

User-Centered Evaluation of Question Answering Systems

Chorng-Shyong Ong, Min-Yuh Day, Kuo-Tay Chen, Wen-Lian Hsu, *Fellow, IEEE*

Abstract— With the rapid growth of the Internet and database technologies in recent years, question answering systems (QAS) have emerged as important applications. As most evaluation models focus on system-centered evaluation, user-centered evaluation has attracted little attention. Although many QAS have been implemented, little work has been done on the development of a user-centered evaluation for QAS. User-centered evaluation is used to understand a user's needs and identify important dimensions and factors in the development of an information system in order to improve its acceptance. The purpose of this study is to develop a user-centered evaluation model for QAS from the user's perspective. The proposed user-centered evaluation model provides a framework for the design of question answering systems from the user's perspective to enhance user satisfaction and acceptance of QAS.

Index Terms— Evaluation; Measurement; Question answering system; User satisfaction

I. INTRODUCTION

Evaluation of an information system's success and user satisfaction are important research issues in the field of information management, especially for online service systems on the Internet. User-centered evaluation models are used to understand users' needs and identify important dimensions and factors in the development of systems in order to broaden their acceptance. With the rapid growth of the Internet and database technologies in recent years, question answering systems (QAS) have emerged as important applications. Hence, they have received a great deal of attention from information systems researchers, particularly those in the information retrieval and natural language processing communities [4-6, 8, 10].

Extensive studies have been made on the evaluation of question answering systems. Appropriate evaluation would motivate research by providing suggestions for the overall improvement of the architecture and behavior of QAS. As most evaluation models focus on system-centered evaluation,

user-centered evaluation has attracted little attention. However, if we are to build a practical question answering system, we must achieve a performance level that satisfies the majority of users.

The purpose of this study is to develop a user-centered evaluation model for question answering systems based on a literature review of information system models and theories. Based on the user's perspective, we propose an integrated reference model for the design of QAS to enhance user satisfaction and acceptance.

II. LITERATURE REVIEW

To better understand the user-centered evaluation of question answering systems, we systematically review five major IS theories, namely, the Theory of Reasoned Action (TRA) [3]; the Technology Acceptance Model (TAM)[2]; the Theory of Planned Behavior (TPB)[1]; the Unified the Theory of Acceptance and Use of Technology (UTAUT)[7]; and the Theoretical Integration of User Satisfaction and Technology Acceptance (TIUSTA)[9].

The Theory of Reasoned Action (TRA) [3], which is drawn from social psychology, is one of the most fundamental and influential theories of human behavior. Davis et al. [2] applied TRA to individual acceptance of technology and found that the variance in acceptance was largely consistent with studies that had employed TRA in the context of other behaviors. The Technology Acceptance Model (TAM) is an adaptation of TRA. It was designed specifically for modeling and predicting information technology acceptance and usage in the work environment. Ajzen [1] refined TRA and proposed the Theory of Planned Behavior (TPB). The major difference between TRA and TPB is that the latter incorporates a third determinant of behavioral intention, namely, perceived behavioral control.

Venkatesh et al. [7] observed that information systems research has long studied how and why individuals adopt new information technologies. There are several streams of IS research, one of which focuses on individual acceptance of technology by using intention or usage as a dependent variable. Other streams focus on implementation success at the organizational level, or on task-technology fit. In addition, Wixom and Todd [9] indicated that research on perceptions of information systems (IS) success can be categorized in to two primary research streams – user satisfaction literature and technology acceptance literature. However, as these two approaches have been developed in parallel, they have not been

Chorng-Shyong Ong is with the Department of Information Management, National Taiwan University, Taiwan (e-mail: ongcs@im.ntu.edu.tw).

Min-Yuh Day is with the Department of Information Management, National Taiwan University, Taiwan and Institute of Information Science, Academia Sinica, Taiwan (corresponding author, phone: 886-227883799 ext. 1366; fax: 886-226518660; e-mail: myday@iis.sinica.edu.tw).

Kuo-Tay Chen is with the Department of Accounting, National Taiwan University, Taiwan (e-mail: ktchen@management.ntu.edu.tw).

Wen-Lian Hsu is with the Institute of Information Science, Academia Sinica, Taiwan (e-mail: hsu@iis.sinica.edu.tw).

reconciled or integrated.

III. PROPOSED USER-CENTERED EVALUATION FOR QAS

Information systems research has resulted in the development of several theoretical models, with roots in information systems, psychology, and sociology, to explain user satisfaction and an individual's intention to use technology. However, as Wixom and Todd [9] indicate, there is a need to develop a refined understanding of the relationships between the theoretical integration of user satisfaction and technology acceptance.

In this paper, we propose a user-centered evaluation model for question answering systems based on a review and synthesis of existing models of IS user satisfaction and technology acceptance. The proposed user-centered evaluation model for question answering systems is shown in Figure 1. There are ten dimensions in the proposed user-centered evaluation model for question answering systems, namely, information quality, system quality, service quality, information satisfaction, system satisfaction, service satisfaction, perceived usefulness, perceived ease of use, intention to use, and use behavior.

The fundamental concept of the proposed model is inspired by the TRA believe-attitude-intention-behavior theory, which is one of the most influential theories of human behavior [3]. We adopt three dimensions of quality from the updated DeLone and McLean Information Systems (IS) Success Model, namely, information quality, systems quality, and service quality. Like quality, satisfaction also has three major dimensions: information satisfaction, systems satisfaction, and service satisfaction. Quality will affect subsequent satisfaction.

Information quality is defined as the quality of the information provided by a QAS. Similarly, system quality is defined as the quality of the system provided by a QAS, while service quality is defined as the user's judgment about a QAS service's overall excellence or superiority.

Information satisfaction is defined as the extent to which an individual's attitude influences the gap between expectations and the perceived performance of the information provided. Similarly, system satisfaction is defined as the extent to which an individual's attitude influences the gap between expectations and the perceived performance of the system; while service satisfaction is defined as the extent to which an

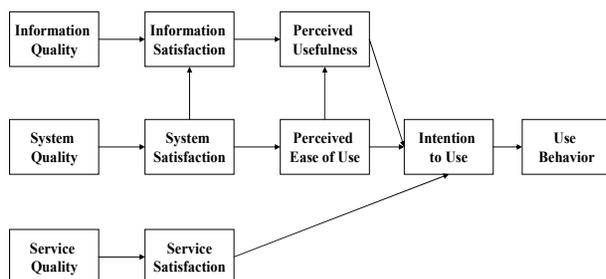


Figure 1. Proposed user-centered evaluation model for Question Answering Systems

individual's attitude influences the gap between expectations and the perceived performance of the service.

Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance," while perceived ease of use is defined as "the degree to which a person believes that using a particular system would be free of effort".

Intention to use is defined as the extent to which an individual intends to perform a specific behavior. Use behavior is defined as the actual use behavior of QAS.

IV. CONCLUSION AND FUTURE RESEARCH

To enhance user satisfaction and the success of question answering systems, we have proposed a user-centered evaluation model for QAS, based on a review and synthesis of existing IS user satisfaction and technology acceptance models. We believe the proposed user-centered evaluation model provides a framework for the design of question answering systems from the user's perspective and that it could help increase user acceptance of QAS.

In the future, we will use the model to conduct an empirical study of question answering systems and analyze the theoretical and managerial implications of this study's results.

ACKNOWLEDGMENT

This research was supported in part by the National Science Council of Taiwan under Grants NSC 95-2416-H-002-047 and NSC 95-2752-E-001-001-PAE.

REFERENCES

- [1] I. Ajzen, "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, pp. 179-211, Dec 1991.
- [2] R. D. Davis and R. P. Bagozzi, "User acceptance of computer technology: a comparison of two theoretical models," *Management Science*, vol. 35, pp. 982-1003, 1989.
- [3] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. MA: Addison-Wesley, 1975.
- [4] R. P. Schumaker and H. Chen, "Leveraging Question Answer technology to address terrorism inquiry," *Decision Support Systems*, vol. 43, pp. 1419-1430, Aug 2007.
- [5] R. P. Schumaker, M. Ginsburg, H. C. Chen, and Y. Liu, "An evaluation of the chat and knowledge delivery components of a low-level dialog system: The AZ-ALICE experiment," *Decision Support Systems*, vol. 42, pp. 2236-2246, Jan 2007.
- [6] R. P. Schumaker, Y. Liu, M. Ginsburg, and H. C. Chen, "Evaluating the efficacy of - A terrorism question/answer system," *Communications of the Acm*, vol. 50, pp. 74-80, Jul 2007.
- [7] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology: Toward A Unified View," *MIS Quarterly*, vol. 27, p. 425, Sep 2003.
- [8] N. Wacholder, D. Kelly, P. Kantor, R. Rittman, Y. Sun, B. Bai, S. Small, B. Yamrom, and T. Strzalkowski, "A model for quantitative evaluation of an end-to-end question-answering system," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1082-1099, Jun 2007.
- [9] B. H. Wixom and P. A. Todd, "A theoretical integration of user satisfaction and technology acceptance," *Information Systems Research*, vol. 16, pp. 85-102, Mar 2005.
- [10] H. Yu, M. Lee, D. Kaufman, J. Ely, J. A. Osheroff, G. Hripscak, and J. Cimino, "Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians," *Journal of Biomedical Informatics*, vol. 40, pp. 236-251, Jun 2007.