

A Conversation Repository for Participatory Librarianship

Abstract

The goal of this study is to model conversation data from digital reference services and reconstruct the processes of the reference interviews as a conversation repository. By doing so, I am hoping that the study will provide an infrastructure for applications such as an interactive information retrieval system, a question answering system, or a recommender system for information intermediaries. While the current study focuses on the digital reference conversation, observations may be generalizable, to a certain extent, to different online conversations such as blogs, question answering sites, social networking sites, or product review systems, and thus contribute to a wider range of knowledge sharing and knowledge creation through information technologies.

This work is a part of the Participatory Librarianship research initiative, proposed by Lankes et al. (2006), which seeks an implementation of library as a facilitator of technology-enhanced human communication and knowledge creation, based on Gordon Pask's Conversation Theory. (Pask, 1975, 1976)

Motivation

While fully automated Web search engines have become the primary tool for information searching of the majority of Web users, human-in-the-loop systems such as product review systems (e.g. Amazon), community-based QA systems (e.g. Yahoo! Answer), and online chat services with information experts (e.g. help desk and virtual reference services) have also been increasing in popularity in recent years. (Resnick et al., 2000; Madden and Fox, 2006) These systems, which became available largely through the development of Web 2.0 technologies, provide an environment for information seeking dialogues and social networking opportunities by employing easy-to-use interfaces for interactive communications among the users in a public domain or a closed shared domain. Since the questions and answers are exchanged in natural language texts, as opposed to query terms and hyperlinks as in the case of Web search engines, the messages exchanged in these systems are richer and more complex in their information. Moreover, since the communications are made online, the transcripts of the conversations can be stored for future reference. However, as rich and complex as the natural language conversations are, searching through an archive of conversations is a difficult task, because of its linguistic structure being different from other kind of text data such as books or documents, where information retrieval has a long tradition of research.

For example, in an information-seeking conversation, even the description of the information sought in the conversation may not always be available. It is known that one of the challenges of a reference interview is to find out what the user is asking for. (Taylor, 1968; Belkin et al., 1982). According to the ASK Hypothesis (Belkin et al., 1982), users are often unable to identify their information need at the beginning of an interview, and the information requirement becomes overt only through a successful information-seeking dialogue, which consists of iterative interactions between the user and the information intermediary with a certain structure. Early studies in information behavior (Oddy, 1977; Brooks and Belkin, 1983; Daniels et al., 1985; Brooks et al., 1986; Belkin et al., 1987) confirmed this hypothesis, by revealing the discourse structure of various reference interviews. For example, Daniels et al. (1985) analyzed face-to-face interviews at the reference desk in various research institutes at London University and identified the *problem structure* of reference conversation, which organized the goals of utterances in the interviews hierarchically. However, the focus of the majority of IR system research has been experimental studies based

on the “single-shot” search and evaluation, and a few studies have incorporated the notion of iterative interaction. Relevance feedback is an active field of study in IR, which employed iterative feedbacks from the user, but as Bates (1989) pointed out, the method assumes that the user’s information need never changes throughout the search process and thus sits well in the classic model of IR systems.¹

On the other hand, the library has long been a place for information-seeking interactions, most explicitly through reference services, but also through other functions or events such as building collections based on community needs or organizing book groups and speaker series. Searching books by browsing through book shelves is also a form of interactive information searching: the user can look for books in different directions based on the evolving information need: from a broad topic to a narrower one or old books to newer ones. Collecting information piece by piece, known as *berrypicking* (Bates, 1989), which is an important capacity of information-seeking interactions, is also available in this mode of information searching.

The Participatory Librarianship research initiative, proposed by Lankes et al. (2006), seeks an implementation of library as a facilitator of technology-enhanced information interactions. It is motivated by Gordon Pask’s conversation theory (Pask, 1975, 1976), which posits that knowledge is created through conversations. A conversation repository is an infrastructure to facilitate such conversations by incorporating theories from information behavior study, (ASK hypothesis, berrypicking model), cybernetics (conversation theory), and discourse linguistics (Grosz and Sidner model) into the emerging participatory information systems that has been enabled by the Web 2.0 technologies.

Current Approach: Discourse Analysis for Conversation Repository

While the intentions of these theories which support knowledge creation and information discovery through iterative interactions resonate with the spirit of Participatory Librarianship, the technical aspects of them have not developed enough to be compatible with the emerging participatory information technologies. However, we believe that the methods of discourse analysis can complement and modernize the earlier work by revealing the structure of information-seeking conversations and making it available as an input to information systems to enhance information retrieval processes.

Discourse analysis is an approach for studying instances of language beyond the sentential level, such as utterances in conversation. Linguist, Harris (1952) defined the term as a formal method for linguistically analyzing “connected speech (or writing)”, but the approach has been applied to various fields in social science, where the focus is on the relations between the language and its effects to the society. For understanding of conversation in the information seeking context, Grosz and Sidner (1986) developed a framework that analyzes the discourse structure by three distinctive (but interrelated) components: linguistic structure, intentional structure, and attentional state. Their approach has been directly or indirectly applied in numerous studies. (Daniels et al., 1985; Carberry and Lambert, 1999; Carberry et al., 2006; McTear, 2002)

The study will analyze the discourse structure of the virtual reference interviews based on their framework. Specifically, the study will seek answers for the following questions:

- What are the linguistic characteristics of digital reference interviews?

¹The HARD track for TREC (Allan, 2005) and studies in Cognitive Information Retrieval (Spink and Cole, 2005) are some of the exceptions.

- What are the intentions of utterances in digital reference interviews and how are they related to the linguistic characteristics?
- How do the attentions change in digital reference interviews and how are they related to the linguistics characteristics?
- How is the discourse structure of digital reference interviews different from other types of conversation?

In addition, the study will examine the following questions in order to contribute to implementing conversation-based information systems.

- What are the machine learning techniques that suit for learning the discourse structure of information seeking dialogues?
- How can the automatic detection of discourse structures be utilized in the information seeking process?

Data

Our current data, provided by the Online Computer Library Center (OCLC), is a log of 450 virtual reference service interviews, consisting of 8066 lines of messages exchanged by various users and librarians from all over the United States. The reference interviews took place in the form of on-line chat sessions with a co-Web-browsing capability.

Conclusive Remarks

The combination of conversation theory, information behavior study, discourse analysis, and emerging information technologies have enabled us to design our research in the following ways:

- Pask's conversation theory provides a conceptual framework to build an information repository that emphasizes knowledge creation through conversations.
- The early studies of information behavior provides a theoretical motivation for incorporating the notion of conversation into information retrieval technologies, enabled by the emerging participatory information systems.
- Discourse analysis methods can be used to reveal the structure of conversation.
- These discourse models will inform the development of conversation repository, an infrastructure for conversation-based information systems.

While the study currently focuses on the digital reference conversation, the observed discourse structure may be generalizable, to a certain extent, to different kinds of conversation. Especially, the conversations through the participatory information technologies, such as community-based question answering systems, blogs, social networking sites, or product review systems, may provide a wide range of complex and detailed daily-life information that has not been easily searchable. Therefore, incorporating the observations from this study to these conversations will contribute to a wider range of knowledge sharing and creation through information technologies.

References

- Allan, J. (2005). Hard track overview in trec 2005 high accuracy retrieval from documents. In Voorhees, E. M. and Buckland, L. P., editors, *Proceedings of the Fourteenth Text REtrieval Conference Proceedings*.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5):407–424.
- Belkin, N. J., Borgman, C. L., Brooks, H. M., Bylander, T., and Croft, W. B. (1987). Distributed expert-based information systems: an interdisciplinary approach. *Inf. Process. Manage.*, 23(5):395–409.
- Belkin, N. J., Oddy, R. N., and Brooks, H. M. (1982). Ask for information retrieval. part i: Background and theory. *Journal of Documentation*, 38(2):61–71.
- Brooks, H. M. and Belkin, N. J. (1983). Using discourse analysis for the design of information retrieval interaction mechanisms. In *SIGIR '83: Proceedings of the 6th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 31–47, New York, NY, USA. ACM Press.
- Brooks, H. M., Daniels, P. J., and Belkin, N. J. (1986). Research on information interaction and intelligent information provision mechanisms. *J. Inf. Sci.*, 12(1-2):37–44.
- Carberry, S., Elzer, S., and Demir, S. (2006). Information graphics: an untapped resource for digital libraries. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 581–588, New York, NY, USA. ACM Press.
- Carberry, S. and Lambert, L. (1999). A process model for recognizing communicative acts and modeling negotiation subdialogues. *Comput. Linguist.*, 25(1):1–53.
- Daniels, P. J., Brooks, H. M., and Belkin, N. J. (1985). Using problem structures for driving human-computer dialogues. In *In RIAO 85, Acts of the Conference: Recherche d'Informations Assistee par Ordinateur*, pages 645–660, Grenoble. I.M.A.G.
- Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Harris, Z. S. (1952). Discourse analysis. *Language*, 28(1):1–30.
- Lankes, R. D., Silverstein, J., and Nicholson, S. (2006). *Participatory Networks: The Library as Conversation*. Information Institute of Syracuse.
- Madden, M. and Fox, S. (2006). Riding the waves of “web 2.0”: More than a buzzword, but still not easily defined.
- McTear, M. F. (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Comput. Surv.*, 34(1):90–169.
- Oddy, R. N. (1977). Information retrieval through man-machine dialogue. *Journal of Documentation*, 33(1):1–14.

- Pask, G. (1975). *Conversation, cognition and learning*. Elsevier, New York.
- Pask, G. (1976). *Conversation Theory Applications in Education and Epistemology*. Elsevier.
- Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.
- Spink, A. and Cole, C. (2005). *New Directions in Cognitive Information Retrieval*. Springer.
- Taylor, R. T. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries*, 29:178–194.