

Building a Multi-Lingual Interactive Question-Answering System for the Library Domain

Manuel Kirschner

KRDB, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy
kirschner@inf.unibz.it

1 Introduction

Lately, there has been increasing interest in how to best enrich Question-Answering (QA) applications with dialogue capabilities¹. While classical QA is concerned with questions posed in isolation, its interactive variant keeps track of the QA process and supports the user in finding the exact solution via natural-language dialogue. The context of each utterance must be considered for handling clarification sub-dialogues and to resolve anaphora, ellipses or fragmentary utterances.

We have started to develop an Interactive QA system for the university library domain. We see it as a fusion of the QA scenario with robust dialogue systems techniques. From the QA point of view, the system's task is to retrieve answers to user questions using a knowledge base, where the user holds the initiative throughout the exchange. Research in discourse structure and dialogue management can provide us with models for properly dealing with co-reference, keeping track of the current topic, and even temporally switching to system initiative when the user is lost. We adopt a bottom-up approach, starting with data collection, and building a simple baseline system. As more data become available, we plan to iteratively add support for more sophisticated discourse phenomena.

2 Library domain

The University library is striving to extend services that help their users find information over the web site in their preferred language. To this end, an interactive QA system will provide permanent and instant multi-lingual access to library-specific information. As the experiences of other libraries have shown, these systems surpass FAQ lists by guiding the user towards answers when initially she does not know how to formulate (and search for) the exact question.

¹E.g., Workshop on Interactive Question Answering (IQA'06), in HLT-NAACL'06.

In cooperation with library staff, we drew up a specification of the scope of our conversational agent, i.e., the topics for possible human-computer dialogues. Our librarians analyzed archived e-mail with past user queries and compiled a typology of frequent questions. Existing FAQ lists from the library web-site were also incorporated. The resulting library domain is structured into topics and sub-topics as follows²:

General: library buildings, organization, services (e.g., opening hours, library card, technical infrastructure, library web-site), FAQ
Inquiry: catalog query, general, books, journals, articles/theses, literature topics
Lending: lending, inter-library loan

3 Wizard-of-Oz study

Under the bottom-up paradigm described in the introduction, we first need conversation data, gained through conducting a Wizard-of-Oz (WoZ) study with actual library users. The study has two goals:

1. Verify topic scope of knowledge base
2. Explore dialogue-specific discourse phenomena present in WoZ data

We designed a simple chat-like interface that was accessible to all users on the university intranet via a link from the main library web-site. The system establishes a text-based conversation between users and library staff, while the users are made believe that they are actually talking to an automatic dialogue system. The users were given only a rough specification of possible query topics as a guideline, but were not provided with a task description, as is usually the case in WoZ studies.

In a first run, the WoZ system was online during office hours for two weeks. We collected an overall of 64 dialogues, with an average of 6 turns per

²As a side task, the librarians are annotating electronically available library documents with the above topic structure. The marked-up documents will prove useful for providing users with additional source evidence, and for later Information Retrieval experiments.

dialogue. We plan to conduct more WoZ runs in the near future.

Topic scope Analyzing the coverage of our knowledge base with regard to the collected WoZ dialogues, roughly two thirds are within the scope of library topics we wanted to cover (see box). The remaining logs exhibited out-of-scope user queries, about half being nonsense words, the other half typical “small talk” dialogue. Thus, if only to keep the conversation going and to allow for a playful exploration of the task domain, the dialogue system should exhibit some small talk skills as well.

Observed discourse phenomena Of the initially collected 64 dialogues, 12 contain some kind of discourse phenomenon. This ratio (18,75%) is on the lower end of the spectrum reported in the literature, where information-seeking tasks in particular have been shown to contain the lowest share of context-dependent turns (Bertomeu et al., 2006). Also, in that study, participants were given a complex task: asking for additional information about previously retrieved entities from a database. In our view, such differences in system and user tasks can explain the lower number of discourse phenomena in our data. We annotated these data using the classification scheme for fragments/referential devices proposed in Bertomeu et al. (2006). In our study, the three most frequent discourse phenomena were anaphoric pronouns³ (11 times), elided NPs⁴ (10) and fragments⁵ (5).

4 Implementation status and next steps

Having implemented the WoZ application, our next goal is to build a baseline dialogue system. We have developed software tools that enable our librarians to construct a hierarchical knowledge base of library topics in XML (with main topics as listed in section 2). For each topic (and sub-topic), this hierarchy contains keywords for mapping a user query to an appropriate canned-text answer or to a clarification request (if the query does

³User: “wie funktioniert Fernleihe? [how does inter-library loan work?]” – Wizard: “(…)” – User: “und wieviel kostet es? [and how much does it cost?]”

⁴U: “Nur die Diplomarbeiten der Informatik-Fakultät? Nicht auch ∅ von der Wirtschaftsfakultät? [Only the masters theses from the faculty of computer science? Not also ∅ from the faculty of Economics?]”

⁵W: “Zum Kopieren verwendet man die Student Card. [For making photocopies use your student card.]” – U: “zum Drucken? [for printing?]”

not contain all the keywords required to traverse the hierarchy until reaching a terminal topic).

The baseline system will cover a sub-set of the discourse phenomena encountered in our current WoZ data. The underlying notion is that in our hierarchy, the topic that matches a user query represents the current conversational context. Follow-up questions as in footnotes 3 and 5 can lead to a correct answer if the identity of the previous topic is kept as the current dialogue state. A follow-up question is then matched against the keywords stored under that specific topic. For the example of footnote 3, the active topic just before the follow-up question is *inter-library loan*; then, provided that *cost* is defined as a sub-topic, the follow-up question (containing “cost”) matches a local keyword, yielding the desired topic-related answer.

Because of the dual role of our topic hierarchy—knowledge base and data structure for keeping dialogue state—we see it as a simplified version of the *topic structure* described in Stede and Schlangen (2004). At the time of writing, library staff have started filling the knowledge base with topics, keywords and answers. By analyzing log-files gained from the baseline system (i.e., checking where user input failed to match (correct) patterns in the knowledge base), librarians will be able to iteratively improve the coverage by adding new keywords or new topics.

We plan to support the full regular expression syntax for defining keywords, and to let the librarians define procedural sub-dialogues to complement the baseline keyword search algorithm for the topic hierarchy⁶. Through the analysis of log files, we will study how users keep/change topic focus, and how mixed-initiative dialogue can help the user in finding the answer to a complex query.

References

Núria Bertomeu, Hans Uszkoreit, Anette Frank, Hans-Ulrich Krieger, and Brigitte Jörg. 2006. Contextual phenomena and thematic relations in database QA dialogues: results from a Wizard-of-Oz Experiment. *Proc. of HLT-NAACL'06*, New York City, NY.

Manfred Stede and David Schlangen. 2004. Information-Seeking Chat: Dialogues Driven by Topic Structure. *Proc. of Catalog'04 (Semdial'04)*, Barcelona, Spain.

⁶These features were deemed important in “Stella” (<http://www.sub.uni-hamburg.de/>), a proprietary German dialogue system built for the Hamburg University library (Anne Christensen, pers. comm.).