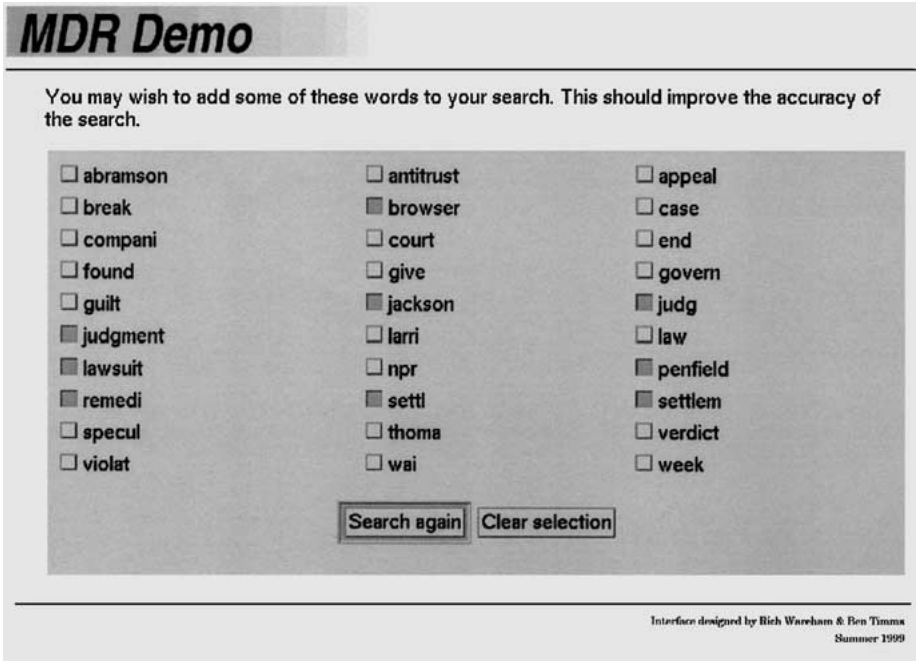

Figure 4. List of suggested terms based on a set of documents that has been marked as relevant for the query "Keep me up to date on the Microsoft anti-trust suit".

content words have been systematically misrecognised. If this is the case, a word is likely to have been replaced by one of a small number of phonetically similar words. This situation can occur if a word is not in the recogniser vocabulary. For example, when looking for stories about Elian Gonzales's return to Cuba, it was found that the word "Elian" was not in the recogniser vocabulary, and had been frequently misrecognised as "alien". This word was offered in the query expansion stage along with other spellings of his name such as "Gonzalez", which enabled the user to recover from these transcription errors.

6.3.2. Semantic Posets. With the help of semantic posets (Jourlin et al., 2000) the user can expand a query directly from the query words. In this case some prior information about which words are topically related is exploited. For instance, for the query "Tell me about the elections in Zimbabwe" the demo system would suggest "Harara, Rhodesia" as words to be added to the original query. The posets for geographical place names were taken from a travel WWW server, and unambiguous nouns were derived from WordNet (Fellbaum, 1998).

7. Informal System Evaluation

So far, the CU-MDR demo system has been shown to many visitors to Cambridge University and at two conferences (Tuerk et al., 2000a, b). Although we have not carried out a rigorous evaluation of the performance of the demo system, the large number of user comments gives us a good indication as to which aspects of the system users liked and what they felt should be improved.

As long as the queries related to either world news or major U.S. or British news, the CU-MDR demo system was able to present documents to the user that contained relevant information even without performing query expansion. Typical queries in this category were "Tell me about the Middle East peace process" and "What's going on in the U.S. baseball league". The query expansion was only found to be helpful in the cases where the user wanted to narrow the search, e.g., if they were interested specifically in the latest news from the Golan Heights instead of the whole Middle East peace process. For queries where database coverage was patchy users found query expansion to be useful and sometimes necessary. In general, relevance feedback seemed to be the preferred method of query

expansion, since it attempted to infer what the user's interest was, based on the documents that the user had earlier marked as relevant. Query expansion based on posets was found to be useful when the user wanted to add geographic information to the query.

Although high recognition accuracy is not crucial for the purpose of maximising a standard IR performance metric (Garofolo et al., 2000), users found it important that the recognition accuracy was relatively high, when interacting with the system. This is due to the fact that text can be downloaded and browsed much faster than audio. A good automatic transcription can therefore reduce the time a user has to spend going through a list of spoken documents to find all the information relevant to a query. The importance of high accuracy transcription in a practical application illustrates the fact that the expectations of an actual user may go beyond a good system performance according to a standard IR performance metric. This issue has, for instance, been addressed in Spärck Jones (2001). Due to the high quality of large parts of the transcriptions in the CU-MDR system, users also found it helpful to be able to select (by highlighting the corresponding text) for playback only the sections of the audio where the transcription was thought to be inaccurate.

The relatively small number of names in the dictionary meant that, for some of the queries about persons or places that appeared in the news only recently, few or no relevant documents were found. It was sometimes possible to recover from this problem by doing relevance feedback and choosing acoustically similar words to the out-of-vocabulary query word. However, this method is not very reliable and might add more noise to the query. It is therefore desirable to have a mechanism to update the recogniser dictionary and language model on a regular basis to include new names in the automatic transcriptions.

8. Conclusions

This paper described the components of the CU-MDR demo system and showed how the user can interact with the features of the web-based user interface to find documents of interest. Future work will involve the extension of the CU-MDR demo database to sources other than NPR, PBS and BBC and the development of a mechanism to update the vocabulary and language model of our recogniser automatically.

Acknowledgments

This work is in part funded by an EPSRC grant reference GR/L49611. We would also like to thank Ben Timms and Richard Wareham for their substantial contribution to the demo interface. Finally our thanks go to the THISL project for making the transcribed BBC acoustic training data available to us.

Notes

1. see http://www.realnetworks.com.
2. see http://www.npr.org.
3. see http://www.pbs.org.
4. see http://www.bbc.co.uk/worldservice/index_stat.shtml.
5. see http://www.realnetworks.com.
6. measured on a 550 MHz Pentium III.

References

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Databse*. Cambridge, MA: MIT Press.

Garofolo, J., Auzanne, C., and Voorhees, E. (2000). The TREC spoken document retrieval track: A success story. *Proc. RIAO*, vol. 1, p. 1–20.

Johnson, S.E., Jourlin, P., Spärck Jones, K., and Woodland, P.C. (2000a). Audio indexing and retrieval of complete broadcast news shows. *Proc. RIAO*, Paris, France.

Johnson, S.E., Jourlin, P., Spärck Jones, K., and Woodland, P.C. (2000b). Spoken document retrieval for TREC-8 at Cambridge University, *Proc. TREC-8*, NIST Gaithersburg, MD.

Johnson, S.E. and Woodland, P.C. (1998). Speaker clustering using direct maximisation of the MLLR-adapted likelihood. *Proc. ICSLP*, pp. 1775–1779.

Jourlin, P., Johnson, S., Spärck Jones, K., and Woodland, P. (2000). Spoken document representations for probabilistic retrieval. *Speech Communication*, *32*(12):21–36.

Leggetter, C.J. and Woodland, P.C. (1995). Maximum likelilhood linear regression for speaker adaptation of continuous density HMMs. *Computer Speech and Language*, *9*:171–186.

Odell, J.J., Woodland, P.C., and Hain, T. (1999). The CUHTK entropic 10 × RT broadcast news transcription system. *Proc. DARPA Broadcast News Workshop*, Herndon, VA, pp. 271–275.

Pallett, D.S., Fiscus, J.G., Garofolo, J.S., Martin, A., and Przybocki, M. (1999). 1998 Broadcast news benchmark test results: English and non-English word error rate performance measures. *Proc. DARPA Broadcast News Workshop*, Herndon, Virginia.

Pallett, D.S., Fiscus, J.G., Martin, A., and Przybocki, A. (1998). 1997 Broadcast news benchmark test results: English and non-English. *Proc. DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, Virginia, pp. 5–11.

Porter, M.F. (1980). An algorithm for suffix stripping. *Program*, *14*:130–137.

Renals, S., Abberley, D., Robinson, T., Kirby, D., and Marks, M. (2000). The THISL Broadcast news retrieval system. *Proc. RIAO*, vol. 3, pp. 39–40.

Robertson, S.E. and Spärck Jones, K. (1997). Simple, proven approaches to test retrieval (Technical Report TR356). Cambridge, UK: Cambridge University Computer Laboratory.

Spärck Jones, K. (2001). Automatic language and information processing: Rethinking evaluation. *Natural Language Engineering*, *7*:1–18.

Spärck Jones, K., Walker, S., and Robertson, S.E. (1998). A probabilistic model of information retrieval: Development and status (Technical Report TR446). Cambridge, UK: Cambridge University Computer Laboratory.

Tuerk, A., Johnson, S.E., Jourlin, P., Spärck Jones, K., and Woodland, P.C. (2000a). The Cambridge University multimedia document retrieval demo system. *Proc. RIAO*, vol. 3, pp. 14–15.

Tuerk, A., Johnson, S.E., Jourlin, P., Spärck Jones, K., and Woodland, P.C. (2000b). The Cambridge University multimedia document retrieval demo system. *Proc. SIGIR*, p. 111.

Van Thong, J., Goddeau, D., Litvinova, A., Logan, B., Moreno, P., and Swain, M. (2000). SpeechBot: A speech recognition based audio indexing system for the web. *Proc. RIAO*, vol. 1, pp. 106–115.

Witbrock, M. and Hauptmann, A. (1997). Speech recognition and information retrieval: Experiments in retrieving spoken documents. *Proc. DARPA Speech Recognition Workshop*, Chantilly, Virginia.