Language Technologies: Question Answering in Speech Transcripts

Jordi Turmo\textsuperscript{1}, Mihai Surdeanu\textsuperscript{1}, Olivier Galibert\textsuperscript{2}, Sophie Rosset\textsuperscript{2}

\textsuperscript{1} TALP Research Center, Universitat Politècnica de Catalunya, Barcelona, Spain
\textsuperscript{2} LIMSI-CNRS, Orsay, France

The question answering (QA) task consists of providing short, relevant answers to natural language questions. Most QA research has focused on extracting information from text sources, providing the shortest relevant text in response to a question. For example, the correct answer to the question, “How many groups participate in the CHIL project?” is “15”, whereas the response to “Who are the partners in CHIL?” is a list of them. This simple example illustrates the two main advantages of QA over current search engines: First, the input is a natural-language question rather a keyword query; and second, the answer provides the desired information content and not simply a potentially large set of documents or URLs that the user must plow through.

One of the aims of the CHIL project was to provide information about what has been said during interactive seminars. Since the information must be located in speech data, the QA systems had to be able to deal with transcripts (manual or automatic) of spontaneous speech. This is a departure from much of the QA research carried out by natural-language groups, who have typically developed techniques for written texts that are assumed to have a correct syntactic and semantic structure. The structure of spoken language is different from that of written language, and some of the anchor points used in processing such as punctuation must be inferred and are therefore error-prone. Other spoken-language phenomena include disfluencies, repetitions, restarts, and corrections. If automatic processing is used to create the speech transcripts, an additional challenge is dealing with the recognition errors. The response can be a short string, as in text-based QA, or an audio segment containing the response.

This chapter summarizes the CHIL efforts devoted to QA for spoken language carried out at UPC and at CNRS-LIMSI. Research at UPC was directed at adapting a QA system developed for written texts to manually and automatically create speech transcripts, whereas at LIMSI an interactive oral QA system developed for the French language was adapted to the English language. CHIL organized the pilot track on question answering in speech transcripts (QAST), as part of CLEF 2007, in order to evaluate and compare QA technology on both manually and automatically produced transcripts of spontaneous speech.
8.1 Question Answering

Two main paradigms are used to search for information: document retrieval and precise information retrieval. In the first approach, documents matching a user query are returned. The match is often based on some keywords that were extracted from a query, and the underlying assumption is that the documents best matching the topic of the query provide a data pool from which the user might find information that suits their need. This need can be very specific (e.g., “Who is presiding the Senate?”), or it can be topic-oriented (e.g., “I’d like information about the Senate”). The user is left to filter through the returned documents to find the desired information, which is quite appropriate for general topic-oriented questions, and less well-adapted to more specific queries.

The second approach to search, which is better suited to the specific queries, is embodied by so-called question answering systems, which return the most probable answer given a specific question (e.g., the answer to “Who won the 2005 Tour de France?” is “Lance Armstrong”).

In the QA and information retrieval domains, progress has been assessed via evaluation campaigns [1, 10, 18, 13, 5]. In the question answering evaluations, the systems handle independent questions and should provide one answer to each question, extracted from textual data, for both open or restricted domains.

Recently, there has been growing interest in extracting information from multimedia data such as meetings, lectures, etc. Spoken data are different from textual data in various ways. The grammatical structure of spontaneous speech is quite different from written discourse and includes various types of disfluencies. The lecture and interactive meeting data of interest to CHIL are particularly difficult due to run-on sentences and interruptions.

Figure 8.1 shows an example of a standard architecture for a QA system. Typical QA systems (mostly textual) are composed of question processing, document processing, document or passage retrieval, and answer extraction components [17, 9].