Deploying Web-based Question Answering System to Local Archive

Zhiping Zheng
Department of Computational Linguistics
Saarland University, Germany
zheng@coli.uni-sb.de

Huiyan Huang, Sven Schmeier
Xtramind Technologies GmbH, Germany
{huang,schmeier}@xtramind.com

Abstract:

AnswerBus (http://www.answerbus.com/, [2,3]) is a Web-based open-domain Question-Answering (QA) system. It successfully uses NLP/IR techniques and reaches very high correct answer rate. Although it is not designed for TREC, it still correctly answers over 70% of TREC-8 questions with Web resources. The question remains whether the techniques for a web-based QA system can be deployed to a local archive.

In experiments with Xtramind (http://www.xtramind.com/) to answer this question we chose part of DUC conference corpus as the local archive. The selected corpus is in general business domain but we treat it as an open domain archive. We tried to keep all techniques used in AnswerBus. These include QA specific dictionary, dynamic named entities extraction, answer candidate ranking system, pseudo-answer detection etc. But the sentence-question matching formula, proved effective in AnswerBus, is not suitable for the QA system on local archive. Hence we developed a new algorithm to judge if a sentence is a possible answer or not.

The first step in deploying AnswerBus to local archive, a local search engine to do the search is needed. We use Seven Tones (http://www.seventones.com/, [4]) search engine but drop the features related to specific domain. The benefits of this search engine include: 1) The indexing is very fast; 2) The index can be partially and logically modified; and 3) It is scalable to a large size.

To evaluate the new system, we refer to the milestones described in [1] and provided questions, which covered 16 Arthur Graesser's questions categories that ranged from easy to very difficult. The test result is very encouraging and the accuracy of Top 1 is 72%.

References:

[1] John Burger et al. Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A). NIST. 2001.

[2] Zhiping Zheng. AnswerBus Question Answering System. *Human Language Technology Conference (HLT 2002)*. San Diego, CA. March 24-27, 2002.

[3] Zhiping Zheng. Developing a Web-based Question Answering System. *The Eleventh World Wide Conference (WWW 2002)*. Honolulu, HI. May 7-11, 2002.

[4] Zhiping Zheng. Seven Tones: Search for Linguistics and Languages. *The 2nd Meeting of the North American Chapter of Association for Computational Linguistics (NAACL 2001)*. Pittsburgh, PA. June 2-7, 2001.

Related materials and links:

http://www.coli.uni-sb.de/~zheng/qa-local/