Web Fact Seeking Meta Engine for Business Information Needs

Dmitri Roussinov, Arizona State University, Dmitri.Roussinov@asu.edu
Michael Chau, The University of Hong Kong, mchau@business.hku.hk

Based on the results of our surveys and interviews with business intelligence practitioners and success of meta search engines, we advocate a meta approach to automated question answering which does not rely on a single system and provides the necessary accuracy and responsiveness. We present our prototype of a fact seeking meta engine, the first known of this kind, and report its empirical evaluation. Our results support the added value of the meta approach: the performance of the combined system surpasses the underlying performances of its components.

The goal of Question Answering (QA) is to locate, extract, and represent a specific answer to a user question expressed in natural language. A QA system would take as input "CNN is owned by whom?" and it should output a precise answer "Time Warner" and/or a link to a source page that provides the answer. Precise, timely, and factual answers are especially important when the communication channel is limited. A growing number of Internet users are currently accessing it through mobile devices such as cell phones and Personal Digital Assistants (PDA), which do not have the luxury of large "real estate" screen space. Military, first-responder, and security systems frequently put their users under such time constraints that every additional second spent browsing through Google search results may put lives at risk.

By following the example of meta search engines on the Web (Selberg & Etzioni, 1995), *we advocate combining several fact seeking engines into a single "Meta" approach.* The practical benefit of the meta approach to question answering (fact seeking) is obvious. *It does not rely on a single system which may fail or may simply not be designed for a specific type of tasks (questions).* The meta approach promises *higher coverage and recall of the correct answers* since different QA engines may cover different databases or different parts of the Web. In addition, the Meta approach *can reduce subjectivity* by querying several engines; like in the real-world, one can gather the views from several people in order to make the answers more accurate and objective. The speed provided by several systems queried in parallel can also significantly exceed those obtained by working with only one system, since their responsiveness may vary with the task and network traffic conditions.

In the first version of our prototype, we have included several freely available demonstrational prototypes and popular commercial engines that have some QA (fact seeking) capabilities or provide at least are capable of providing precise information on a given subject, specifically START, AskJeeves, BrainBoost, Wikipedia and ASU QA. The meta-search part of our system was based on the MetaSpider architecture (Chau et al., 2001). Multi-threads are launched to submit the query to fetch the candidate answers from each service. After these results are obtained, the system performs answer extraction, triangulation and semantic verification of the results, based on the algorithms from Roussinov and Robles (2004).

REFERENCES

Chau, M., Chen, H., and Zeng, D. (2001). Personalized Spiders for Web Search and Analysis. *Proceedings of the 1st Joint Conference on Digital Libraries*, Roanoke, Virginia, June 2001, pp. 79-87.

Roussinov, D., & Robles, J. (2004). Web Question Answering: Technology and Business Applications. *Proceedings of the American Conference on Information Systems, 2004.*