A Just-in-Time Document Retrieval System for Dialogues or Monologues

Andrei Popescu-Belis, Majid Yazdani, Alexandre Nanchen, and Philip N. Garner
Idiap Research Institute
Rue Marconi 19, Case Postale 592
1920 Martigny, Switzerland
{apbelis,myazdani,ananchen,pgarner}@idiap.ch

Abstract

The Automatic Content Linking Device is a just-in-time document retrieval system that monitors an ongoing dialogue or monologue and enriches it with potentially related documents from local repositories or from the Web. The documents are found using queries that are built from the dialogue words, obtained through automatic speech recognition. Results are displayed in real time to the dialogue participants, or to people watching a recorded dialogue or a talk. The system can be demonstrated in both settings.

1 Introduction

The Automatic Content Linking Device (ACLD) is a system that analyzes speech input from one or more speakers using automatic speech recognition (ASR), in order to retrieve related content, in real time, from a variety of repositories. This paper describes the main components of the system and summarizes evaluation results. The remainder of this section introduces scenarios of use and previous systems with similar goals.

The first scenario of use involves people taking part in meetings, who often mention documents containing facts that are relevant to the current discussion, but cannot search for them without interrupting the discussion flow. Our goal is to perform such searches automatically. In a second scenario, search is performed for live or recorded lectures, for instance in a computer-assisted learning environment. The ACLD enriches the lectures with related course material, receiving real-time feedback from the user.

The ACLD improves over past systems by using speech, by giving access to multimedia documents, and by using semantic search. Its first precursors were the Fixit query-free search system (Hart and Graham, 1997), the Remembrance Agent for just-in-time retrieval (Rhodes and Maes, 2000), and the Implicit Queries system (Dumais et al., 2004). A version of the Remembrance Agent called Jimminy was conceived as a wearable assistant for taking notes, but ASR was only simulated (Rhodes, 1997). Watson monitored the user's operations in a text editor, and selected terms for web search (Budzik and Hammond, 2000). Another authoring assistant was developed in the A-Propos project (Puerta Melguizo and al., 2008). Recently, several speech-based search engines have been proposed, as well as systems for searching spoken documents. For human dialogues in meetings, the FAME interactive space (Metze and al., 2006) provided multi-modal access to recordings of lectures via a table top interface, but required specific voice commands from one user only, and did not spontaneously follow a conversation as the ACLD does.

2 Description of the ACLD

The architecture of the ACLD comprises modules for: (1) document preparation and indexing; (2) input sensing and query construction; (3) search and integration of results; (4) user interaction.

2.1 Document Preparation and Indexing

The preparation of the local database of documents available for search requires text extraction from various file formats (like MS Office or PDF), and

document indexing, here using Apache Lucene. Past meetings, when available, are automatically transcribed, then chunked into smaller units, and indexed along with the other documents. For searching the Web, the system does not build indexes but uses the Google Search API.

2.2 Sensing the User's Information Needs

The ACLD uses the AMI real-time ASR system for English (Garner and al., 2009), which has an acceptable accuracy for use with conversational speech in the ACLD. When processing past recordings, the ASR system can run slower than real-time to maximize its accuracy. If one or more pre-specified keywords (based on domain knowledge) are detected in the ASR output, then their importance is increased for searching. Otherwise, all the words from the ASR (except stopwords) are used for constructing the query.

2.3 Querying the Document Database

The Query Aggregator component uses the ASR words in order to retrieve the most relevant documents from a given database. The latest version of the ACLD makes use of semantic search (see below), but earlier versions used keyword-based search from Apache Lucene for local documents. Queries are formulated and launched at regular time intervals, typically every 15-30 seconds, or on demand. The search results are integrated with previous ones, using a persistence model that smoothes variations in time by keeping track of the salience of each result. Salience is initialized from the ranking of search results, then decreases in time, or increases if the document appears again among results. A history of all results is also accessible.

2.4 Semantic Search over Wikipedia

The goal of semantic search is to improve the relevance of results with respect to the spoken words, and to make search more robust to noise from ASR. The method used here is adapted from a graph-based measure of semantic relatedness between text fragments (Yazdani and Popescu-Belis, 2010). Relatedness is computed using random walk in a large network of documents, here about 1.2 milion Wikipedia articles from the WEX data set (Metaweb Technologies, 2010). These are linked by directional hyperlinks, and also by lexical similarity links that we construct upon initialization. The random walk model allows the computation of the visiting probability (VP) from one document to another, and then of the VP between sets of documents. This functions as a measure of semantic relatedness, and has been applied to several NLP problems by projecting the text fragments to be compared onto the documents in the network (Yazdani and Popescu-Belis, 2010).

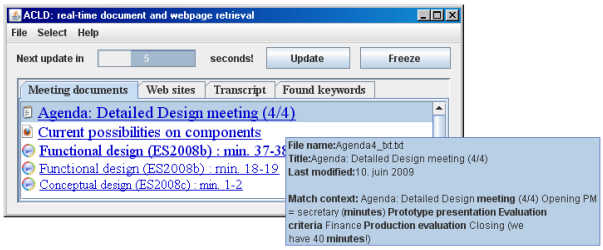

Figure 1: Unobtrusive UI of the ACLD displaying document results. The pop-up window shows more details for the first results.

For the ACLD, the use of semantic relatedness for document retrieval amounts to searching, in a very large collection, the documents that are the most closely related to the words obtained from the ASR in a given time frame. Here, we set the document collection to Wikipedia (WEX). As the search is hard to perform in real time, we made a series of justified approximations to make it tractable.

2.5 The User Interface

The goal of the UI is to make ACLD information available in a configurable way, allowing users to see more or less information according to their own needs. The UI displays up to four widgets, which can be arranged at will, and contain: (1) ASR words with highlighted keywords; (2) tag-cloud of keywords, coding for recency and frequency; (3) links to the current results from the local repository; (4) links to the current Web search results.

Two main arrangements are intended: an informative full-screen UI (not shown here from lack of space) and an unobtrusive UI, with superposed tabs, shown in Figure 1 with the document result widget. When hovering over a document name, a pop-up window displays metadata and document excerpts that match words from the query, as an explanation for why the document was retrieved.

3 Evaluation of the ACLD

Four types of evidence for the relevance and utility of the ACLD are summarized here. Firstly, the ACLD was demonstrated to about 50 potential users (industrial partners, focus groups, etc.), who found the concept useful, and offered positive verbal evaluation, along with suggestions for smaller and larger improvements.

Secondly, a pilot experiment was conducted with a group using an earlier version of the UI. Two pilot runs have shown that the ACLD was consulted about five times per meeting, but many more runs are (still) needed for statistical significance of observations.

Thirdly, the UI was tested in a usability evaluation experiment with nine non-technical subjects, who rated it as 'acceptable' (68%) on the System Usability Scale, following a series of tasks they had to perform using it. Additional suggestions for changes were received.

Finally, we compared offline the results of semantic search with the keyword-based ones. We asked eight subjects to read a series of nine meeting fragments, and to decide which of the two results was the most useful one (they could also answer 'none'). Of a total of 36 snippets, each seen by two subjects, there was agreement on 23 (64%) snippets and disagreement on 13 (36%). In fact, if 'none' is excluded, there were only 7 true disagreements. Over the 23 snippets on which the subjects agreed, the result of semantic search was judged more relevant than that of keyword search for 19 (53% of the total), and the reverse for 4 only (11%). Alternatively, if one counts the votes cast by subjects in favor of each system, regardless of agreement, then semantic search received 72% of the votes and keyword-based only 28%. Hence, semantic search already outperforms keyword based one.

4 Conclusion

The ACLD is, to the best of our knowledge, the first just-in-time retrieval system to use spontaneous speech and to support access to multimedia documents and to websites, using a robust semantic search method. Future work should aim at improving the relevance of semantic search, at modeling context to improve the timing of results, and at inferring relevance feedback from users. The ACLD should also be applied to specific use cases, and an experiment with group discussions in a learning environment is under way.

Acknowledgments

We are grateful to the EU AMI and AMIDA Integrated Projects and to the Swiss IM2 NCCR (Interactive Multimodal Information Management) for supporting the development of the ACLD.

References

Jay Budzik and Kristian J. Hammond. 2000. User interactions with everyday applications as context for just-in-time information access. In *IUI 2000 (5th International Conference on Intelligent User Interfaces)*, New Orleans, LA.

Susan Dumais, Edward Cutrell, Raman Sarin, and Eric Horvitz. 2004. Implicit Queries (IQ) for contextualized search. In *SIGIR 2004 (27th Annual ACM SIGIR Conference) Demonstrations*, page 534, Sheffield.

Philip N. Garner and al. 2009. Real-time ASR from meetings. In *Interspeech 2009 (10th Annual Conference of the International Speech Communication Association)*, pages 2119–2122, Brighton.

Peter E. Hart and Jamey Graham. 1997. Query-free information retrieval. *IEEE Expert: Intelligent Systems and Their Applications*, 12(5):32–37.

Metaweb Technologies. 2010. Freebase Wikipedia Extraction (WEX). http://download.freebase.com/wex/.

Florian Metze and al. 2006. The 'Fame' interactive space. In *Machine Learning for Multimodal Interaction II*, LNCS 3869, pages 126–137. Springer, Berlin.

Maria Carmen Puerta Melguizo and al. 2008. A personalized recommender system for writing in the Internet age. In *LREC 2008 Workshop on NLP Resources, Algorithms, and Tools for Authoring Aids*, pages 21–26, Marrakech.

Bradley J. Rhodes and Pattie Maes. 2000. Just-in-time information retrieval agents. *IBM Systems Journal*, 39(3-4):685–704.

Bradley J. Rhodes. 1997. The Wearable Remembrance Agent: A system for augmented memory. *Personal Technologies: Special Issue on Wearable Computing*, 1:218–224.

Majid Yazdani and Andrei Popescu-Belis. 2010. A random walk framework to compute textual semantic similarity: a unified model for three benchmark tasks. In *ICSC 2010 (4th IEEE International Conference on Semantic Computing)*, pages 424–429, Pittsburgh, PA.