Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2012 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

7-15-2012

"Who Needs To Know?": How Different Aspects Of The User's Situation Are Important For Answering Different Query Types

David Bodoff
Graduate School of Management, University of Haifa, Haifa, Israel, dbodoff@univ.haifa.ac.il

“WHO NEEDS TO KNOW?”: HOW DIFFERENT ASPECTS OF THE USER'S SITUATION ARE IMPORTANT FOR ANSWERING DIFFERENT QUERY TYPES

David Bodoff, Graduate School of Management, University of Haifa, Haifa, Israel,
dbodoff@univ.haifa.ac.il

Abstract

Prior research has established that various aspects of the user’s situation, collectively called the user model, affect what information is relevant. The purpose of our research is to refine this idea by exploring how different aspects of the user model are salient for different question types. Our methodology follows tradition in studying real intermediary elicitations for clues about what aspects of the user model are important, except that we analyze how this differs across question types. We find that there are more elicitations about the background of the user’s task and about the relevance of particular information for longer-answer questions than for short-answer questions, but surprisingly, no more elicitations regarding the sufficiency of particular information. The practical application of our research is to guide human or automated respondents to focus on the user details that are most important for different types of question.

Keywords: User Modelling, Question Answering.

1 INTRODUCTION

Research on user-centered information retrieval has identified various aspects of the user context that influence which information is relevant to the user at a particular time. Taken together, these aspects are known as the user model. In this paper, we investigate whether different parts of the user model are important for different types of questions. Our research draws on two branches of theory, the nature of questions, and user modelling.

The nature of questions has been a subject of study in disciplines ranging from philosophy and logic to information science and computer science (Saracevic, Kantor et al. 1988). Much of the prior research about question characteristics has studied the questions brought to a traditional library reference desk. The world wide web offers new question-answering settings, including online reference interviews with traditional librarians, as well as brand new structures including social question-answering sites such as Yahoo Answers (Harper, Weinberg et al. 2010), web-based research services such as Wikipedia reference desk (Shachaf 2009), online forums, and others. Furthermore, questions submitted to web-based research and reference services are quite varied (Harper, Weinberg et al. 2010). The variety of question types and new question-answering settings motivates our renewed interest in the practical importance of correctly characterizing and classifying users' questions. The ultimate purpose of classifying questions is to improve the answers that the human or automatic system provides to the questioner. In this paper, we investigate how question classification can be used to illuminate what aspects of the user model must be known in order to appropriately answer his/her question. This brings us to the literature on user modeling.

A central tenet of user modeling in information retrieval is that two different users who submit the identical query may consider different information to be relevant, due to elements of the users' situations that are not reflected in their queries (Belkin 1993; Saracevic 1997). These elements may include the user's background, the purpose for which the answer is needed, and so on (Belkin, Seeger et al. 1983; Brooks and Belkin 1983; Daniels, Brooks et al. 1985; Belkin, Brooks et al. 1987; Saracevic, Mokros et al. 1990; Saracevic, Spink et al. 1997; Spink, Goodrum et al. 1998). The purpose of user modeling is that by modeling these aspects, the respondent – whether human or automated -- can provide information that is more relevant to a specific user's situation. The purpose of *research* in user modeling is to determine which elements of the user situation are important for the respondent to model.

The primary methodology in research that investigates these elements, is studying intermediated search sessions, in which a user interacts with a human search intermediary (Belkin, Brooks et al. 1987; Spink, Goodrum et al. 1995; Saracevic, Spink et al. 1997; Wu and Liu 2003). The idea is that when the intermediary is observed eliciting (i.e. asking the user) about a detail of the user's situation, it shows that – at least in the mind of that intermediary – that detail is important for the respondent to know, if he/she is to provide relevant information. From these "elicitations", researchers construct a user model schema, i.e. a set of user characteristics that are hypothesized to affect what information the user would find useful, and which respondents should therefore try to determine for each specific user.

The work reported here relates the literature on questions types with the literature on user modeling. It explores how, for different question types, different details of the user situation are important to know —or to elicit from the user if they are not known—in order to adequately answer the user's question. The practical application of our results is for interactive question answering. Whether the respondent is a human or an automated or semi-automated question answering system, it may try to improve its responses by determining various aspects of the used model. Our research aims to guide this process by illuminating which elements of the user model are most important for different types of questions. The primary theoretical contribution of our work is in our integration of ideas from user modeling and from question typologies.

The setting of our empirical study is an online question-answering service called Google Answers (GA). GA is comparable to the Yahoo Answers (YA) service, and although GA has seen been retired from use by Google, we chose GA over YA as the source for our data because unlike YA, GA was a for-pay service, and its questions tended to be of a more serious nature with explicit economic value to the questioners. GA users submit questions online to GA "experts" who are experts in searching for information, not domain experts. In this regard, the Web-based research services that we study are an online version of traditional mediated search.

2 QUESTION TYPOLOGY AND ELICITATION TYPOLOGY

2.1 Question Typology

We identify four prominent typologies that have been proposed in the literature: (1) Lehnert's (1978) typology organizes questions by the function of the expected answer; (2) Sears (2001) and Katz (2002), who organize questions based on the kind of reference service that they require; (3) White and Iivonen (2001) classify questions according to whether they are open or closed, and according to whether one could predict where relevant information could be found; and (4) Graesser et al. (1994). Among these, we adopt the Graesser et al. typology as most fitting to our theory and setting. An informal analysis of our dataset, before we had seen the Graesser et al. typology, led us to identify and distinguish (among other distinctions) between Causation, Who-what-where-when questions, and How-to questions, categories that appear in this taxonomy, indicating that the Graesser taxonomy is theoretically pertinent for our purposes. Also, this taxonomy was developed with regard to questions asked in a variety of settings, making it widely appropriate. Moreover, its categories can be logically related to user model schema, and so are directly applicable to our case.

The categories developed in Graesser et al. are defined "primarily on the basis of the content of the information sought" (Graesser et al. 1994 p. 520). The full taxonomy is reproduced here as Table 1. Our research only utilizes four of these categories -- Concept Completion, Enablement, Instrumental/procedural, and Quantification -- which comprised the majority of examples in our data sample. Each of these four is adapted to our setting, as we now explain.

Question category	Abstract specification
Short Answer	
Verification	Is a fact true? Did an event occur?
Disjunctive	Is X or Y the case? Is X, Y, or Z the case?
Concept completion	Who? What? What is the referent of a noun argument slot?
Feature specification	What qualitative attribute does entity X have?
Quantification	What is the value of a quantitative variable? How many?
Long Answer	
Definition	What does X mean?
Example	What is an example label or instance of the category?
Comparison	How is X similar to Y? How is X different from Y?
Interpretation	What concept or claim can be inferred from a static of active pattern of data?
Causal antecedent	What state or event causally led to an event or state?
Causal consequence	What are the consequences of an event or state?
Goal orientation	What are the motives or goals behind an agent's action?
Instrumental/procedural	What instrument or plan allows an agent to accomplish a goal?
Enablement	What object or resource allows an agent to perform an action?
Expectational	Why did some expected event not occur?
Judgmental	What value does the answerer place on an idea or advice?

Assertion	The speaker makes a statement indicating he lacks knowledge or does not understand an idea
Request/directive	The speak wants the listener to perform an action

Table 1: Graesser et al. (1994) taxonomy

Concept Completion questions ask Who-What-Where-When. They "are very much like fill-in-the-blank questions, insofar as they specify a particular event with one missing component" (Lehnert 1978 p. 69). Although the original definition does not include "Where" as one of the "w" questions we include cases where the questioner has in mind a particular item and seeks information about where to find it, or how to obtain it. The original Graesser et al. calls this "short answer", but we include all such questions, regardless of whether the answer is short or long in terms of how many words of explanation are required.

Quantification: This category includes questions that seek quantitative answers. An example is GA 370677: "*How much does a residential architect charge on average for a whole house remodel project?*"

Graesser et al. define Enablement questions as asking what object will allow an action, we include questions that ask what object meets certain criteria. This type of question was very common in GA. An examples of such a question is GA 369119: "*I live in NY and travel to LA often - every 4 to 6 weeks. I keep a car in LA. A month parking at the airport is 300 - 400.00…I was hoping to spend 150 or so for safe, covered parking…*". Graesser et al. do not explicitly discuss questions with this structure, but our interpretation is that they fit within the Enablement category, since the reason a user poses this type of question is that only such an item will allow him/her user to act.

Instrumental/Procedural questions ask what instrument or plan will allow the agent to accomplish a goal. How-to questions coded in this category. An example is GA 378422: "*Different business, hardware stores, farm supplies etc., once used elevators which were, by a series of ropes pulleys and counter weights, moved up or down with little effort by the person inside the car. They were very safe and stable and used no power, electrical or otherwise, but only manpower…I am trying to recreate one of these in my farm to traverse between two floors*". When a question asks what instrument will allow accomplishment of a goal, it borders on Enablement questions that ask what object will allow an action. The subtle distinction as we apply it is that instrumental/procedural questions include an emphasis on the questioner's more ultimate goal, whereas Enablement questions were more focused on items that meet fixed criteria.

2.2 Elicitation Typology

While section 2.1 developed our question typology, this section presents our typology of user situation characteristics. We are interested in identifying attributes of the user situation that affect which information is relevant; we will then propose that for some of them, their importance depends on the question type.

One way of developing an inventory of user attributes that affect relevance is though studying mediated search, in which a human mediator is engaged in a *visible*, intelligent, and interactive process of user modeling (Belkin, Brooks et al. 1987; Saracevic, Spink et al. 1997). If the search intermediary elicits a characteristic (of the user, his/her need, his/her context, etc.), this indicates that, at least in the intermediary's view, that characteristic affects what information will be relevant to the user. Thus, a standard methodology for developing user model schema is to study "elicitations", or requests for information, that intermediaries pose to users. Elicitations are categorized according to the user characteristic they aim to clarify, and the resulting categories comprise a user model schema or inventory of user situation characteristics. We adopted this methodology to develop an elicitations typology and user model schema for our setting.

As a baseline, we adopted a taxonomy that was developed in a series of studies by Saracevic, Spink, and colleagues (Saracevic, Mokros et al. 1990; Spink, Goodrum et al. 1995; Saracevic, Spink et al. 1997; Spink, Goodrum et al. 1998) based partly on earlier work by Belkin et al. (1987). The final categories and their exact definitions were refined through an iterative process of working with the theory that is relevant to our setting, and to the nature of our data. A summary of the process is provided in methodology section below.

The final typology we developed for our setting consists of eleven types. This means that intermediary elicitations were inquiring or commenting on eleven different aspects of the user situation. The full taxonomy is presented in Authors (2011). The hypotheses in the following section that relate elicitations to question type will utilize a subset of three of these elicitations types, in which the intermediary asks about:

(a) the background of the problem including the user's task;

(b) the relevance of particular information that the respondent offers as a potential answer or part of it; and

(c) the sufficiency of particular information that the respondent offers as comprising a sufficiently complete answer.

3 HYPOTHESIS DEVELOPMENT: IMPACT OF QUESTION TYPES ON ELICITATIONS

The independent variables in our study are the question categories developed in section 2.1, i.e. the four types from Graesser et al. that comprise the bulk of our sample, i.e. Instrumental/procedural, Quantification, Enablement, and Concept Completion. The dependent variables in our study are the three elicitation types highlighted in section 2.2, i.e. background, relevance, and sufficiency. In this section, we develop hypotheses about the relationships between these, based on observations of the nature of the question types.

In terms of the pattern of intermediary elicitations that we expect, we view Quantification questions ("how many" or "how much") as similar to Concept Completion questions (who-what-when-where). Due to their nature as short-answer questions, we expect that it will be naturally clearer to the intermediary what answer is sufficient, as compared with other question types:

H1a: There will be more elicitations about sufficiency of the answer for Enablement and Instrumental questions than for Concept Completion and Quantitative questions

Regarding task background and relevance, the situation is less clear. The reason is that even if the answer is to take a short form, e.g. the name of a supplier for a needed part, there may still be numerous points to clarify about the nature of the user's task and whether or not a particular item fits the need. And on the other hand, although Enablement questions are "long answer", they often specify rigid criteria that may not be open to much ambiguity. For this reason, although it seems straightforward enough to hypothesize fewer elicitations of all kinds for short as opposed to long-answer questions, the logic is not persuasive a priori. For exploratory purposes, we state two propositions in the affirmative:

H1b: There will be more elicitations about the user's task-background for Enablement and Instrumental questions than for Concept Completion and Quantitative questions

H1c: There will be more elicitations about the relevance of a proposed answer for Enablement and Instrumental questions than for Concept Completion and Quantitative questions

Although Enablement and Instrumental questions are both "long answer", and although they are even conceptually related as discussed above, we see clear reason to hypothesize differences in the user model. The reason is that since (as we have defined the categories) Instrumental questions are more

focused on solving an ultimate problem than Enablement questions, there is reason to expect that the user's task background will play a more important role.

H2: There will be more elicitations about the user's task-background for Enablement questions than for Instrumental questions

4 EMPIRICAL STUDY

4.1 Setting and Method

Google Answers (GA) was a Web-based research service established in April 2002 and decommissioned in December 2006. Only certified GA experts were able to offer answers or official elicitations. A user would pose a question and a proposed price they would pay for an answer. Any certified expert could attempt to answer the question directly, and/or could post one or more publicly visible "requests for clarification" (RFC's) to the questioner. These RFC's contained the expert intermediaries' elicitations.

Our empirical study covers 103 randomly selected GA questions. Detailed technical questions, especially about fixing computer problems were omitted because their elicitations (e.g. "what happens if you hold down the Alt key while rebooting?") defied the usual categories, which were developed for more conceptual questions.

Questions were categorized into one of the Graesser et al. categories. Two coders conducted training to achieve reliability of their codings. A batch of randomly chosen GA questions was chosen for each training iteration. The two coders would independently code each question into one of the Graesser et al. categories. Then the two coders would compare their codings and calculate inter-coder reliability. Training continued in this way until satisfactory inter-coder reliability of .7 was achieved using Cohen's kappa (Neuendorf 2002). Then the process commenced of categorizing the 103 questions that were used for actual data analysis, with 30 of the 103 being coded by both coders to re-confirm reliability, and the other 73 being coded by a single coder. Inter-rater reliability for the 30 was .67.

The process of deriving elicitation types was more iterative, being driven by the data as well as by theory. The methodology was patterned on the iterative process used throughout the literature e.g. (Belkin, Brooks et al. 1987; Spink, Goodrum et al. 1998). An initial set of categories was proposed based on the Saracevic taxonomy (1997), but the categories were adapted for our setting based on working with the data. This process repeated itself until inter-coder reliability was sufficiently high and the categories suffice to cover the data.

Regarding elicitations, a coder's job was not limited to categorizing elicitations, but also involved an earlier step of "chunking" the text that an intermediary posted into segments called "utterances". This is because the intermediaries post their "request for clarification" (RFC) from the questioner in a text box, and nothing prevents him/her from including many separate elicitations into a single RFC text box. There are a number of approaches regarding the necessity and measurement of reliability for unitizing. The reliability measure for unitizing is sometimes defined as the extent of agreement over the number of units to be coded (Guetzkow 1950), but does not consider whether the two coders agreed about the actual substance of those units. To be more conservative, we compared the two coder's unitizations of the RFC's, and considered as a "match" only those cases where the two coders completely agreed on the beginning and end words of a particular text fragment as constituting an utterance. We divided this by the greater of the total number of utterances that the two coders identified, and considered this as the reliability measure.

Training for the coding of elicitations continued until the category definitions were stabilized and until satisfactory inter-coder reliability was achieved, then the process of categorizing elicitations for the 103 questions commenced. There were 458 utterances from 240 RFC's for these 103 GA questions.

RFC's of 24 of the 103 questions, which contained 143 utterances, were coded by both coders to re-confirm reliability. Reliability of the categorization of those 143 utterances into elicitation categories was measured using Cohen's kappa, and was .77. For reliability of unitization, there was complete agreement on 122 out of 173 chunks, yielding a reliability of .71.

The hypotheses regard the relationship between question type and elicitations. To prepare the data for testing these hypotheses, a data record for a single question was created by merging the coded question types with the frequency counts of different types of elicitation for that single question. For example, GA 360432 asks about a source for purchasing men's gray-on-gray striped regimental ties. The question was coded as Concept Completion. This was then combined with the cumulative data showing the total numbers of utterances according to their elicitation type, for that particular question. An example is shown in Figure 1.

Question #	Question type (Graesser et al.)	Number of elicitations on task-background	Number of elicitations on review and relevance	Number of elicitations on sufficiency
GA 360432	Concept Completion	0	3	2

Figure 1: Example data record

Hypotheses were tested by comparing means between two samples. The data is non-normal, variances are unequal, and sample sizes are small to medium. Under these conditions, a t-test assuming unequal variances is best. Non-parametric tests – which are otherwise suitable for non-normal data – are not suggested with variances are unequal (Cribbie and Keselman 2003; Ruxton 2006), but results are similar.

4.2 Results

Descriptive statistics of questions by type are shown in Table 2. Descriptive statistics of elicitations are shown in Table 3. Results of tests of propositions are presented in Table 4.

Graesser et al. (1994) type	Frequency
Concept Completion	34
Enablement	20
Instrumental/Procedural	14
Quantification	13
Interpretation	4
Verification	5
Comparison	3
N/A	3
Definition	2
Causal Consequence	2
Causal Antecedent	2
Judgmental	1
Total	103

Table 2: 103 Questions by type

Elicitations by Type	Count	% of Elicitations	% of all utterances
Sufficient Conditions	89	45%	19%

Conceptual Clarification	39	20%	9%
Review and Relevance	34	17%	7%
Task – Background	14	7%	3%
Identity	13	7%	3%
Already Known	7	4%	2%
Search Tactics Elicitation	2	1%	0%
Total Elicitations	198	100 %	43%

Statements by Type		% of Statements	% of all utterances
Report of Actual Results	140	60%	31%
Status of Results	83	36%	18%
Report on Search Tactics	10	4%	2%
Total Statements	233	100%	51%

Statements by Type		% of Statements	% of all utterances
Report of Actual Results	140	60%	31%
Status of Results	83	36%	18%
Report on Search Tactics	10	4%	2%
Total Statements	233	100%	51%

Other	26		6%

Grand Total	458		100 %

Table 3: Elicitations by Type

	Proposition	Significance
Enablement and Instrumental versus Concept Completion and Quantitative	Hypothesis 1: There will be more elicitations about sufficient conditions for Enablement and Instrument questions than for Concept Completion and Quantitative questions	Not significant
	H1b: There will be more elicitations about the user's task-background for Enablement and Instrumental questions than for Concept Completion and Quantitative questions	$p < .05$
	H1c: There will be more elicitations about the relevance of a proposed answer for Enablement and Instrumental questions than for Concept Completion and Quantitative questions	$p < .05$
	H2: There will be more elicitations about the user's task-background for Enablement questions than for Instrumental questions	Not Significant

Table 4: Results of statistical tests of propositions

The data indicate that contrary to our expectation, who-what-where-when and how-many questions engendered as many inquiries about sufficiency as Enablement and Instrumental questions. Post-hoc

analysis reveals that even though the format of the answer is succinct, this does not always mean it is clear what sort of succinct-answer will be sufficient. Consider GA 389579, a Quantitative question that asks: "*I am looking for information on the number of claims adjusters employed by large insurance companies such as State Farm and Allstate. I need information on the number of field agents that are appraising and adjusting claims. If the number can only be found for one of the above, that is fine. Thanks*". Although the format of the requested answer is a number, i.e. a succinct format, experts made a number of inquiries about the sufficiency of the (partial) information they were able to find. We conclude that short-answer questions evidently involve no less ambiguity about sufficient conditions than longer-answer questions.

On the other hand, it was found that for the short-answer questions, there was less need to inquire whether certain information is relevant, and less need to inquire about the background reasons behind why the person is asking his/her question.

Finally, no difference was found between the two related question types of Instrumental and Enablement, in terms of the need to learn about the user's background.

4.3 Discussion

The World Wide Web continues to present new formats and sources of information, including the question answering services that we study here. This has renewed interest in the extent to which it is possible for human and/or automated systems to respond to users based simply on an input question, or whether it is necessary to get a deeper picture of the user's situation and characteristics before his/her question can be satisfactorily answered. Our work is premised on the presumption that this may depend on the type of question.

The practical application our work is to guide development of automated and semi-automated question answering systems. Fully automated question answering is not yet feasible (2005). "Modelling the user" is a laudable goal, but little is known about how to do user modelling in a way that contributes to the performance of question answering systems. At the same time, Burger et al. (2001) argue that a first step towards answering any question is understanding it, and that classification is one aspect of understanding or interpreting a question. However, research on fully disambiguating a questions' focus is still in its early stages. Our work adopts a combined approach, in which question type helps the human or automated system know which specific aspects of the user model are worth modelling. On the theoretical side, by relating the two fields, our work furthers our understanding of questions, as well as our understanding of user models. Various question typologies have been offered in the literature, but our work suggests a method of defining which typology is "good" or "useful". For the purposes of automated question answering, a useful question typology is one that highlights which elements of the user model are pertinent. If a distinction – such as between short- and long-answer questions – does not tell us which user elements are needed, then the distinction lacks value. In this manner, our work implies a basis for assessing the various question typologies that appear in the literature. In summary, our work aims to further our understanding of basic notions in information science and computer science, such as the nature of questions and the user model, while also providing a focused basis for development of question answering systems with guided user modelling capabilities.

References

Belkin, N. J. (1993). Interaction with texts: information retrieval as information-seeking behavior. Information Retrieval '93. Von der Modellierung zur Anwendung. Konstanz, Universitaetsverlag Konstanz.

Belkin, N. J., H. M. Brooks, et al. (1987). "Knowledge elicitation using discourse analysis." International Journal of Man-Machine Studies **27**: 127-144.

Belkin, N. J., T. Seeger, et al. (1983). "Distributed expert problem treatment as a model for information system analysis and design." Journal of Information Science **5**(5): 153-167.
Brooks, H. M. and N. J. Belkin (1983). Using Discourse Analysis for the Design of Information Retrieval Interaction Mechanisms. SIGIR '83 - 6th annual international ACM SIGIR conference on Research and development in information retrieval ACM**:** 31-47.
Burger, J., C. Cardie, et al. (2001). Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A), NIST.
Cribbie, R. A. and H. J. Keselman (2003). "The effects of nonnormality on parametric, nonparametric and model comparison approaches to pairwise comparisons." Educational and Psychological Measurement **63**: 615-635.
Daniels, P. J., H. M. Brooks, et al. (1985). Using problem structures for driving human-computer dialogues. RIAO '85. Grenoble**:** 645-660.
Graesser, A. C., C. L. McMahen, et al. (1994). Question Asking and Answering. Handbook of Psycholinguistics. M. A. Gernsbacher. San Diego, Academic Press.
Guetzkow, H. (1950). "Unitizing the categorizing problems in coding qualitative data." Journal of Clinical Psychology **6**: 47-58.
Harper, F. M., J. Weinberg, et al. (2010). "Question types in social Q&A sites." First Monday **15**(7).
Lehnert, W. G. (1978). The process of question answering. Hillsdale, NJ, Lawrence Erlbaum Associates.
Neuendorf, K. A. (2002). The Content Analysis Guidebook. Thousand Oaks, Sage.
Pomerantz, J. (2005). "A Conceptual Framework and Open Research Questions for Chat-Based Reference Service." Journal of the American Society for Information Science and Technology **56**(12): 1288-1302.
Ruxton, G. D. (2006). "The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test." Behavioral Ecology **17**(4): 688-690.
Saracevic, T. (1997). "Users lost: reflections on the past, future, and limits of information science." SIGIR Forum **31**(2).
Saracevic, T., P. Kantor, et al. (1988). "A Study of Information Seeking and Retrieving. 1. Background and Methodology." Journal of the American Society for Information Science and Technology **39**(3): 161-176.
Saracevic, T., H. Mokros, et al. (1990). Nature of Interaction Between Users and Intermediaries in Online Searching: A Qualitative Analysis. 53rd Annual Meeting of the American Society for Information Science.
Saracevic, T., A. Spink, et al. (1997). Users and Intermediaries in Information Retrieval: What are They Talking About? User Modeling: 6th International Conference UM97, Springer-Verlag**:** 43-54.
Sears, J. (2001). "Chat Reference Service: An Analysis of one Semester's Data." Issues in Science and Technology Librarianship **32**.
Shachaf, P. (2009). "The paradox of expertise: is the Wikipedia Reference Desk as good as your library?" Journal of Documentation **65**(6): 977-996.
Spink, A., A. Goodrum, et al. (1995). Search Intermediary Elicitations During Mediated Online Searching. 58th ASIS Annual Meeting ASIS '95. T. Kinney. Chicago, IL, ASIS**:** 97-102.
Spink, A., A. Goodrum, et al. (1998). "Elicitation Behavior During Mediated Information Retrieval." Information Processing and Management **34**(2/3): 257-273.
Wu, M.-M. and Y.-H. Liu (2003). "Intermediary's Information Seeking, Inquiring Minds, and Elicitation Styles." Journal of the American Society for Information Science and Technology **54**(12): 1117-1133.