Can I also have a question, please? On Perspectives of Interactive Clustering-Driven Question-Answering

Marcin SKOWRON and Kenji ARAKI
Graduate School of Information Science and Technology, Hokkaido University

1 Introduction

Question Answering (QA) is an active research area that gained an interest of many researchers, including but not limited only to the TREC community[1]. The systems that participate in the TREC QA track are designed to provide answers to factoid questions, from 2 gigabytes of text, collection that includes mostly newspaper or newswire articles[3]. Most of the systems that participate in the TREC evaluation, uses a variety of linguistic resources and employs sophisticated text processing techniques, which help to retrieve a sufficient set of answer candidates by resolving anaphor and synonyms, uncovering complex lexical, syntactic and semantic relations between a question and answer candidates. Since the beginning of TREC evaluation one can observe a constant improvement of the systems performance, maturing of commonly used methods and increasing sophistication of used tools and linguistic resources.

Users expectations concerning a naturalness, reliability and intuitiveness of access to information as well as ever-expanding search space of various types of resources, including dynamically growing textual documents influence the environment and working conditions for wide range of tools such as databases querying systems, Internet search engines, QA-Systems etc. The amount of text available online has been rapidly increasing; in 2005, a popular search engine indexed 19.2 billion web pages[2] and their numbers continue to grow. In [1], Brill demonstrated that web redundancy can be effectively used in QA-Systems to retrieve an answer rich set of passages, reducing the overall complexity of the system and its dependency on external tools and linguistic resources. We think that a similar effect can be achieved by redefining Question-Answering by introducing the means for a system-user interaction. As we demonstrate below, such interaction while beneficial for a system by providing the means to adjust the search space and select an answer rich set of documents closely related to a question addresses also some of users expectations concerning naturalness and intuitiveness of accessing sought-after information. This approach aims to reassemble a part of a conversation between two human beings, when a person who is supposed to answer a given question, asks for more details to eliminate potential ambiguity of a question or to specify the proper answer scope expected by other person.

We think that research on interactive QA is also valuable for the development of dialog and chat systems. So far, most of these systems employs various discussion strategies to avoid a direct answer to a interlocutor questions, and by all means try to continue a chat with a user. In many situations a meaningful answer could be relatively easily provided, thus making such systems more useable, enjoyable for a user, and in case of Turing-like tests, extend the illusion of conversation with an intelligent human-interlocutor.

2 System Processing and Preliminary Experiments

The flowchart of the System that implements Interactive Clustering-Driven Question-Answering is presented on Figure 1.

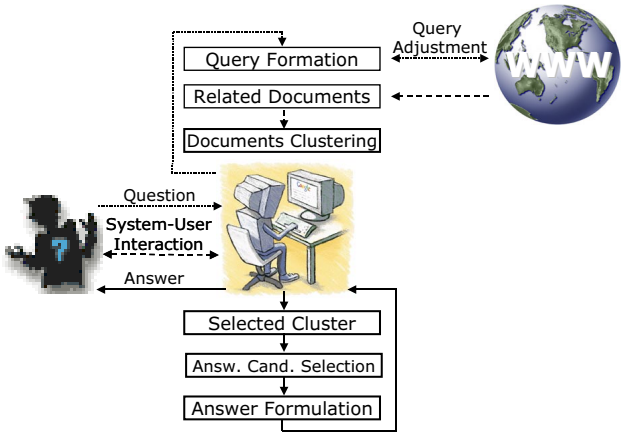

Figure 1 System Flowchart

The system is initiated by receiving an input in a form of a question that is processed with the aim to extract useful keywords to form an initial query. For this task our system uses a simple Perl module Lingua::EN::Keywords, which was slightly modified by providing the additional list of stop-words, and by limiting the maximum number of words composing a phrase (a sequence of words used as an exact match when querying a search engine, by embracing it with the quotation marks) to two words.

In the next step, the system uses information from the search engine (in our experiments we used Yahoo) on the number of pages that contain words and phrases used in a query to gradually adjust a query to obtain an answer-rich set of documents. For this purpose we set two thresholds. The minimal number of pages that needs to be retrieved was set to 30. If less pages existed, the system gradually removed words from a query, starting from the most general ones selected using frequency information obtained from the BNC Corpus[3]. The maximum number of pages, was set to 100,000. If more pages existed the system gradually added words from the question that were not used to form a query before, starting with the ones which were the most specific (again information on the frequency of words occurrence from the BNC Corpus was used).

To provide a wide coverage of various domains our system aimed to retrieve 100 documents related to a given query for each of five domains (.com,.edu,.net,.org,.net). In order to initially evalu-

[1] http://trec.nist.gov/data/qa.html
[2] http://www.yahoo.com, as of August 2005.
[3] http://www.natcorp.ox.ac.uk/

ate our system, we decided to use only snippets of pages retrieved by the search engine. Although, this decision resulted in retrieving significantly smaller size documents, and reduced the number of answer candidates we found such settings sufficient to perform the preliminary evaluation of the proposed method and to compare the system with a generic QA-System.

In the next step the system uses Cluto Toolkit[2] to cluster the retrieved set of snippets, and to extract the most descriptive words for each of created clusters. The initially number of clusters is set to four. The system uses the extracted words to form questions in a natural language form that are used to interact with a user to select a cluster, most relevant to a user question. Our assumption is that such cluster usually provides an answer rich set of documents, with the highest proportion of correct answers to other answer candidates. For example, for the question "What are the names of the space shuttles?" the system generated the following clusters, which can be represented using the 3 most descriptive words extracted from each of them:

- book, article, encyclopedia
- Earth, STO, edge
- NASA, Atlantis, Columbia
- station, flight, technology

These words were used, to form questions like: "Did you ask, because you are interested in a book or an article?", "You mean the one related to Earth and STO?", "Do you want to know more about NASA and Atlantis?', etc. As we could notice, in some cases a correct answer or a part of it was already present at this step among descriptive words describing clusters and used to form a question to a user. If a user did not find anything related to his/her question, the system generated new set of clusters, by gradually increasing their number and providing more detailed divisions between various topics and domains covered in a retrieved set of snippets. The maximum number of clusters generated by the system was set to 7.

The analysis of the clusters selected by a user revealed that in 41 out of 50 cases (82%) they had higher proportion of correct answers to other answer candidates, compared to the whole set of retrieved documents. This demonstrates that frequently, a user posses more knowledge on a topic he/she is interested in than is expressed in an initial question. The interactive approach to QA, provides the means to use a user intuition, commonsense or partial knowledge on a topic to supply the system with additional information, which cannot be done in a standard approach limited to a single question and answer.

Imposing domain or context-driven constraints on the scope of documents considered for the answer candidates selection process resulted in reduction of the processing time required to analyze, extract and verify large number of answer candidates, as well as increased the chances to chose a correct answer among numerous answer candidates in a statistical-based answer selection process. To measure the proportion of correct answers to all answer candidates retrieved for a given query, we used the Correct Answers Density Ratio (CADR) as showed in formula 1:

$$CADR = \frac{NumberofCorrectAnswers}{NumberofAnswerCandidates} * 100 \quad (1)$$

Table 1 CADR for the Generic and Interactive QA-System.

	CADR
Generic QA System	2.781
Interactive QA System	**5.060**

For the initial evaluation of the systems we used a set of 55 TREC 2004 list questions and answers[4]. The system was able to find at least partial answers to 50 (90%) questions. Table 1 presents the comparison between the Generic and the Interactive QA-System using the CADR measure. The application of the proposed method resulted in an increase of the CADR value of 81,95%.

3 Conclusions and Future Work

The described method demonstrated that by adapting QA-System to a limited interactive mode, a significant performance improvement could be obtained without the need to use highly sophisticated tools or linguistic resources, which are often developed in a costly, laborious and time-consuming process. Additional advantages of this approach compared with the standard QA, included:

- Reduction of a search space; ability to select an answer rich set of documents from often noisy set of passages retrieved with an initial query.
- Existence of the means to deal with the ambiguity of questions expressed in natural language form.
- More natural settings for human-computer interaction.
- Existence of the means to use a user intuition, commonsense or partial knowledge on a topic to supply the system with additional information which leads to simplification of the answer extraction and verification process.

In our future work, we plan to extensively evaluate the described method and adopt it to more sophisticated version of our QA-System, which includes the state-of-art question classification and answer candidates extraction and verification modules. We intend also to perform tests on a group of Internet users to measure the usability, naturalness and intuitiveness of system usage. We think that research on interactive QA-System could be also valuable for the development of dialog systems, by bridging the gap between standard question-answering and chat-systems and by providing a synergy between these two research areas.

Acknowledgement This research was partially supported by Grant from Ministry of Internal Affairs and Communications, Strategic Information and Communications R&D Promotion Programme (SCOPE).

References

[1] E. Brill, J. Lin, M. Banko, S. Dumais, and A. Ng. Data-Intensive Question Answering. Proceedings of the 2001 Text REtrieval Conference (TREC 2001), 2001.

[2] G. Karypis. A Clustering Toolkit. Available at http://www.cs.umn.edu/~cluto. 2002.

[3] E.M. Voorhees. Overview of TREC 2004. Proceedings of The Thirteenth Text Retrieval Conference (TREC 2004). 2004.

[4] http://trec.nist.gov/data/qa/2004_qadata/04.list_answers.txt