

Automated Question Answering in Webclopedia – A Demonstration

Ulf Hermjakob, Eduard Hovy and Chin-Yew Lin

USC Information Sciences Institute

4676 Admiralty Way

Marina del Rey, CA 90292-6695

{ulf,hovy,cyl}@isi.edu

Tel: 310-448-8476, 310-448-8731 and 310-448-8711

1. Overview

In this demonstration we present work in our Webclopedia (Hovy et al. 2000, 2001) project on semantics-based answer pinpointing. Through a live interface, users can type in their questions in the query window or select a TREC10 question (Figure 1). The system will then parse the question and return the top 5 answers (Figure 1).

Webclopedia achieved 31% in MRR (mean reciprocal rank) score in TREC9 Q&A track and 45% in TREC10 Q&A track. The architecture includes:

- ❑ **Question parsing:** Using BBN's IdentiFinder (Bikel et al., 1999), the CONTEX parser produces a syntactic-semantic analysis of the question and determines the QA type.
- ❑ **Query formation:** Single- and multi-word units (content words) are extracted from the analysis, and WordNet synsets are used for query expansion. A series of Boolean queries is formed.
- ❑ **IR:** The IR engine MG (Witten et al., 1994) returns the top-ranked N documents.
- ❑ **Selecting and ranking sentences:** For each document, the most promising $K \ll N$ sentences are located and scored using a formula that rewards word and phrase overlap with the question and its expanded query words. Results are ranked.
- ❑ **Parsing segments:** CONTEX parses the top-ranked 300 sentences.
- ❑ **Pinpointing:** Each candidate answer sentence parse tree is matched against the parse of the question. It sometimes also considers the preceding sentence. As a fallback the window method is used.
- ❑ **Ranking of answers:** The candidate answers' scores are computed and the topmost 5 are output as final answers.

Webclopedia classifies desired answers by their semantic type, using the approx. 140 classes. These types include common semantic classes such as PROPER-PERSON, EMAIL-ADDRESS, LOCATION, and PROPER-ORGANIZATION, but also classes particular to QA such as WHY-FAMOUS, YES:NO, and ABBREVIATION-EXPANSION. They have been taxonomized as the Webclopedia QA Typology (Hermjakob et al. 2002).

The system increasingly makes use of syntactic and semantic (world) knowledge to improve the accuracy of its results. We will identify the strength and weakness of our approach during the demonstration with examples.

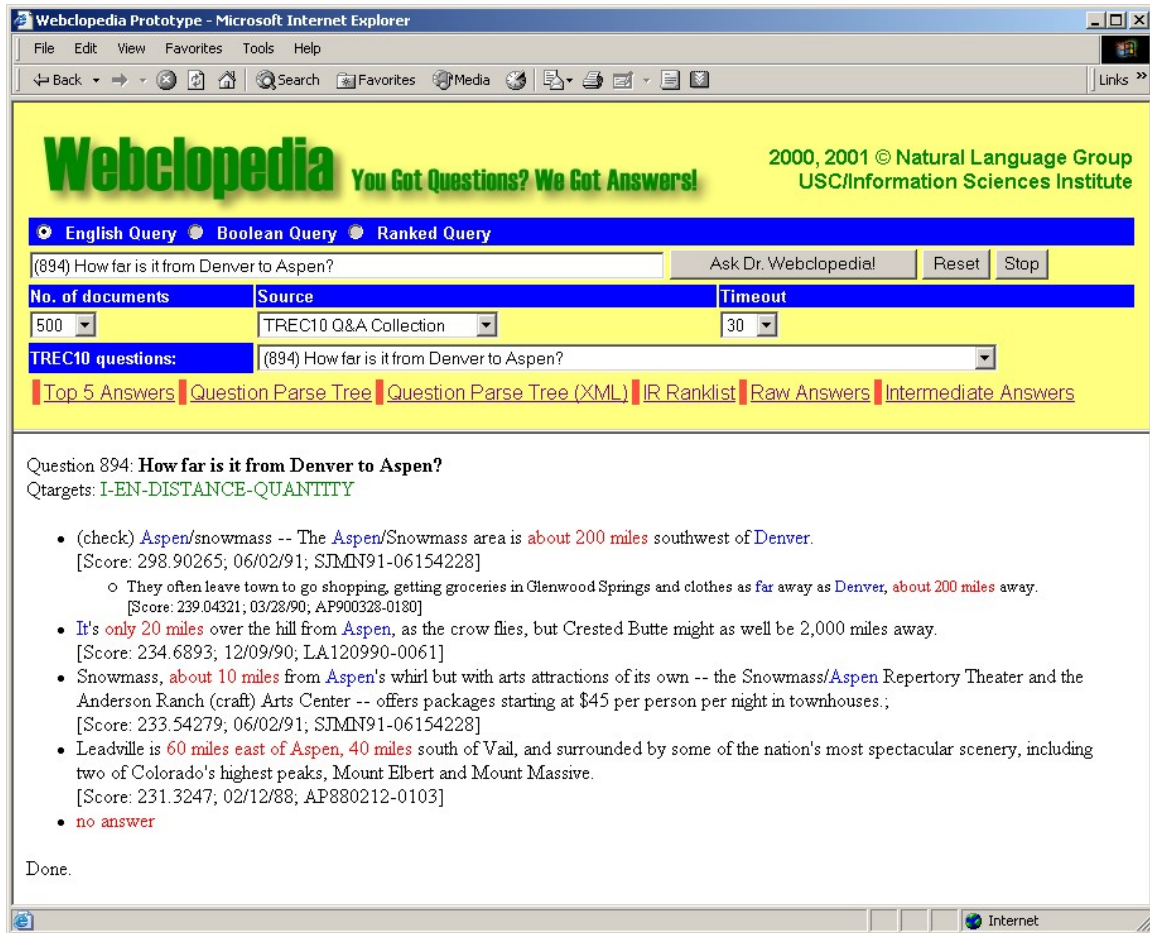


Figure 1. Webclopedia web interface.

2. References

- Hermjakob, U. 1997. *Learning Parse and Translation Decisions from Examples with Rich Context*. Ph.D. dissertation, University of Texas at Austin. file://ftp.cs.utexas.edu/pub/~mooney/papers/hermjakob-dissertation-97.ps.gz.
- Hermjakob, U., E. Hovy and D. Ravichandran 2002. *A Question/Answer Typology with Surface Text Patterns*. Submitted to Human Language Technology Conference 2002.
- Hovy, E.H., L. Gerber, U. Hermjakob, M. Junk, and C.-Y. Lin. 2000. Question Answering in Webclopedia. *Proceedings of the TREC-9 Conference*. NIST, Gaithersburg, MD. November 2000.
- Hovy, E.H., L. Gerber, U. Hermjakob, C.-Y. Lin, and D. Ravichandran. 2001. Toward Semantics-Based Answer Pinpointing. *Proceedings of the DARPA Human Language Technology Conference (HLT)*. San Diego, CA. March 2001.
- Witten, I. H. and A. Moffat and T. Bell 1994. *Managing Gigabytes*. New York: Van Nostrand Reinhold.