December 2006

Web Question Answering Systems: Measuring Task Performance

José Antonio Robles-Flores
Arizona State University

Web Question Answering Systems: Measuring Task Performance

José Antonio Robles-Flores
Arizona State University
Jose.Robles@asu.edu

ABSTRACT

This research seeks to understand the user perspective in evaluating web question answering technology. Previous research concentrated in the design and development of question answering systems but left out the user's perspective. Web question answering technology is important because it represents a way to search for information on the web and it is argued that it can improve task performance. This research proposes to study the relationship between the use of search technology and task performance. A research model is presented and a set of experiments are designed to test the hypotheses.

Keywords

Question Answering, WWW, Task Performance, Knowledge Intensive Tasks.

INTRODUCTION

The web represents a rich source of information and allows people potential access to knowledge that was not readily available before (Katz, Felshin, Lin and Marton, 2004). Despite recent improvements made by general purpose search engines (GPSE), like Google and Yahoo, they are not designed to deal with natural language questions. These systems treat questions as a "bag of words" and the query "Who is the largest producer of software?" will be treated the same as "software largest producer", which produces unexpected results.

The goal of question answering (QA) technology is to locate, extract, and represent a specific answer to a user question expressed in natural language. For example, a QA tool takes as input "What is mad cow disease?" and should output "Mad cow disease is a fatal disease of cattle that affects the central nervous system. It causes staggering and agitation." and a link to a source web page that provides the answer. This definition fits one of the characteristics of knowledge management systems: support the retrieval of information and knowledge (Alavi and Leidner, 2001). Therefore, it could be argued that QA technology may be treated as a knowledge management tool and, consequently it is an IT artifact designed for a specific purpose. Considering the domain of the system, QA systems may be either closed-domain or open-domain systems. In this research, we look at open-domain QA technology. The domain is the World Wide Web (Web). We call this type of systems WebQA.

Previous research noted that a significant proportion of queries on search engines typically have a question in mind even if they were not entered as natural language questions themselves (Radev, Qi, Zheng, Blair-Goldensohn, Zhang, Fan and Prager, 2001). Also, by investigating applications of QA tools, it was found that they can help to locate malevolent content (Roussinov and Robles-Flores, 2005a) and were useful in several business intelligence tasks (Roussinov and Robles-Flores, 2005b). In addition, asking questions and finding the corresponding answers is an essential activity for learning (Bruner, 1960; Pask, 1975). However, regardless of these advances with WebQA technology, keyword searching is still the most widely available and used technology.

QA technology is seen today also as a revenue-making tool for companies like Google, Microsoft's MSN, Yahoo, etc. QA technology is receiving more attention since the acquisition of AskJeeves, a search engine specialized in QA, for which the media giant IAC paid more than $1.8 billion.

Preliminary investigations on business processes and their tasks, led to the concept of knowledge-intensity. Three dimensions define it: contingency, knowledge worker participation, and knowledge characteristics (Robles-Flores and Kulkarni, 2005). It is expected that knowledge-intensive tasks lead to different types of questions compared to those from less knowledge-intensive tasks. Therefore this research proposes to study the effect of knowledge-intensity on task performance.

Because English is the predominant language on the Internet, being able to find answers in documents written in English is important. However, English is not the predominant language in the World's population and, as connectivity to the Internet grows in non-English speaking countries, more non-native English speakers will be accessing the Web. Therefore, it is believed that the tools that specially target non-native English speakers in order to facilitate the search and retrieval of documents in English will be particularly helpful. It is proposed in this research that one such tool is the QA technology because asking a question in a natural language is an easier task than coming up with the appropriate keywords, which requires more in-depth knowledge of English vocabulary.

Those observations motivate this research. Several QA systems have been proposed as research prototypes (Katz, 1988, 1997; Nyberg, Mitamura, Carbonell, Callan, Collins-Thompson, Czuba, Duggan, Hiyakumoto, Hu, Huang, Ko, Lita, Murtagh, Pedro and Svoboda, 2002; Roussinov, Robles-Flores and Ding, 2004). Other online QA systems are available as commercial portals (AskJeeves, Brainboost, AnswerBus, etc). The Text Retrieval Conference (TREC) co-sponsored by the National Institute of Standards and Technology (NIST) the Defense Advanced Research Projects Agency (DARPA) and the Advanced Research and Development Activity (ARDA), provides a competition-like environment and has a track to evaluate QA systems. However, most of the research has left out the user's perspective in evaluating these systems and the efforts have concentrated on the algorithms and the technology itself. The AAAI Spring Symposium on QA (Maybury, Day, Prange, Pusteyovsky and Wiebe, 2003) suggested a roadmap for QA and indicated the need for evaluation methods involving users. This research proposes to study open-domain, web based, QA technology and evaluate them with greater emphasis on the user perspective, particularly when QA is used for business tasks.

For the purposes of this research, a prototype of a WebQA system is already available. It has been tested and preliminary experiments have been conducted.

RESEARCH QUESTIONS

Prior research exploring the value of information technology has found that there is a positive correlation between investments in IT and organizational performance. These results were obtained both through quantifiable measures (ROI, market share, cost, productivity, profitability) as well as qualitative measures (i.e. efficiency, responsiveness, organizational structure, coordination, flexibility, change, quality, decision-making). A summary is provided by ISWorld in http://business.clemson.edu/ISE/html/organizational_impact.html. Following the model that links information technology and business process performance (Melville, Kraemer and Gurbaxani, 2004) and the definition of a business process (Davenport and Short, 1990; Sharp and McDermott, 2001; Smith and Fingar, 2003), this research looks at the impact of WebQA technology on task performance in a business environment. Business tasks are the specific and atomic activities of a business process. Thus, the first question that this study explores is:

> *To what extent does WebQA technology affect performance of business tasks?*

This research also looks at the effects of knowledge-intensity of business processes and tasks as described in (Robles-Flores and Kulkarni, 2005), therefore the second research question is:

> *To what extent does knowledge-intensity of the task affect performance of the task when it is supported by WebQA technology?*

The increased popularity of smaller devices like PDA's and cell phones with access to the internet has led to research about the limitations imposed by reduced screens (Buchanan, Jones and Marsden, 2002; Buchanan and Matt, 2000; Jones, Buchanan and Thimbleby, 2002). It is believed that WebQA technology may be helpful in overcoming the constraints and therefore this study also explores the following research question:

> *To what extent does WebQA technology affect task performance when using reduced screen devices (such as PDA's and cell-phones) ?*

Applications of WebQA technology in helping users with different kinds of limitations such as blind users and non-native English speakers are another issue of interest in this research. Therefore, the third research question is:

> *To what extent does WebQA technology affect task performance under certain user conditions such as blind users and non-native English speakers?*

METHODOLOGY

The research will be conducted in four phases:

Phase 1: Explore the applicability of QA technology to business tasks. A survey and interviews will be constructed in a similar way as in a preliminary research. The survey will be used to select between 5 and 10 questions which each of the interviewees may find most relevant to their professional responsibilities based on their personal experience. The original set of questions, from which the interviewees select questions, will be obtained by random drawing of questions from the combination of the set of TREC questions and the Excite set. Both sources have been previously used in QA research (Radev et al., 2001) and in a pilot study conducted within this research. The selection of questions by the interviewees will produce a final set which will be analyzed to draw conclusions regarding the types of questions selected. The interviews will help to verify the conjectures (e.g. the types of questions that are important in a given business scenario) and gain insights (e.g. most of the questions refer to specific scenarios – for example, more than half of the questions are related to the business intelligence process), and eventually result in preparing scenarios to be used in the subsequent phases. Subjects considered for this phase include full-time MBA, evening MBA, executive MBA, and Masters of Science in Information Management Program students.

Phase 2: Building scenarios for the experiments. Based on the results from phase 1, two types of scenarios will be constructed. One relates to high knowledge-intensive business tasks while the other represents low knowledge-intensive business tasks. Possible examples of scenarios may be:

> *Business Intelligence for New Product Development:* For example, within the new product development (NPD) process there is need to understand the market and the customers in order to have a successful process. An example from previous research (Robles-Flores and Kulkarni, 2005) is a NPD process distinguishing different knowledge items that are needed in order to understand (have knowledge about) the market and the customers. Sample questions that need to be answered as part of the NPD process are: a) what is the most popular color of a car? b) How many cars are sold each year in Phoenix?, c) What is the average price of a vehicle?, d) What is the gas guzzler tax?
>
> *Troubleshooting in a Help Desk for computer-related problems in an Education Institution:* The IT Help Desk is usually the single point of contact for all information technology users within the organization. For example, in a higher education institution, the IT help desk assists multiple types of users (e.g. administrators, professors, students) with different types of settings (e.g. personal desktop computers, wireless laptops, servers, communications equipment, networks, etc.) and multiple types of services (e.g. networking/communications, software, backups, updates, etc.). The main purpose of the help desk is to centralize requirements related to IT problems and to find solutions to these problems or derive requests to other alternative support areas. The IT help desk staff needs to quickly find the root cause of the problems and, once identified, they need to find the appropriate solution. For example, when diagnosing a desktop computer reported to send unsolicited emails, the help desk staff may need to answer some of the following questions: a) What virus causes spam emails?, b) What error message is produced by spam email?, c) What antivirus prevents receiving emails?, d) How does antivirus software prevent virus?

Phase 3: Experiments: The scenarios constructed and pilot tested in the previous phase will be used to provide the context for a set of experiments. The subjects involved in the experiment will be instructed to consider the scenario in which they are introduced and will be asked to come up with questions related to that scenario. Then the subjects will be instructed to find the answers to those questions using the keyword searching tool for half of the questions and the WebQA tool for the other half of the questions, switching turns to minimize learning effects. In order to have more realistic measures, the users will be allowed to click on the links returned by the tool and read the web pages found, but will not be allowed to follow the links from those pages any further. One experiment explores the difference in task performance between the use of keyword searching and WebQA controlling for the user's experience. A second experiment incorporates the variable representing the knowledge-intensity of the task to explore how it affects the relationship between the technology and task performance. A third experiment introduces the communication channel variable in order to explore the effect of using small screen devices.

Phase 4: Exploring User Characteristics. This phase intends to explore different characteristics of users. Two additional groups of subjects participate in two more experiments in order to measure the effect of the user characteristics: visually impaired users and non-native English speakers. Visually impaired users can not quickly glance over snippets, thus it is essential for them to get results in the most concise format possible, which is provided by WebQA technology. Non-native English speakers have difficulty finding the appropriate keywords for a search engine and writing a question in natural language may be an easier task.

CONTRIBUTION

This research addresses issues concerning WebQA technology and has implications for both academics and practitioners. From a practitioner standpoint, this research provides a better understanding of the contribution of WebQA technology on task performance and its applicability in business scenarios. It provides guidance about the tasks on which applying WebQA technology increases performance. From and academic perspective, the research proposes an evaluation model which is replicable and based on the user's perspective. Current research efforts in the information retrieval field, particularly on QA technology, use evaluations based on restricted domains. Evaluations of web-based systems do not take into consideration the user's role and perception. The proposed methodology to evaluate WebQA technology mainly focuses on the user's point of view and the impact on task performance.

REFERENCES

1. Alavi, M., and Leidner, D. (2001) Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues, MIS Quarterly, 25,1, pp 107-136.
2. Bruner, J. (1960) The process of education, Harvard University Press, Cambridge, MA.
3. Buchanan, G., Jones, M., and Marsden, G. (2002) Exploring Small Screen Digital Library Access with the Greenstone Digital Library, Proceedings of Sixth European Conference on Research and Advanced Technology for Digital Libraries (ECDL), September 16-18, Rome, Italy.
4. Buchanan, G., and Matt, J. (2000) Search Interfaces for Handheld Mobile Devices (Poster paper), Proceedings of Ninth International World Wide Web Conference, May 15-19, Amsterdam.
5. Davenport, T.H., and Short, J. (1990) The New Industrial Engineering: Information Technology and Business Process Redesign, MIT Sloan Management Review, 31,4, pp 11-27.
6. Jones, M., Buchanan, G., and Thimbleby, H. (2002) Sorting out Searching on Small Screen Devices, Proceedings of Fourth International Symposium Mobile HCI 2002, September 18-20, Pisa, Italy.
7. Katz, B. (1988) Using English for Indexing and Retrieving, Proceedings of RIAO Conference on User-Oriented Content-based Text and Image Handling, March 21-24, Cambridge, MA.
8. Katz, B. (1997) Annotating the World Wide Web using natural language, Proceedings of RIAO Conference on Computer Assisted Information Searching on the Internet (RIAO 97), June 15-27, Montreal, Canada.
9. Katz, B., Felshin, S., Lin, J., and Marton, G. (2004) Viewing the Web as a Virtual Database for Question Answering, in: New Directions in Question Answering, M. Maybury (ed.), MIT Press, Cambridge, MA.
10. Maybury, M., Day, D., Prange, J., Pusteyovsky, J., and Wiebe, J. (2003) New Directions in Question Answering, Proceedings of Papers from 2003 AAAI Spring Symposium, Menlo Park, CA.
11. Melville, N., Kraemer, K., and Gurbaxani, V. (2004) Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value, MIS Quarterly, 28,2, pp 283-322.
12. Nyberg, E., Mitamura, T., Carbonell, J.G., Callan, J.P., Collins-Thompson, K., Czuba, K., Duggan, M., Hiyakumoto, L., Hu, N., Huang, Y., Ko, J., Lita, L.V., Murtagh, S., Pedro, V., and Svoboda, D. (2002) The JAVELIN Question-Answering System at TREC 2002, Proceedings of Text Retrieval Conference, November 19-22, Gaithersburg, Maryland.
13. Pask, G. (1975) Conversation, Cognition, and Learning, Elsevier, New York.
14. Radev, D., Qi, H., Zheng, Z., Blair-Goldensohn, S., Zhang, Z., Fan, W., and Prager, J. (2001) Mining the web for answers to natural language questions, Proceedings of ACM CIKM 2001: Tenth International Conference on Information and Knowledge Management, Nov 5-10, Atlanta, GA.
15. Robles-Flores, J.A., and Kulkarni, U. (2005) Knowledge Management Systems: A Business Value Perspective, Proceedings of Ninth Pacific Asia Conference on Information Systems (PACIS 2005), July 7-10, Bangkok, Thailand, pp. 325-337.
16. Roussinov, D., and Robles-Flores, J.A. (2005a) How Question Answering Technology Helps to Locate Malevolent Online Content, Proceedings of Eleventh Americas Conference on Information Systems (AMCIS 2005), August, 11-14, Omaha, NE.

17. Roussinov, D., and Robles-Flores, J.A. (2005b) Web Question Answering: Technology and Applications to Business Intelligence, International Journal of Internet and Enterprise Management, 3,1, pp 46 - 62.

18. Roussinov, D., Robles-Flores, J.A., and Ding, J. (2004) Experiments with Web QA System and TREC2004 Questions, Proceedings of TREC 2004 Conference, November 16-19., Gaithersburg, MD.

19. Sharp, A., and McDermott, P. (2001) Workflow Modeling: Tools for Process Improvement and Application Development, Artech House, Norwood.

20. Smith, H., and Fingar, P. (2003) Business Process Management (BPM): The Third Wave, Meghan-Kiffer Press, Tampa.