View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by D-Scholarship@Pitt

AUTOMATED QUESTION TRIAGE FOR SOCIAL REFERENCE: A STUDY OF ADOPTING DECISION FACTORS FROM DIGITAL REFERENCE

by

Jong Do Park

B. A., Chung-Ang University, 1996

M. A., Chung-Ang University, 1998

Submitted to the Graduate Faculty of

School of Information Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCES

This dissertation was presented

by

Jong Do Park

It was defended on

December 14, 2012

and approved by

Dr. Xia Lin, Professor, College of Information Science and Technology, Drexel University

Dr. Leanne Bowler, Assistant Professor, School of Information Sciences

Dr. Jung Sun Oh, Assistant Professor, School of Information Sciences

Dissertation Advisor: Dr. Daqing He, Associate Professor, School of Information Sciences

AUTOMATED QUESTION TRIAGE FOR SOCIAL REFERENCE:
A STUDY OF ADOPTING DECISION FACTORS FROM DIGITAL REFERENCE

Jong Do Park, Ph.D.

University of Pittsburgh, 2013

Copyright © by Jong Do Park

2013

AUTOMATED QUESTION TRIAGE FOR SOCIAL REFERENCE:

A STUDY OF ADOPTING DECISION FACTORS FROM DIGITAL REFERENCE

Jong Do Park, Ph.D.

University of Pittsburgh, 2013

The increasing popularity of Social Reference (SR) services has enabled a corresponding growth in the number of users engaging in them as well as in the number of questions submitted to the services. However, the efficiency and quality of the services are being challenged because a large quantity of the questions have not been answered or satisfied for quite a long time. In this dissertation project, I propose using expert finding techniques to construct an automated Question Triage (QT) approach to resolve this problem. QT has been established in Digital Reference (DR) for some time, but it is not available in SR. This means designing an automated QT mechanism for SR is very innovative.

In this project, I first examined important factors affecting triage decisions in DR, and extended this to the SR setting by investigating important factors affecting the decision making of QT in the SR setting. The study was conducted using question-answer pairs collected from Ask Metafilter, a popular SR site. For the evaluation, logistic regression analyses were conducted to examine which factors would significantly affect the performance of predicting relevant answerers to questions.

The study results showed that the user's answering activity is the most important factor affecting the triage decision of SR, followed by the user's general performance in providing good answers and the degree of their interest in the question topic. The proposed algorithm,

implementing these factors for identifying appropriate answerers to the given question, increased the performance of automated QT above the baseline for estimating relevant answerers to questions.

The results of the current study have important implications for research and practice in automated QT for SR. Furthermore, the results will offer insights into designing user-participatory DR systems.

ACKNOWLEDGEMENTS

It would not have been possible to write this doctoral dissertation without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Above all, I would like to thank my wife Dr. Saem Han for her personal support and great patience at all times. My mother, parents-in-law, and brothers have gave me their unequivocal support throughout, as always, for which my mere expression of thanks likewise does not suffice.

I would like to express the deepest appreciation to my advisor, Dr. Daqing He, who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regard to research and scholarship. Without his patience, guidance and persistent help this dissertation would not have been possible.

I also would like to thank my committee members of Dr. Xia Lin, Dr. Leanne Bowler, and Dr. Jung Sun Oh for their support with constructive comments.

Lastly, I am most grateful to my beloved daughter, Joanna, for keeping me smile during the last few miles of my dissertation marathon race.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **VI**

LIST OF TABLES ..... **XIII**

LIST OF FIGURES ..... **XV**

1 INTRODUCTION **1**

1.1 KEY CONCEPTS **2**

1.1.1 Digital Reference (DR) **2**

1.1.2 Expert and expertise **3**

1.1.3 Expert finding **3**

1.1.4 Reference **4**

1.1.5 Social Reference (SR) **4**

1.1.6 Question **5**

1.1.7 Question Answering (QA) **5**

1.1.8 Question Triage (QT) **6**

1.1.9 Subject and topic **6**

1.1.10 Element, attribute, and factor **7**

1.2 BACKGROUND **8**

1.2.1 Reference: the library's approaches to question answering **8**

1.2.2 Automation of reference work: computerize expertise of reference librarians as answerers 9

1.2.3 DR: question asking and answering online 9

1.2.4 SR: user-participatory online reference 10

1.2.5 Challenges in SR services 10

1.2.6 The need for QT 11

1.3 PROBLEM STATEMENT 12

1.4 SCOPE OF THE STUDY 13

1.4.1 Focuses on human-mediated question answering in reference service 13

1.4.2 Focuses on automated QT 13

1.4.3 Focuses on expertise 14

1.4.4 Subject interest, performance, and contribution as expertise 14

1.4.5 Not system-building 15

1.5 RESEARCH GOAL 15

1.6 RESEARCH QUESTIONS 16

1.7 CONTRIBUTION OF THE STUDY 16

2 LITERATURE REVIEW 18

2.1 APPROACHES TO QA 18

2.1.1 Human-mediated QA 18

2.1.2 Approaches to automated QA 20

2.2 APPROACHES TO ACCESS EXPERTISE 28

2.2.1 Expertise seeking: the human-centered approach 28

2.2.2 Expertise retrieval: the system-centered approach 29

2.3 DR **31**

2.3.1 DR Model **32**

2.3.2 QT **34**

2.4 SR **37**

2.4.1 The Social Web **37**

2.4.2 What is SR? **37**

2.4.3 SR Models **38**

2.4.4 Three types of QA Sites **40**

2.4.5 Approach to research on SR **41**

2.5 OVERVIEW OF DIFFERENT APPROACHES TO QA **45**

2.5.1 Key elements of the QA process **45**

2.6 CHAPTER SUMMARY **47**

3 RESEARCH DESIGN AND METHODOLOGY **49**

3.1 INTRODUCTION **49**

3.1.1 Purpose of the study **49**

3.1.2 Research questions **50**

3.1.3 Overview of the research design **50**

3.1.4 Methodological challenges **52**

3.2 PHASE 1: FACTOR IDENTIFICATION FOR QT FOR SR **53**

3.2.1 Identification of the key elements of the QT process **53**

3.2.2 Selection of attributes influencing QT **54**

3.2.3 A framework of expertise: subject interest, performance, and contribution **67**

3.2.4 Research steps in Phase 1 73

3.3 PHASE 2: EXPERIMENT AND EVALUATION 74

3.3.1 Logistic regression analysis 75

3.3.2 Research steps in Phase 2 77

4 EXPERIMENTAL DESIGN 79

4.1 HYPOTHESES 80

4.1.1 Test level of topic interest (x_1) 80

4.1.2 Test the scope of subject area of interest (x_2) 80

4.1.3 Test the scope of topic area of interest (x_3) 81

4.1.4 Test contribution to answer providing (x_4) 81

4.1.5 Test Contribution to Question Submitting (x_5) 81

4.1.6 Test performance to providing relevant answers (x_6) 82

4.1.7 Test Response Time of Answer Providing (x_7) 82

4.1.8 Test quota per day (x_8) 82

4.2 A FRAMEWORK FOR AUTOMATED QT FOR SR 83

4.2.1 Question analysis module 83

4.2.2 Document retrieval module 84

4.2.3 Candidate recommendation module 84

4.3 PROCEDURES 85

4.3.1 Data preparation 85

4.3.2 Data processing 89

4.3.3 Candidate recommendation 102

4.3.4 Generating dataset for logistic regression analysis 106

4.3.5 Evaluation 110

5 RESULTS 115

5.1 PHASE 1: FACTOR IDENTIFICATION FOR QT FOR SR 115

5.1.1 Commonality and differences between DR and SR 115

5.1.2 Selecting attributes affecting to triage decision 116

5.1.3 Framework of user expertise 117

5.2 PHASE 2: EXPERIMENT AND EVALUATION 118

5.2.1 The logistic regression model 118

5.2.2 Imbalanced dataset 119

5.2.3 Data Normalization 120

5.2.4 Training 121

5.2.5 Selecting the best fit logistic model 130

6 DISCUSSION AND IMPLICATIONS 139

6.1 INTRODUCTION 139

6.2 THE AUTOMATED QT PROCESS 140

6.3 REPRESENTATION OF USER EXPERTISE 140

6.4 ATTRIBUTES THAT AFFECT THE TRIAGE PROCESS 140

6.4.1 Attributes of the question 141

6.4.2 Attributes of the answerer 142

6.4.3 Summary 146

6.5 IMPLICATIONS FOR RESEARCH DESIGN 147

6.5.1 Theoretical Implications 147

6.5.2 Practical implications 148

7 CONCLUSION 149

7.1 REFLECTIONS ON ACCOMPLISHING THE RESEARCH GOALS 149

7.1.1 RQ1. What are the elements of automated QT for SR, and are they different from that of DR? 149

7.1.2 RQ2. What are the factors that affect the performance of automated QT for SR? 150

7.1.3 Summary 151

7.2 LIMITATIONS OF THE STUDY 152

7.3 REMAINING QUESTIONS 152

7.4 FUTURE WORK 153

7.4.1 Selection of identifying additional factors affecting QT for SR 153

7.4.2 Question recommendation for SR sites 154

7.4.3 Study on automated question filtering 154

7.4.4 Study on user-participated library reference (SR) 154

BIBLIOGRAPHY 156

LIST OF TABLES

Table 1. Types of information systems and their approaches to QA 21

Table 2. Focuses for finding best answerer or experts 44

Table 3. Elements of QA process 46

Table 4. Fifteen factors affecting QT for DR (Pomerantz, 2003) 55

Table 5. Mapping the attribute of the question from DR into SR 57

Table 6. Mapping the attribute of the answerer from DR into SR 58

Table 7. Mapping the service attributes onto the answerer 59

Table 8. Mapping the triage factors of DR to SR 61

Table 9. Comparison of subject with topic 63

Table 10. The framework of expertise 70

Table 11. Research steps in phase 1 73

Table 12. Variables in the logistic regression analysis 77

Table 13. Research step in phase 2: building new models for QT in SR 78

Table 14. Summary of the dataset 87

Table 15. Summary of categories selected 88

Table 16. Summary of dataset statistics for training and testing 89

Table 17. Most frequent top 10 tags in 5 categories chosen. 91

Table 18. Selected number of topics for each category 94

Table 19. Example of the first four topics of computer-internet category using LDA model..... 97

Table 20. Variables used in the evaluation 112

Table 21. Logistic regression analysis of TAG-B0F1N20 dataset using a single factor for baseline 119

Table 22. Ratio between relevant cases and non-relevant case in the training dataset..... 120

Table 23. Comparison of data type: raw, log, and z-transformation 121

Table 24. Performance comparison (MAP) of each model that used different set of training data with different conditions. 124

Table 25. Performance comparison between F0 and F1 models 126

Table 26. Performance comparison between the proposed model and baseline (training dataset) 128

Table 27. Performance comparison between the proposed model and baseline (testing dataset) 129

Table 28. Logistic regression analysis of factors affecting to the estimation of relevant answerers (Dataset of TAG-B0F0N20)..... 132

LIST OF FIGURES

Figure 1. Conceptual framework of subject and topic 7

Figure 2. Overview of a generic QA architecture (Umbert, 2012) 26

Figure 3. General DR model (Pomerantz, Nicholson, Belanger, & David Lankes, 2004) 34

Figure 4. Model of answering behaviors in social Q&A: motivations & strategies (Oh, 2010) .. 42

Figure 5. Conceptual frameworks of different approaches to QA 48

Figure 6. An overview of the research methodology 51

Figure 7. Proposed framework of QA process for SR 54

Figure 8. Map of user's subject areas of interest 64

Figure 9. Map of user's topic areas of interest 64

Figure 10. A process framework for automated question triage for social reference 84

Figure 11. Overview of the experiment procedure 86

Figure 12. Number of topics and perplexity in education category 93

Figure 13. Example of document-topic distribution: question 3905, question 3947 95

Figure 14. User-topic distribution (user 14648 and 1104) 96

Figure 15. Pseudo code for determining topics of the question (topic-clustering approach) 104

Figure 16. Pseudo code for candidate selection - document-based approach 106

Figure 17. Raw distribution of answering activity 108

Figure 18. Z-distribution of answering activity 109

Figure 19. Changes in MAP as candidate pool size (N) is increasing (baseline) 123

1 INTRODUCTION

Information seeking is a common activity of human life. People sometimes rely on other human beings to solve their information problems. A familiar type of social interaction of information seeking is question asking. Question asking and answering are universal features of human communication (Goldman, 1999). The main purpose of asking questions is to learn the answer from the respondent to meet the questioner's information need.

The development of Web 2.0 technologies, often referred to as the *participatory Web*, resulted in the growth of Social Reference (SR) services that enable users to interact with each other in the form of question asking and answering in online communities. The increasing popularity of SR services, in recent years, has enabled a corresponding growth in the number of users of SR services on the Web. This increase in the use of SR services has led to increases in the number of questions received by these services, thus the efficiency and quality of the services have become issues in this field.

In the field of libraries, reference librarians have practiced Question Triage (QT), or question routing, especially for Digital Reference (DR) to be able to handle increasingly large number of questions received, since the quality of the answers provided is directly affected by the amount of questions assigned to a reference service or an expert in DR services. Similarly,

there is an obvious need to investigate QT for SR services, in order to increase efficiency and to improve the quality of the answers provided.

This research seeks to address the need for automatic QT for SR for the purpose of improving its practice, as well as to inform the design of SR systems and services that exploit users' expertise. Specifically, this study investigates how to identify the best candidates to respond to each question.

1.1 KEY CONCEPTS

In this section, key concepts used in this dissertation are defined for better understanding of the study.

1.1.1 Digital Reference (DR)

The term *digital reference* is used here as a generic term for question answering interactions between users and reference librarians that occur entirely online through Internet communication modes such as chat, e-mail, instant messaging and Web forms. The terminology in this field has not yet stabilized and there are many other variants in use such as *virtual reference*, *live online reference*, *e-mail reference*, and *chat reference* but all of these terms can be considered here to be subsumed within the broader category of DR. In the early- to mid-1990s, DR services began to appear that were not affiliated with any library. In the literature, these DR services are often

referred to as *ask-a* or *expert* services such AllExperts[1] and Ask A Scientist[2]. In this study, DR refers to library-affiliated online reference service, and the term *Ask-A* is used to mean a new type of online reference services that are not affiliated with any library.

1.1.2 Expert and expertise

The term *expert* is used in this research to refer a person who has expertise in a particular area. In this research, an expert is defined as a user who is able to provide relevant answers to the question submitted. The main purpose of using the term is to refer to relevant answerers among users to a specific question in SR.

The term *expertise* is used in this research to explain the specialty of an expert. In order to define *expertise*, *knowledge*, *skill*, and *experience* are often used as the aspects of expertise (Ericsson, 2007). In this study, *knowledge*, *skill* and *experience* are reworded as *subject interest*, *performance*, and *contribution*, to refer answerers' expertise in SR services.

1.1.3 Expert finding

Assessing expertise of an expert is a key to success of QT for the study. While researchers in the field of library and information science have provided theoretical foundation for assessing expertise of a specific expert or service for QT, other practical approaches to assessing expertise of a specific expert have been made in the field of information retrieval. *Expert finding* is one of information retrieval tasks. The goal of expert finding task is to identify experts, who are

[1] http://www.allexperts.com/
[2] http://www.hhmi.org/askascientist/

relevant to a given specific topic, and recommend them as candidates. While factors that can be used for assessing expertise of an expert are often limited to subject or topic of interest of the expert, those approaches have implication to QT for SR in that the expertise of a user, as an expert on a specific topic, is assessed automatically.

1.1.4 Reference

The term *reference* is used in this study as a generic term for question-answering interactions between users and experts, usually reference librarians, while its original concept is much broader than this. This definition employs two key concepts: (1) *question answering* and (2) *human-mediated*. The first concept is employed for two reasons: question answering is one of main activity or role in the process; the main focus of the research is question answering. The second concept of *human-mediate* is employed to distinguish it from automated question answering system, in which human is not involved in answering process.

1.1.5 Social Reference (SR)

The term "*social reference* (Shachaf, 2010)" is used in the study to refer to online QA services that are provided by communities of volunteers on question and answer (Q&A) sites. This usage has been selected in this study as it conveys the concept of reference in itself, and the researcher seeks to do research in the context of reference. The terminology in this emerging field has not yet stabilized and there are many other variants of the term in use such as "*social Q&A* (Gazan, 2007; Harper, Moy, & Konstan, 2009; Kim, Oh, & Oh, 2007)", "*community Q&A* (Lee, Rodrigues, Kazai, Milic-Frayling, & Ignjatovic, 2009; Li & King, 2010; Nam, Ackerman, &

Adamic, 2009; Shah & Pomerantz, 2010)" or "*community-based QA* (Jeon, Croft, & Lee, 2005; Shah & Pomerantz, 2010)". SR is similar to library reference, but at the same time, it is significantly different from the traditional (and digital) dyadic reference encounter since it involves a collaborative group effort and uses wikis and other Web 2.0 infrastructure (Shachaf, 2010).

1.1.6 Question

The term *question* refers to a very heterogeneous concept (Flammer, 1981). There are open questions and closed questions, real questions and rhetorical questions, wh-questions and disjunctive questions, spoken questions and written questions, and so on. In this study, the term *question* is used to mean information-seeking question that is "an observable behavioral act reflecting information need (Horne, 1983)." In this sense, questions are interpreted as indicators of the asker's information need.

1.1.7 Question Answering (QA)

The term *question answering* is used here as a generic term to mean the activity of giving answers to a question. While the concept of question answering is limited to *automated* answering a question posed in natural language in the field of question answering, it is not necessary to limit its concept to *automated*.

1.1.8 Question Triage (QT)

The term *question triage* is used here to mean the assignment of a question to an expert who has ability to provide an answer to the question. In the setting of library reference, the expert has usually been regarded as a reference librarian or a subject specialist who provides answers to the question submitted by the user in libraries. The term *triage* is originated from the French verb *trier*, meaning to select, divide, or sort. The word *triage* means “the action of assorting according to quality” ("Triage," n.d.). The principles of the term have been in use for centuries. The original meaning of the term is purported to be grading agricultural products, such as coffee beans (Boyd, 1983). The more commonly understood meaning of the word these days, the assignment of injured persons to treatment according to the urgency of their injuries, was acquired during the World War I (Mitchell, 2008). Regardless of its medical orientation, the term has been used in the domain of libraries to refer question routing to a reference librarian or another service. The Virtual Reference Desk, a project dedicated to the advancement of DR, has defined triage as the assignment of a question to a reference or subject expert answerer (Pomerantz, Nicholson, & Lankes, 2003). Pomerantz (Pomerantz, 2003) defined it as “the sorting out and classification of questions to determine the proper place of treatment,” replacing the word *casualties* with *questions* in the definition of triage that is borrowed from the On-Line Medical Dictionary.

1.1.9 Subject and topic

While subject and topic are often interchangeably used to point to the matter being considered in a speech or written composition ("Subject," n.d.), they are used in contrasting ways in this study, for an intended purpose. In this study, the term *subject* is used to refer a general or broad concept

of the primary matter of discussion, while the term *topic* is used as a sub-class of the subject, referring to a specific or narrow area within the primary matter of discussion. In this study, a category to which users post their questions in SR is called the subject or subject area, and a sub-category in that category would be called a topic or topic area.

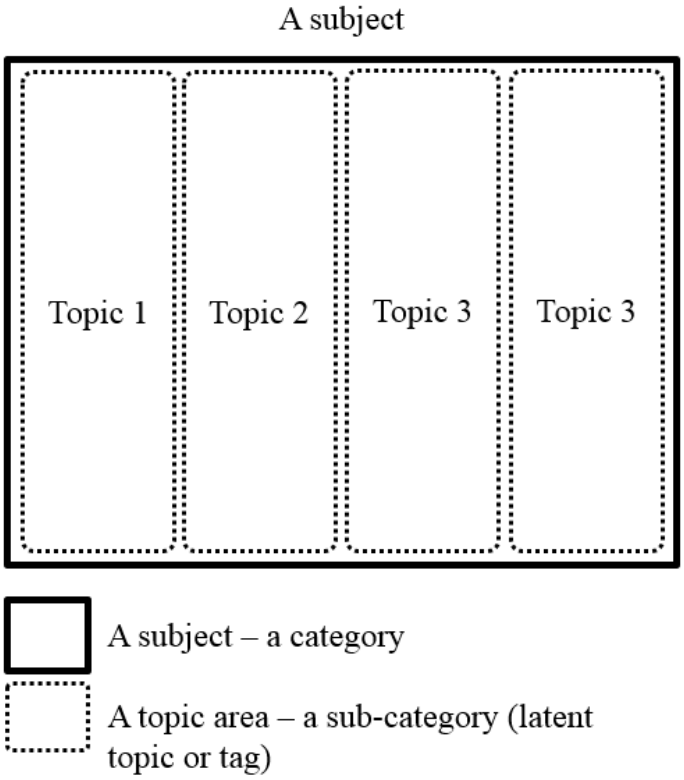

Figure 1. Conceptual framework of subject and topic

1.1.10 Element, attribute, and factor

In this study, the term *element* and *attribute* are used in order to address factor identification within QT. An *element* is defined here as a basic building block in the process of QT. An

attribute is defined here as a feature that is characteristic to an element. The term *factor* used in this study refers to one of the statistically important *attributes* of *elements* that contribute to the decision-making of QT.

1.2 BACKGROUND

This section introduces important background information for a better understanding of the scope of this research.

1.2.1 Reference: the library's approaches to question answering

The library has been a well-known information seeking place for centuries. Libraries have provided reference services that enable users to seek information by asking questions to library staff, human experts, to satisfy their information need. Reference librarians have practiced well as reference services, and researchers in the field of library and information science have developed theoretical frameworks for this reference since Green laid the foundation of modern reference work in 1876 (Green, 1876). Reference service is the personal assistance provided to users in pursuit of information (Bunge, 1999). This service can range from general directional questions to specific in-depth queries. Reference librarians are able to help users identify answers to specific questions or interpret information in specific library resources. Thus, reference librarians are the main experts who provide answers to the users' questions in the library reference setting.

1.2.2 Automation of reference work: computerize expertise of reference librarians as answerers

Throughout the history of libraries, the evolution of reference work has been greatly influenced by the transformation of the library environment, such as the advent of new technologies and the change in users' information-seeking behavior. Librarians have tried to automate reference work by using computing power, and have introduced expert systems, or knowledge-based systems, for reference work in the late 1960's (Shera, 1964). Approaches to the automation of reference work are meaningful in that researchers and system builders have attempted to capture reference librarians' expertise or knowledge when designing these knowledge-based systems for reference work.

1.2.3 DR: question asking and answering online

With the rise of the Internet and World Wide Web to popularity in the mid-1990's, reference services evolved to DR, or virtual reference, in order to reach out to new types of library users—the remote online users who want to access the library's services via the Internet and submit a reference question through online communication tools, such as email or chat. For the study, 'digital reference' is defined as a library-affiliated online reference service; thus it is different from other online reference services that are not affiliated with any library. Some of the theoretical background of QT has been established in the setting of DR rather than desk reference: it is a 'question' that is forwarded to another reference service or expert in a DR service whereas a 'user' is referred to another reference service or expert in desk reference. Though reference services have evolved quite a bit while moving from desk reference to DR, the

central purpose of reference services—to answer and provide resources to enable users to answer their own questions—and the main role of reference librarians—as human experts who provide answers—has always been constant.

1.2.4 SR: user-participatory online reference

On the other hand, the development of Web 2.0 technologies, often referred to as the *participatory Web*, resulted in the growth of SR services in which users can interact with each other in the form of question asking and answering in online communities. This new type of online reference services can be differentiated in that (1) the source of provided answers is other users, self-declared experts; and (2) users may be involved in two distinctive roles, questioner and answerer, in the process of question answering. Since research on SR is in an early stage of development, it is necessary for researchers in this domain to better understand this new type of reference service, user-participatory reference service as being congruent with the broader context of reference services, while in addition being able to take advantage of the users' expertise, in order to design optimal, technologically advanced reference services.

1.2.5 Challenges in SR services

The increasing popularity of SR services has enabled a corresponding growth in the number of users of SR services on the Web. This increase in the use of SR services has led to increases in the number of questions received by these services, thus the efficiency and quality of the services are challenged: SR sites cannot solve users' questions efficiently (Li & King, 2010); and the quality of the submitted answers is uneven (Bouguessa, Dumoulin, & Wang, 2008; Jurczyk &

Agichtein, 2007). According to a blog of Yahoo Answers, more than 800,000 questions are posted daily, and more than 1 billion answers were served as of May 2010 (Yahoo! Answers Team, 2010). The increase of its popularity is amazing. However, a large number of the questions are not answered or are waiting to be answered. There may be many explanations for this phenomenon; for example, it may be a wrong, unclear, or repeated question that cannot be answered, or the question may be posted in the wrong category. However, many of them could still be answered if the question were routed to experts who can give the answer to the question rather than just remaining on an extensive list of questions, waiting for several days to be answered by a random person.

In order to get relevant answers to questions, we must know not only the questioners' information needs, but also who is able and willing to answer their question. Otherwise their information need can be satisfied only by accident (Flammer, 1981).

1.2.6 The need for QT

In the field of libraries, reference librarians have faced similar problems: limited resources (human, resources, and budget) as reference questions increase, especially within DR. Reference librarians for DR services have begun practicing QT in order to handle the increasingly large number of questions received and to increase the quality of the answers provided, thus maximizing availability while improving the quality of the service. Likewise, there is an increased and immediate need for QT in SR services. In order to design algorithms to route a question to relevant users who can answer the question, it is necessary to assess other users' expertise that is relevant to the question.

1.3 PROBLEM STATEMENT

SR services, which enable users to ask and answer questions within a community for their own needs, have emerged as popular, and often effective, means of information-seeking on the web. In the framework of SR services, information seekers can obtain specific answers to their questions by posting these questions to the community, expecting the questions to be answered by other participants.

As SR sites, such as Yahoo! Answers, are growing rapidly in recent years, users of such sites already have submitted millions of questions and received hundreds of millions of answers from other participants. However, it may also take hours, and sometimes days, until a satisfactory answer is posted, and unanswered questions still exist within a longer period of time. Some researchers have focused on this problem and observed that SR sites cannot solve users' questions efficiently: in Yahoo! Answers only 17.6% of questions receive satisfied answers within 48 hours; nearly 20% of unresolved questions receive no response (Li & King, 2010). There may be several reasons for this failure or lag in answering: for example, questions may be wrongly worded, unclear, or spam; questions may posted in the wrong place so that they are not exposed to relevant answerers, etc.

This dissertation research addresses the need for QT, or question routing, for SR, for the purpose of improving the practice and informing the design of SR services.

1.4 SCOPE OF THE STUDY

The research area: "QT for SR" is too broad an area to explore as the subject of a dissertation, since it has many aspects to be studied. For example, this topic is covered in many other fields such as library and information science, information retrieval, sociology, and psychology. Limiting it to library and information, it still has several sub-fields such as information retrieval, information-seeking behavior, etc. To make this study achievable, the boundary of the study had to be narrowed down further.

1.4.1 Focuses on human-mediated question answering in reference service

In the study, the term 'reference' is one of the core terms that explain the direction of the research. As discussed earlier, the concept of reference, in this study, conveys two meanings: (1) human-mediated and (2) question answering. So, the main focus of the study is on human-mediated question answering in reference service. Thus, other reference services, such as collection development and preparation of bibliographies, are beyond the range of the interests of this study. In addition, automated question-answering systems, in which a human is not involved in the answering process is outside of the focus of this study.

1.4.2 Focuses on automated QT

The main focus of the study is on automated QT for SR. The topic of QT has been well practiced in the libraries, and is often studied in the context of DR in the domain of library and information science. This study seeks to introduce the concept of QT to SR, borrowing some ideas of QT

from DR. As practiced in libraries, QT can be performed by a human expert (manually) or a machine (automatically). This study focuses on the automation of QT, using state-of-the-art expert-finding techniques for QT.

1.4.3 Focuses on expertise

The process of QT is involved within several stages, for example, question filtering, question assigning, and question answering. In the filtering stage, the following needs to be determined: which questions the service should accept; which questions can be answered by the answerers; and which questions are out of scope for the service. In the question assigning stage, candidate answerers are located and the question is forwarded to an expert or answerer. In the answering stage, the answer is formulated by the answerer and presented to the questioner. In this study, the researcher focuses on the question assigning stage. In order to assign a question to an expert, in practice at least two points need to be considered: expertise and availability of the expert. In this study, the researcher mainly focuses on the use of expertise for QT, assuming that all experts are always available online at the time of assigning a question.

1.4.4 Subject interest, performance, and contribution as expertise

Indeed, how to assess and profile the responders' expertise is the key to success in QT. In the context of DR, QT can increase service quality by increasing the possibility of getting fast and relevant answers to the question from experts or the best reference services. In the field of libraries, the expertise of an answerer, or reference service, is assessed using its collection of information, experience, subject specialty, etc. In this study, in order to locate an expert who can

provide answers to questions, the following three basic questions are asked: (1) Who is interested in the subject and topic of the question submitted? (2) Who is able to provide 'relevant' answers to the question submitted? and (3) Who is able to participate in providing answers to the question submitted? Further defined, the concepts behind those questions are set to address expertise as: (1) subject interest, (2) performance, and (3) past contribution to the reference service.

1.4.5 Not system-building

This study does not seek to build a QT system for SR, but rather to develop and test some expertise-assessing algorithms, which can be used to develop an automatic QT system in the future, to demonstrate the possibility of automated QT in SR.

1.5 RESEARCH GOAL

The purpose of this research is to present a foundation for incorporating the value of QT into SR. The goal of this research is to describe and gain a further understanding of automated QT in the context of SR. This research has three specific objectives within that goal:

1) to investigate key elements that are currently used,
2) to identify new features and requirements, and
3) to design new conceptual and practical models.

1.6 RESEARCH QUESTIONS

In order to meet the purpose and objectives of this research, the following research questions need to be asked:

1. What are the essential elements of automated QT for SR, and are they different from that of DR?
2. What are the factors that affect the performance of automated QT for SR?

The first question is asked in order to achieve the first objective of this study, seeking to gain an understanding of QT in the context of SR. In order to develop this research, the researcher first tried to investigate the existing models of QT for DR, since researchers in the field of DR have presented the conceptual foundations of QT. It is useful to explore existing models and extend them in the context of SR.

The second question is asked in order to achieve the second and third objectives of this study. While the first question focuses on the elements that exist in the process of QT, the second question focuses on the characteristics of each element in QT, seeking to discern guidelines to be used when designing new conceptual and practical models of automated QT for SR, thus achieving the last objective.

1.7 CONTRIBUTION OF THE STUDY

This study seeks to make contributions to research and practice in automated QT for SR.

Contributions at the research level include:

- augmenting our understanding of SR which allow users to participate in question answering services;
- the development of a framework that illustrates the process of QT for a SR; and
- the identification of areas in need of further investigation, within the field of library and information science, for developing user-participatory reference.

At practice level, this study:

- provides helpful information to improve existing SR services by implementing automated QT;
- gives insight for designing new systems that exploit user expertise for future user-participatory reference (SR) in libraries; and
- provides a set of algorithms that can be utilized as the basis for designing and building a system to automate the process of QT.

2 LITERATURE REVIEW

This chapter provides an overview of the bodies of literature that inform this study. There are six areas of literature drawn upon for this study: reference, DR, QT, expert finding, question answering, and SR. These literatures either better define the study's area of investigation - QT in SR - or aid in the creation of the study's conceptual framework. The conceptual framework of the literature review for this research is visualized in Figure 5. The main focus of the conceptual framework is question answering process.

2.1 APPROACHES TO QA

2.1.1 Human-mediated QA

2.1.1.1 Reference as a service for QA

Green laid the foundations of modern reference work in his highly influential article, "Personal relations between librarians and readers," which was published in *Library Journal* in 1876. Reference services were developed in the late 19th century to help readers find relevant and

pertinent information sources that they required (Rothstein, 1953). Since Green's work, the practices involved in providing service have been refined and changed in various ways.

The literature supports question answering as one of the key roles of reference librarians in library reference service. Green pointed out "answering patrons' questions" as one of reference librarians' goals:

- teaching patrons about the library's functions and resources, as well as how to utilize them;
- answering patrons' questions;
- helping patrons select good reading material and;
- promoting the library to the greater community (Green, 1876).

Low (1996) categorized the duties with which reference services is comprised into three types of services:

- informational services - in which an answer is provided to a user's question;
- library instruction - teaches the patron how to locate information, what reference tools are available and how to use them or teach them how the library is organized and how to use it in the most effective manner;
- readers' advisory services - which assist the patron by suggesting specific titles and subject headings which will be appropriate for the user's particular interest.

The Reference and User Services Association (RUSA) of the American Library Association (ALA) state that "Reference Transactions are information consultations in which

library staff recommend, interpret, evaluate, and/or use information resources to help others to meet particular information needs" (RUSA, 2008).

Since Green defined the relations between librarians and library users, the concept of human-mediated assistance provided by reference librarians has been considered in various ways, but there has been consensus about the central purpose of reference service, which is to answer and provide resources to enable patrons to answer their own questions.

DR and SR can also be categorized into this approach, but those two will be addressed later in separated sections in order to discuss them in detail.

2.1.2 Approaches to automated QA

The idea of answering questions with automated computing power has existed ever since computer technologies began to appear widely in academic and corporate settings. In the late 1960's, expert systems, or knowledge-based systems, emerged when computer-based networks began to appear widely in academic and corporate settings, and libraries began to be interested in using them to facilitate and automate library work.

Table 1 presents an overview comparing the different types of systems (both human and automated) that exist in the field of libraries and information retrieval, and the approaches that these systems take to answering questions.

Table 1. Types of information systems and their approaches to QA

Type of System	Approach to QA	Result
Information retrieval system	Document retrieval	Document
Knowledge-based system for reference	Document retrieval	Document
Desk reference service	Question answering Referring users	Answer, or document Another service
DR service	Question answering, Question triage	Answer, or document Another service
QA system	Answer retrieval	Answer
Expert finding system	Expert retrieval	List of experts

2.1.2.1 Knowledge-based systems for reference

Shera (1964) appears to be one of the earliest researchers who published works discussing the possibility of automation in the service of reference librarianship. In his article, he focuses on establishing the proper role of decision theory and library automation in reference work mentioning "the conviction that automation can raise the intellectual level of the reference librarian does not imply a belief that a machine can make a literature specialist out of a simple button pusher." The first projected attempts to automate a reference service were supposed to assist reference librarians in the selection of which biographical reference book to use (Weil, 1968). Since such earliest systems were programmed using expert knowledge to achieve the

goals, they are often referred to as expert systems or knowledge-based systems. In the literature, the definition of expert systems is varied: “a computer program using expert knowledge to attain high levels of performance in an narrow problem domain (Waterman, 1986),” and “a computer system that uses a representation of human expertise in an specialist domain in order to perform functions similar to those normally performed by a human expert in that domain” (Goodall, 1985). Yet, the British Computer Society (BCS) has provided the most formal definition of it: “An expert system is regarded as the embodiment within a computer of a knowledge-based component from an expert skill in such a form that system can offer intelligent advice or make an intelligent decision about a processing function. A desirable additional characteristic, which many would consider fundamental, is the capability of the system, on demand, to justify its own line of reasoning in a manner directly intelligible to the inquirer. The style adopted to attain these characteristics is rule-based programming” (Naylor, 1987). While, some authors in the field of library and information science indicate a level of complexity with the related terms of expert systems, such as ‘knowledge systems’, ‘knowledge-based systems’, or ‘reference advisory systems’, there is a consensus that ‘knowledge-based system’ is a broader term than just ‘expert systems’ (Bohr, 1995). The features of such expert systems can be summarized as the following:

- It is limited to a specific domain.
- Since it is based on computer, all actions in the domain have to be decomposed into a set of algorithms, which the system employs to perform whatever actions it is designed to perform.
- It utilizes knowledge representation of experts in which those algorithms are applied.

Richardson (1995) provides a comprehensive review of knowledge-based systems for reference work, especially focusing on QA, which is a complex and difficult task within reference services. In his work, the author reviewed 56 knowledge-based systems for reference work, which were selected throughout the literature, to identify the prototype and commercial knowledge-based systems in the field of general reference work. The features of such knowledge-based system for reference work can be summarized:

- The domain of the system is limited to reference work.
- Rule-based algorithms are often applied.
- It uses the knowledge representation of reference librarians to guide queries to reference resources.
- Most of them are developed as a prototype system.

Because of these limitations, knowledge-based systems for reference work in the era of early library automation were not as effective. Some people believe that they can improve decision making; support more consistent decision making; … capture rare or dispersed knowledge (Coats, 1988), since they have been built by one or two domain experts or knowledge engineers and not based on a theoretical model of reference transactions (Richardson, 1995).

While there is some limitation in earlier approaches to automate reference work in libraries, it has some pertinent implications, because researchers tried to capture knowledge from reference librarians, and to represent it using computing power. For this study, it needs to be remarked that such systems do not actually answer questions, however, but rather match a question with an information source that will hopefully contain an answer.

2.1.2.2 QA system

A *QA system* provides direct answers to user questions by consulting its knowledge base. Since the early days of artificial intelligence in the 1960's, researchers have been fascinated with answering natural language questions. The difficulty of natural language processing (NLP), however, has limited the scope of QA to domain-specific expert systems. The combination of Web growth, improvements in information technology, and explosive demands for better information access has reignited the interest in QA systems. The availability of huge document collections, such as the Web, combined with development in information retrieval (IR) and NLP techniques, has attracted the development of a special class of QA systems that answers natural language questions by consulting a repository of documents (Voorhees & Tice, 2000). The achievement of these systems is that it's possible to answer questions over the Web. A QA system utilizing this resource has the potential to answer questions about a wide variety of topics, and will constantly be kept up to date with the Web itself. It therefore, makes sense to build QA systems that can scale up to the Web.

The Text Retrieval Conference (TREC) is a renowned research arena for QA systems. In the late 1990's, researchers in this field were interested in providing direct answers to user questions, and TREC initiated a QA track to take a step closer to information retrieval over document retrieval. The goal of the TREC QA track is to foster research on systems that directly return answers, rather than documents containing answers, in response to a natural language question. Since then, the track has steadily expanded both the type and difficulty of the questions asked:

- Factoid questions: a fact-based, short answer question,

- List questions: ask for different answer instances that satisfy the information need,
- Definition questions: ask for interesting information about a particular person or thing,
- Other questions: ask for interesting information about the target that is not covered by the preceding questions, and
- Complex, interactive QA (ciQA) tasks: “Complex” is concerned with the information need of the question. Information needs, or topics, which consist of a template and a narrative, provide additional context. A template provides the question in a canonical form, and a narrative elaborates on what the user is looking for; “interactive” is about interacting with the system, in which a human assessor is allowed to interact with the system for five minutes for evaluation.

Since TREC aims to evaluate IR systems developed at different institutions for given tasks, to evaluate the achievement of the goals, every prototype of QA systems in TREC is implemented differently. There are, however, important similarities among those systems. The majority of the systems consist of three distinct components:

1. question analysis,
2. document or passage retrieval, and
3. answer extraction.

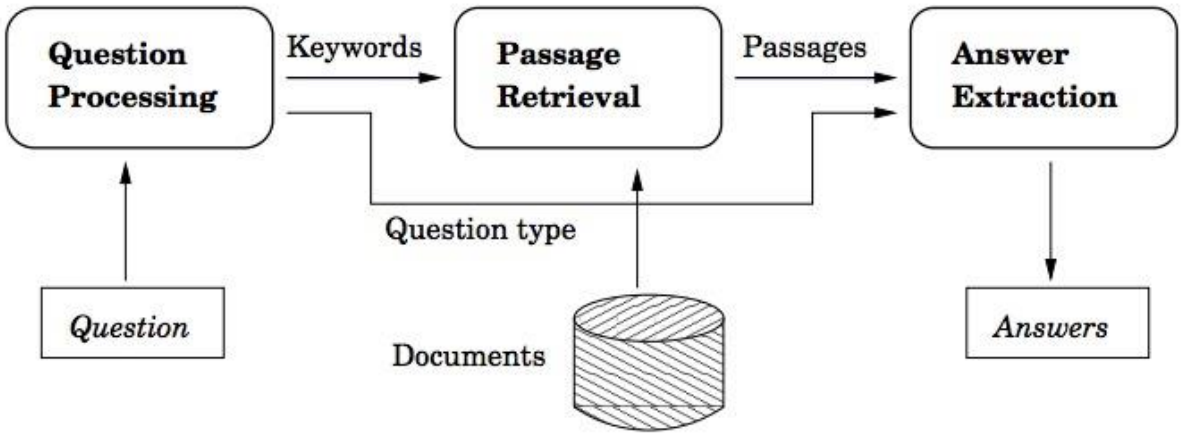

Figure 2. Overview of a generic QA architecture (Umbert, 2012)

While this three-component architecture, as shown in Figure 2, is widely used in the literature, Pomerantz categorized the processes occurred in QA systems into four, also based on the literature:

1. Question processing: This step consists of two sub-processes:
    a. Logical representation of a question, as the opposite meaning to natural language: this could be used as the query to a search engine or a database (Diekema et al., 2000), or a format unique to the QA system (Moldovan et al., 1999); and
    b. Recognition of the question focus, to determine the subject or forms of the expected answer.
2. Document processing: Documents are retrieved from the corpus of all documents as in a traditional IR system and returned documents are ranked according to specific algorithms.

3. Paragraph or segment finding: Documents are then broken up into paragraphs or segments of some size in order to retrieve the segments of the retrieved documents that best match the question.
4. Answer finding: Paragraphs or segments are broken up into even smaller segments. The keywords in the logical representation of the question and the question focus are searched for within a window size of 50 or 250 bytes (default answer length of TREC QA task). Next, these smaller segments are ranked according to system-specific algorithms. The top-ranked segment is considered to be the answer.

For this study, it should be noted that QA systems are:

- different from traditional IR systems in that they provides direct answers rather than a set of documents that contain answers in themselves;
- different from SR services in that the answer provided is extracted from documents, not from human experts in a SR service.

It is obvious that human experts can provide better answers than machines such as QA systems. However, it is good idea to use QA systems to provide answers in response to users' simple factoid questions, which are fact-based questions. The use of QA systems for reference services under limited environments, such as for answering factoid questions or ready reference questions, may lessen the load of work pressure for reference librarians.

2.2 APPROACHES TO ACCESS EXPERTISE

Research on accessing expertise can be classified into two types of approaches in terms of its focus: the human-centered approach and the system-centered approach. Some researchers refer to these as (1) expertise seeking, which studies human behavior in expertise seeking, and (2) expertise retrieval, which focuses on identifying topical matches between needs for expertise and the content of documents associated with candidate experts (Hofmann, Balog, Bogers, & de Rijke, 2010). Since these terms represent the difference between those concepts clearly as the difference between information seeking and information retrieval, these terms are adopted here.

2.2.1 Expertise seeking: the human-centered approach

One of the goals in the human-centered approach has been to develop models of how human information sources are used. In the human-centered approach, several studies have investigated factors that may affect decisions of what expert to contact or recommend. In a study of selection criteria for human information sources, Woudstra and van den Hooff (2008) found that source quality is the most dominant factor in the selection of human information sources. For example, topic of knowledge, as a factor of quality-related factors, was mentioned 51% as a factor for the consideration by the participants. While content-based features, such as topic of knowledge, are the most important, Hofmann, Balog et al. (2010) studied contextual factors for finding experts and found that contextual factors, such as media experience and organizational structure, are also considered when searching for experts. Borgatti and Cross (2003) identified that knowing about an expert's knowledge, valuing that knowledge, and being able to gain access to an expert's knowledge influence decisions of what expert to contact for help. Some researchers have

investigated the use of social network information to search for experts (Shami, Ehrlich, & Millen, 2008; Terveen & McDonald, 2005; Zhang, Tang, & Li, 2010).

2.2.2 Expertise retrieval: the system-centered approach

Expertise retrieval seeks to develop algorithms that can support the search for expertise using information retrieval techniques. The topic has been addressed as the expert finding task of the enterprise track at the annual Text Retrieval Conference (TREC) from 2005 to 2008 (Bailey, Craswell, de Vries, & Soboroff, 2007; Balog et al., 2008; Craswell, de Vries, & Soboroff, 2005; Soboroff, de Vries, & Craswell, 2006). The concept of expert finding is not a new one. For example, researchers in academia often refer to citation indexes to identify publications, which are associated with authors who have the strongest research impact. The literature on expert finding systems supports the fact that a lot of attention on expert finding systems developed in the context of enterprise settings (Balog, Azzopardi, & De Rijke, 2006; Fang & Zhai, 2007; Petkova & Croft, 2006). This trend seems to be affected by the increased interests in the expert finding task at TREC. The expert search task has been included as one of the tasks in the TREC Enterprise Track since 2005, and the goal is to find a list of candidates who are experts in the specific topic with supporting documents. This has encouraged researchers in the field to develop expert finding systems by providing enterprise corpora to them.

Human resources are valuable assets to an organization since they possess a range of knowledge and expertise that can benefit the organization. Expert finding, also called expertise location, has become a challenging problem in the enterprise environment, in order to discover experts and acquire their knowledge. People in organizations often must find others with knowledge and information to a specific topic in order to solve their problems and accomplish

their work goals. Expert Finding Systems (EFS) that enable users to locate individuals or communities of expertise can rapidly improve the efficiency and effectiveness of organizations by connecting them easily.

A common approach to expertise retrieval that has been experimented with is to use information retrieval techniques based on documents associated with experts. For example, evidence from a document is used to estimate associations between experts and documents or experts and topics (Balog et al., 2006; Balog, Azzopardi, & de Rijke, 2009; Serdyukov & Hiemstra, 2008b). A challenge in content-based expertise retrieval is that systems need to go beyond document retrieval since the goal of expertise retrieval is to retrieve experts instead of documents (Hofmann et al., 2010). Serdyukov, Rode, & Hiemstra (2008) introduced graph-based algorithms that increase the evidence of expertise via multi-steps of related documents and experts. Karimzadehgan, White, & Richardson (2009) leveraged the organizational hierarchy (depicting relationships between managers, subordinates, and peers) to enrich information on employees about which little or no information is known. Some researchers explored the Web, such as blogs and social bookmarks, to acquire additional expertise evidence (Amitay et al.; Jiang, Han, & Lu; Serdyukov & Hiemstra, 2008a). Researchers also have been interested in expert finding tasks in the context of the social Web, which relies on users' contribution, since human resources are also valuable assets to a social Web. Some researchers have tried to use social networks for the expert finding task (Chen, Shen, Xiong, Tan, & Cheng, 2006; Zhang et al., 2010). In recent years, some researchers have shown interest on expert finding in the setting of SR sites (Bouguessa et al., 2008; Jurczyk & Agichtein, 2007; Kao, Liu, & Wang; X. Liu, Croft, & Koll, 2005).

The process of expert finding, or expertise location, can be divided into two phases—expertise identification and expertise selection:

1. Expertise identification concerns the problem of knowing what information or special skills other individuals have (who knows what).
2. Expertise selection is about choosing relevant experts from potential experts or people with the requisite expertise (McDonald & Ackerman, 1998).

In order to complete the process, expert finding systems require information about a number of indicators that show the expertise of experts. These information source can be classified into: (1) text-based sources, such as corpora of e-mail, forum threads, publication, documents, web pages, etc., and (2) non-text based sources, such as social network (Chen et al., 2006; Zhang et al., 2010) .

For this research, question and answer pairs associated with user identification will be a good source of information for assessing expertise of experts for routing a question to specific experts. There are, however, some difficulties in obtaining this data from QA services, including SR sites and Ask-an expert sites, primarily due to the lack of any API–like interface (Shah, Oh, & Oh, 2008) or the lack of support due to privacy issues. Thus, there is an immediate need to build a corpora of question and answer pairs associated with user identification for this study.

2.3 DR

The definition of reference service was provided earlier as human-mediated assistance provided to users to help them fulfill their information needs. The evolution of library reference work has

been greatly influenced by the advent of new information and communication technologies. New networking technologies allowed isolated computers to be connected in an enormous network, known as the Internet, enhancing the availability and accessibility of electronic resources. In addition, the arrival of the graphical World Wide Web (WWW, or the Web) greatly facilitated the growth of Internet resources and Internet usage among the general public (Naughton, 2001; Straw, 2001).

As Internet resources expanded, libraries started adopting these technologies to expand the selection of electronic resources and to reinforce the instructional role of reference librarians to meet their users' changing needs. As libraries have made their resources accessible through networks, libraries have also enabled their services to reach and serve users far beyond the library building. Since reference services traditionally have been regarded as one of prime services to users who visit the library, DR services or virtual reference services, have also been offered to assist and guide users online. In this sense, DR services can be seen as an extension of in-person reference services on the Web.

The literature on DR is valuable for this study since it provides some theoretical grounds for this study.

2.3.1 DR Model

The DR model, pictured in Figure 3, is a general-process model developed through an empirical study of high-capacity DR services (Lankes, 1998). This model consists of five steps:

1. Question acquisition includes all issues related to the process of obtaining information from a user. This includes not only the user's question, but also question categorization

and user identification information, via e-mail, web forms, chat, or embedded applications.

2. Triage is the assignment of a question to a reference or subject expert. This step may be automated or conducted via human decision support. Triage also includes the filtering out of repeated or out-of-scope questions.
3. Answer formulation includes all actions taken by the expert to generate a response to a question, including sending the response to a reviewer or directly to the user. Factors for creating "good" answers such as age and cultural appropriateness are included in this step.
4. Tracking is the quantitative and qualitative monitoring of repeat questions for trends. Tracking allows the identification of "hot topics", and may indicate where gaps exist in the collection(s).
5. Resource creation involves the use of tracking data to build or expand collections to better meet users' information needs.

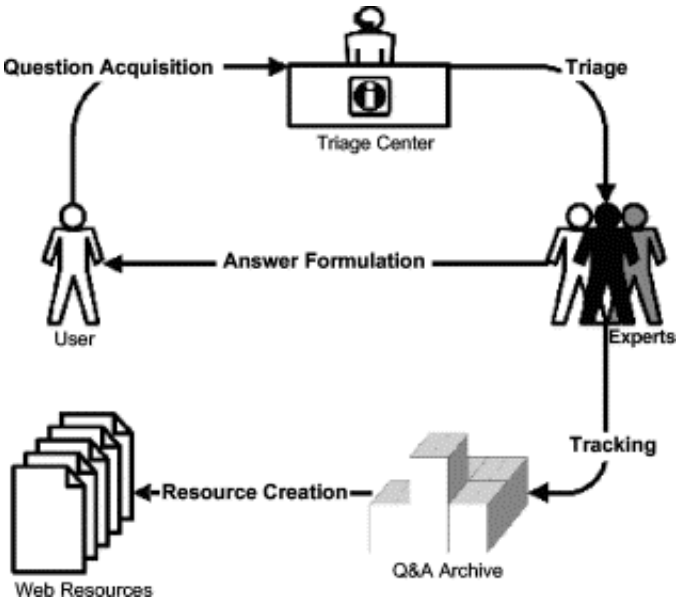

Figure 3. General DR model (Pomerantz, Nicholson, Belanger, & David Lankes, 2004)

The implication of this model for this study is that it illustrates the process of QT in the context of DR. This model is a generic model of DR and has been validated by other researchers in the context of DR (Pomerantz et al., 2004). However, it has limitations when used to illustrate the process of QT in the context of SR; while the user and experts are illustrated distinctively in the model, there is no clear distinction between the user and experts since any user can provide answers to a question in the framework. These limitations make this study complicated and point to the need for a framework for SR. This model can be a starting point to develop a framework for SR for this research.

2.3.2 QT

In desk reference services, the librarian who is in charge of the desk when a patron approaches usually has responsibility to respond to that patron's question. This librarian, however, may not

be the most relevant person to respond to the question regarding subject expertise, reference experience, or other service criteria. In such cases, a reference librarian will refer a patron to another reference librarian or expert, service, or organization. In DR services, a question is, in a similar way, forwarded to other services, often referred to as "triage", when the question cannot be solved by the service in some reason—for example, out of service hours, lack of subject specialty, etc.

While the main goal of both 'referrals' in desk reference and 'triage' in DR is the same—to provide the patron another service to solve the users' information needs, there is a difference in the subject: the patron's question, in DR, is sent from one service to another, while the patron is guided to another service in desk reference services (Pomerantz, 2003). This difference can explain why a lot of work on this topic has been found in the literature on DR rather than desk reference services.

2.3.2.1 Factors that affect the decision of QT

Researchers in the field have also interested in the factors that affect the decision of QT (Pomerantz et al., 2003; Salih, 2007). Pomerantz (2003) found that question subject is the single most important attribute that affects triage decision among fifteen factors determined by Pomerantz, Nichoson, and Lankes:

1. Subject area of the question
2. The service's area(s) of subject expertise
3. The answerer's area of subject expertise

4. Level/depth of assistance available from the service: what is provided as or with an answer: citations only, answers, bibliographic instruction, etc.
5. Number of questions that may be forwarded to the service per unit for times, as set by consortium agreements
6. Response rate of the service: how many of the questions which the service receives get answered
7. The answerer's experience and skill in providing reference service
8. Past performance of the service in providing correct and complete answers
9. The service's turnaround time for answering questions
10. Number of questions that your service may forward to other services per unit of time, as set by consortium agreements
11. Availability of sources to answer the question
12. The answerer's experience and skill with providing customer service
13. Language of the question
14. Scope of the service's collection
15. Question type

These factors provide the researcher with insight on selecting factors that can be used to model the users' expertise for QT in SR. For example, the subject area of question, the answerer's area of subject expertise, the answerer's experience and skill in providing reference service, and past performance of the service in providing correct and complete answers could be used for QT in the SR setting.

2.4 SR

2.4.1 The Social Web

Over the past decade, the Web has constantly evolved. Since the early 2000s, many applications have relied on the users and their relations more than on the information itself. This kind of Web, the second incarnation of the Web or Web 2.0, has been called the 'social Web', since in contrast to Web 1.0, its content can be more easily generated and published by users. The advent of Web 2.0 applications, which provide a framework for interactivity on the Web, support group interaction, and foster a greater sense of community, has emerged to enrich new online social activities, many of which could not been achieved before (Kamel Boulos & Wheeler, 2007). The growth of Web 2.0 and its participatory social sites, such as Flickr, YouTube and Wikipedia, challenges many traditional conceptions of information creation, dissemination, and use (Shachaf & Rosenbaum, 2009). One increasingly popular type of the social Web is the question and answer (Q&A) site such as Yahoo! Answers and Answer.com. In the literature, these services are referred to in different ways by different researchers; for example, "community-based question answering", "question and answer" (Harper, Raban, Rafaeli, & Konstan, 2008), "social question and answer" (Shah, Oh, & Oh, 2009), "online question and answer", and "social reference" (Shachaf, 2010), etc.

2.4.2 What is SR?

Harper, Raban et al. define (2008) that "A question and answer (Q&A) Web sites is purposefully designed to allow people to ask and respond to questions on a broad range of topics." This means

that a Q&A service allows users to express their information needs as questions, in the form of natural language, rather than keywords, and provides answers as relevant snippets or passages, rather than lists of documents. Shachaf prefers to use the term "social reference" to denote social question and answer sites. She defines SR as "online question answering services that are provided by communities of volunteers on question and answer (Q&A) sites," excluding fee-based online question answering services (Shachaf, 2010). Shah, Oh et al. defined "social Q&A" as "a service that involves (1) a method for a person to present his/her information need in natural language (mostly a question, and not a collection of keywords), (2) a place where other people can respond to somebody's information need, and (3) a community build around such service based on participation" (Shah et al., 2009).

These definitions illustrate the concept of community-based question answer sites, the core concept behind those is same: "related or rely on users' participation." Borrowing the concept of "the social Web, or participatory Web," the term "social reference" is used to refer to community-based question answer services.

Shachaf compared SR with library reference services and pointed out the differences between them: in SR the dyadic reference encounter is replaced by teamwork; the traditional boundaries between patrons (questioners) and librarians (answers), are no longer relevant; a bottom up approaches is encouraged to define policies and guidelines; and SR occurs in an online community but traditional reference encounter takes place in a library (Shachaf, 2010).

2.4.3 SR Models

As SR sites are gaining in popularity, an increasing number of SR sites have arisen. Although the main concepts behind SR sites are similar—answers are provided by other users or experts—

different mechanisms for question asking and answering are observed. These mechanisms can be explained with different models, which provide further information on how to implement user participation into a SR service. Different scenarios for SR are observed from existing SR sites.

1) One Question-Many Answers-Many Experts (1Q-nA-nE): This illustrates the process of question and answering in Yahoo Answers[3]. In this scenario, when a user posts a question to the site, then many answers can be provided by many users.
2) One Question-One Answer-Many Experts (1Q 1A nE): This model explains the scenario of the QA process offered by WikiAnswers.[4] In this scenario, when a user posts a question to the site, then only one answer is provided to the question. In this case, many answerers, or experts, collaborate to build and develop the question.
3) One Question-One Answer-One Expert (1Q-1A-nE): This scenario doesn't occur very often, but is important to this researcher. One illustration of this scenario is AllExperts.[5] In this scenario, a user can submit the question to experts who are designated by the questioner.

Exploring these models that address the process of QA is meaningful for this study, which seeks to contribute to the development of user-participatory reference services.

---

[3] http://answers.yahoo.com

[4] http://wiki.answers.com

[5] http://www.allexperts.com

2.4.4 Three types of QA Sites

Based on the literature review, the current online question answer sites can be categorized into three types: “digital reference services”, “expert services”, and “community Q&A sites” (Harper et al., 2008; Shah et al., 2009). Since this section discusses SR, the first two types of online reference are addressed here.

2.4.4.1 DR services

DR services involve “the use of human intermediation to answer questions in a digital environment (Lankes, 2004)”, and represent the online analogue to library reference services (Pomerantz et al., 2004). In DR services, answers are provided by reference librarians who are trained experts. In DRs, one-to-one interactions occur between a reference librarian and a user.

2.4.4.2 Expert services (or Ask-A services)

Expert services are a form of DR services but usually offered by various types of commercial and noncommercial organizations other than libraries. In these services, answers are supposed to be provided by subject experts. Lankes refers to services of this type of reference service as *Ask-A* services.

While social Q&A services are growing in popularity and in number of users, research on social Q&A has not matured. As online Q&A has gained in popularity, many researchers are interested in exploring social Q&A services. Many of the researchers are focusing on information retrieval and information seeking behavior (Gazan, 2007; Shachaf & Rosenbaum,

2009; Shah et al., 2009). An examination of the literature on SR indicates that researchers have not yet fully explored the processes that characterize social Q&A services. Shachaf and Rosenbaum (2009) argue that "while the research is driven by questions and theories in the domain of information retrieval, the interaction among members of the community is largely being neglected."

2.4.5 Approach to research on SR

Since social Web services, following Web 2.0 principles, rely heavily on users' contributions, user participation has emerged as a critical issue for collaborative and social Web systems, as well as for a range of other systems based on the power of a user community.

2.4.5.1 Focus on user information seeking

Many researchers in the field have tried to understand users' information seeking behaviors on SR sites. Adamic and her colleague studied knowledge sharing in Yahoo Answers and found that there is difference in the range of knowledge that users share across the many categories of Yahoo Answers (Adamic, Zhang, Bakshy, & Ackerman, 2008).

Some researchers are interested in investigating user satisfaction in SR. Shah studied effectiveness and user satisfaction in Yahoo Answers, focusing on response time to get answers, and found that "the sooner an answer appears for a question, the higher chances it has being selected as the best answer by the asker" (Shah, 2011). Liu and his colleague focused on user history-asker user history and answerer user history-in order to estimate user satisfaction (Y. Liu, Bian, & Agichtein, 2008).

2.4.5.2 Approach to identify users' motivation

In recent years, SR sites, which allow users to answer questions posted by others, are rapidly growing in popularity. Some researchers in the field are interested in understanding the users' motivation for participating in the collaboration to share their knowledge. Users' motivations can be categorized into two: intrinsic and extrinsic motivation (Farzan & Brusilovsky, 2010). Intrinsic motivation happens when people engage in activities for the activity itself and without any obvious external incentives, such as rewards from others.

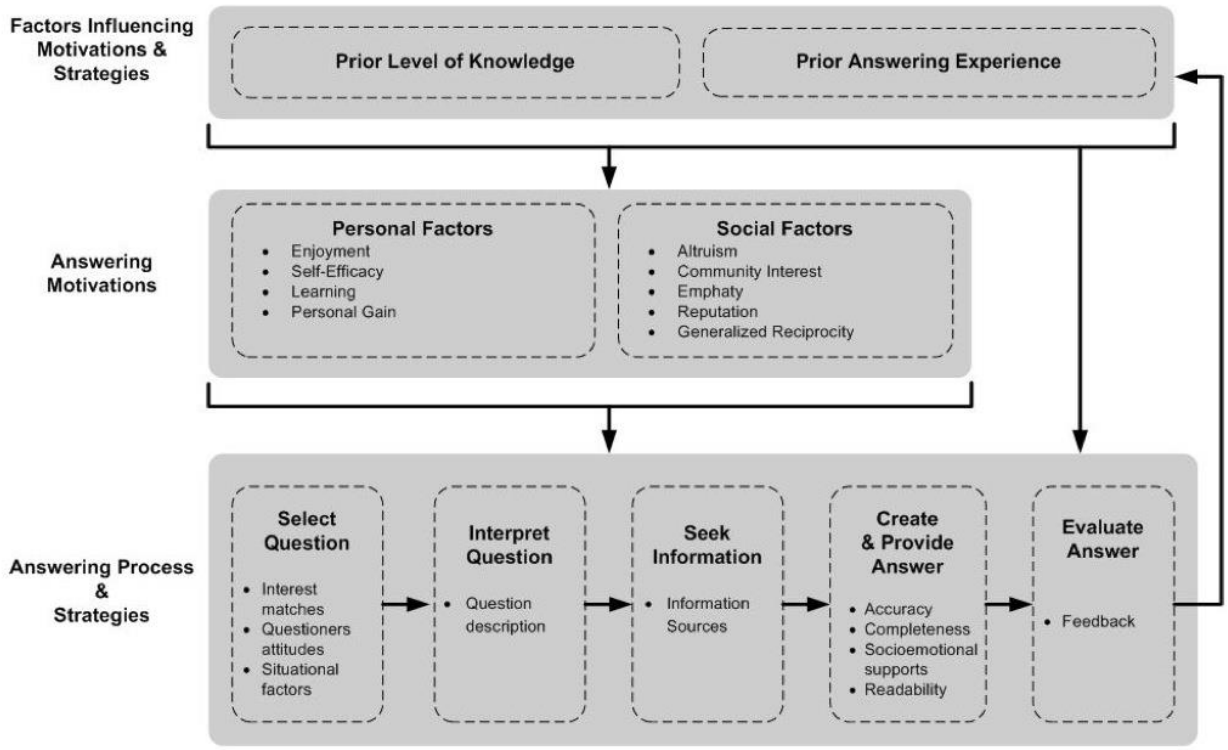

Figure 4. Model of answering behaviors in social Q&A: motivations & strategies (Oh, 2010)

Oh (2010) proposed a model for answerers' motivation in SR (Figure 4). In the model, two fundamental factors which influence the motivations and strategies of answering questions

in social Q&A are assumed: (1) “prior level of knowledge and experience about the topics on which answerers provide information and support,” and (2) “prior answering experience.” This assumption provides the study with implications on assessing expertise of the answerer for expert finding to route the question to the expert. In this study, question-answer pairs of SR are used as an artifact of information for assessing expertise; the experts’ level of knowledge and experience is considered to measure expertise. This model suggests two factors that affect answering motivation: personal and social factors. As an individual, the answerers’ personal needs and interests in social Q&A can motivate them to contribute: for example, self-enjoyment, self-efficacy, learning, and personal gain. In addition, as social beings, answerers want to communicate with questioners and help them by sharing information and support: for example, altruism, empathy, community interest, reputation, and generalized reciprocity (Nam et al., 2009).

2.4.5.3 How to identify the best answerers

The literature provides research on the identification of the best answerers. In approaching this, researchers use different terms to refer to the best answerers (Bouguessa et al., 2008; M. Liu, Liu, & Yang, 2010; Qu et al., 2009), such as experts (Zhang et al., 2010), authoritative users (Bouguessa et al., 2008), etc. Some researchers employed link analysis techniques, such as Page Rank and HITS algorithms to identify authoritative answerers in a SR site (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008; Jurczyk & Agichtein, 2007). Bouguessa and his colleague (2008) proposed a model to identify authoritative actors based on the number of best answers provided by them. Zhang and his colleague (2010) proposed a measure called Z-score which combines the number of answers and questions given by a user to a single value in order to

measure the relative expertise of a user, while other researchers have proposed topic-based models to identify appropriate question-answerers (Inglis, 2003). They introduced latent topic modeling methods for recommending answer providers, and found that combining topic-level information with term-level similarity significantly improves the performance over the term-level only method.

Table 2. Focuses for finding best answerer or experts

Focus	**Researcher**
User link in a community	Bridges, 2003; Inglis, 2003; JH Shera, 1970
Number of best answers by user	Bouguessa et.al., 2003
Combines the number of answers and questions given by a user (Z-score)	Zhang et.al., 2003
Topic	Inglis, 2003
User interest	Qu el.al., 2009

2.5 OVERVIEW OF DIFFERENT APPROACHES TO QA

So far, different approaches to QA or problem solving were discussed above. In this section, an overview of the different approaches to the QA and key elements that are involved in the process is discussed.

2.5.1 Key elements of the QA process

The literature provides overviews of those different approaches (Figure 5). In this framework, key elements that are involved in the process of QA were modeled such as user or questioner, expert or answerer, question, answer, triager, system etc. Among these elements, questioner, answerer, question, and answer were identified as essential elements that are involved in the process of QA process. The questioner is the initiator of the process, when the question is submitted to the service. Thus, the question is a product of the questioner. Answerer is another important actor in the process: the answerer is a person who answers a question. Thus, answer is a product of the answerer. Beyond these, there are subordinate elements, such as triager, system, or information source. In reference services, the submitted question may be transferred to another answerer or service if the question cannot be answered by the service. Triager is a person or system in charge of making decisions about how to route the question to an answerer or service. In some automated QA approaches, the system is involved in the QA process by generating and providing answers to the question automatically (QA system), or directly recommends experts who are able to provide answers to the question (expert finding system). The information source is also involved in the process of QA since reference services, through the triager, sometimes guide the user to an information source that contains the answer in itself,

while, QA systems and expert finding systems also use information sources to generate answers to the question (QA system) or to find evidence for recommending experts to the question (expert finding system).

In the context of SR, the core elements are also important elements in the QA process. However there is a difference in the distinction between questioner and answerer. Although obvious in reference services—answers are provided only by the librarian— the separation between questioners and answers in SR is equivocal since a user can be not only a questioner but also an answerer. The other subordinate elements, such as triager or information source, are not available in the context of SR. The information resource is concealed in the context of SR: the answerer has access to an online Web resource instead of being about to hold an information resource. The element of triage is not available currently since it is not provided by the current online SR services.

Table 3. Elements of QA process

Category	Actor	Action	Product	Note
Core	**Questioner* (user)**	Ask	**Question***	
	Answerer* (expert)	Answer	**Answer***	
Subordinate	Triager	Refer questioner or triage question		Desk reference
	System	Generate answer	Answer	QA system
		Recommend experts		Expert finding system

Note. * indicates essential elements of QA process

2.6 CHAPTER SUMMARY

The literature supports QA as a constant role of reference librarian throughout the evolvement of reference services. This justifies the use of term “reference” with “social” to refer to “social QA”. There have been approaches to automate the process of QA using computing power in the domain, but although there are different approaches, these can be conceptualized in a framework. SR can be characterized by its use of user expertise or contribution to provide answers. The literature also supports the fact that the growing SR services find that their quality and efficiency are increasingly challenged by the significantly increasing number of questions submitted to the services.

The field of digital libraries has developed a theoretical background for QT, which needs to be tested in the context of SR. The researcher identified fifteen important factors affecting decision making in the process of QT found in digital reference settings from a literature review. These factors need to be redefined in order to be used in the context of SR. There is also a need for further study of automated QT. The literature also provides insight into the users’ motivations to participate in answering questions within online communities.

All in all, the literature supports the needs for research into automated QT in the context of SR.

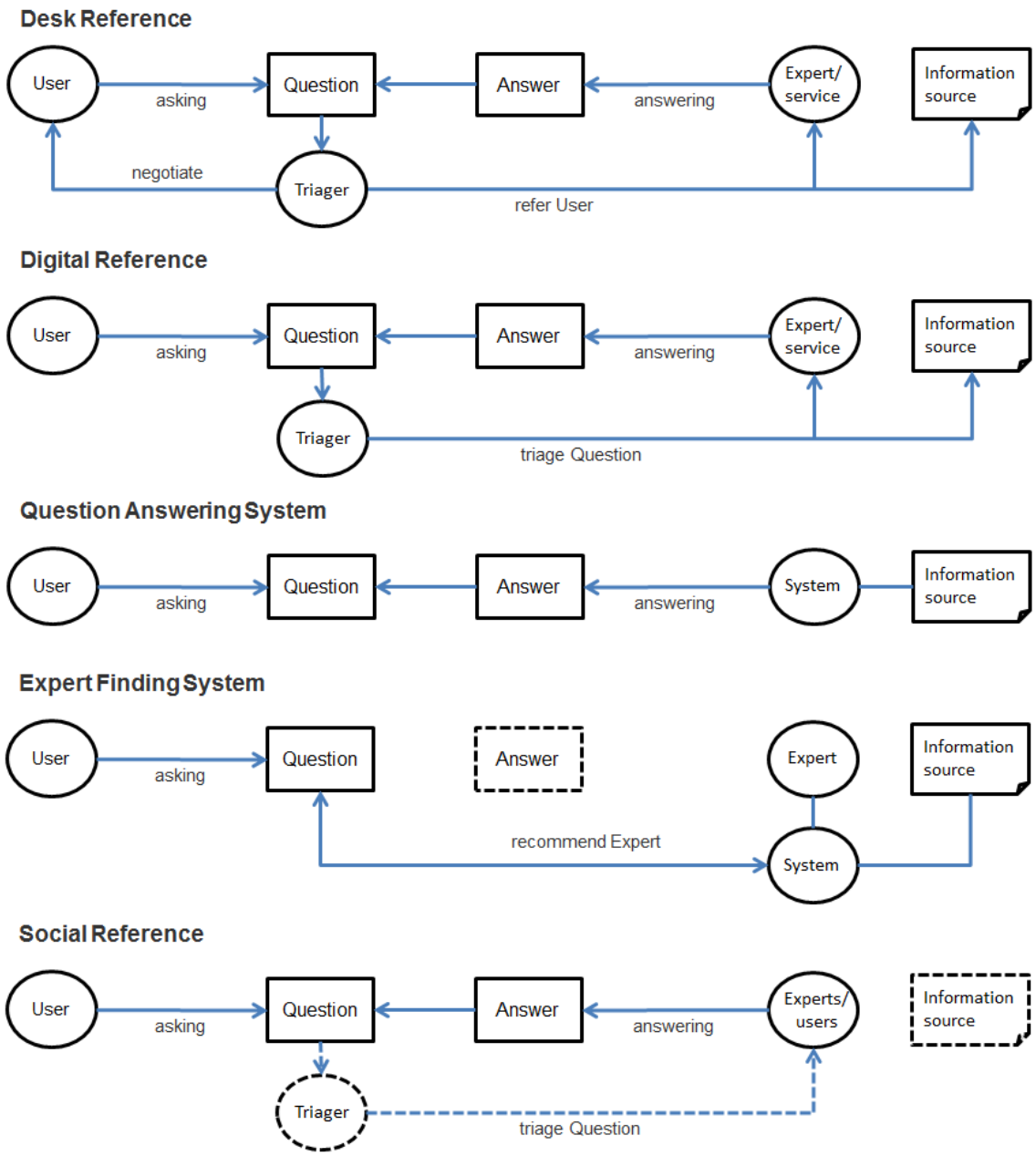

Figure 5. Conceptual frameworks of different approaches to QA

3 RESEARCH DESIGN AND METHODOLOGY

3.1 INTRODUCTION

The review of literature has produced reoccurring themes emphasizing the importance of automated QA for both digital and SR services. QT is an important step in the process of QA for these services. The literature supports that there are available technologies for automated QA. This study investigated automated QT in the SR setting.

3.1.1 Purpose of the study

The goal of the research was to describe and gain a further understanding of automated QT in the context of SR. QT is understood as the process of assigning submitted questions to relevant users who have ability to answer the question assigned to them. Under the goal of investigating SR services, the following three specific objectives were set:

1. to investigate key elements of automated QT,
2. to identify factors and requirements for automated QT, and
3. to design new conceptual models of automated QT.

3.1.2 Research questions

In order to achieve above the objectives, this study was guided by the following two majorin questions:

1. What are the elements of automated QT for SR, and are they different from that of DR?
2. What are the factors that affect the performance of automated QT for SR?

Since the literature supports that researchers in the field of digital libraries have developed the theoretical and practical foundations of QT for DR in the context of digital libraries, the first question used this model as a base to understand QT in the context of SR. The researcher tried to compare the QA processes of social and DR.

The second question sought to investigate factors that affect QT in the context of SR, in order to build an automated QT model for this area.

3.1.3 Overview of the research design

In order to answer the research questions, a sequential two-phase design approach was employed by this study, conducting first a qualitative and then a quantitative phase. The steps in the study methodology are represented in Figure 6.

In the first phase, the literature review was employed for qualitative research. Since attributes and elements that were used for developing the automated QT model for SR were mainly selected from previous studies in the field of library science, literature review was useful

for this analysis. In this phase, a qualitative content analysis was conducted in order to identify the elements of QT, not only in the context of DR but also in the context of SR.

In the second, quantitative phase, an experiment was conducted and evaluated, in order to answer the second research question. The experiment was designed and conducted in order to investigate factors that would affect the performance of an automated QT model for SR, and then an evaluation was conducted in order to assess the model proposed.

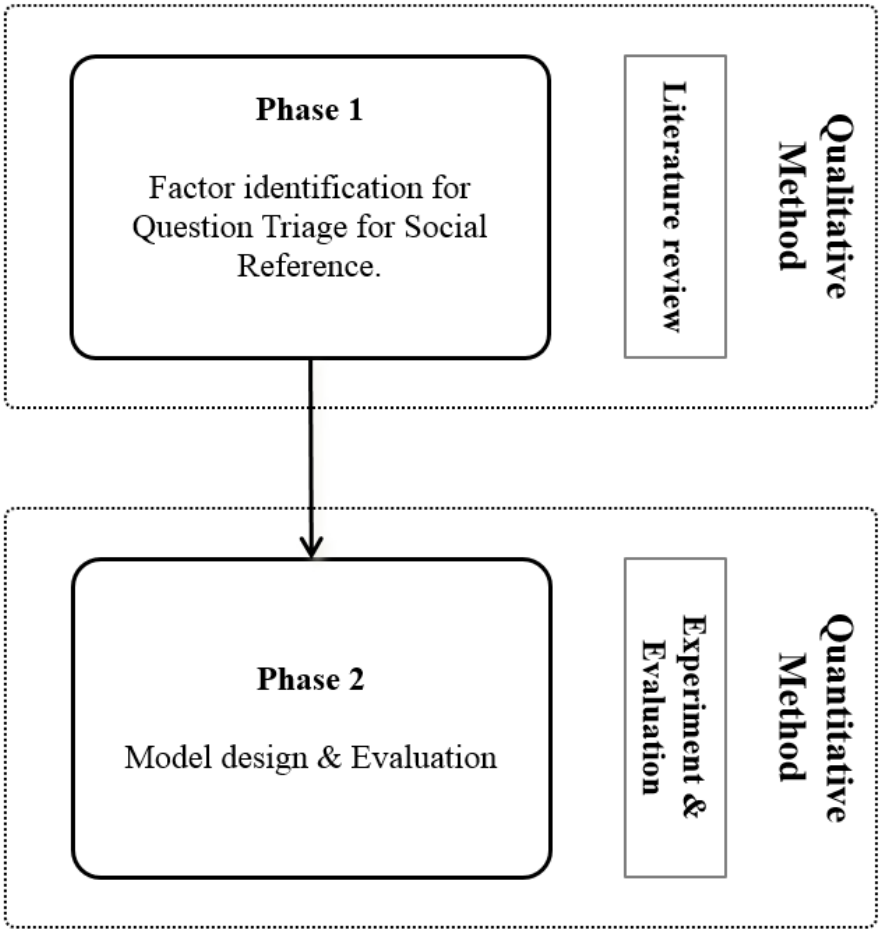

Figure 6. An overview of the research methodology

3.1.4 Methodological challenges

3.1.4.1 Limitations in eliciting all factors affecting QT for SR

The literature suggests many factors that are considered in the decision making of QT in DR. However, it was hard to elicit all those factors in the context of SR since some of those factors, such as factors attributed from the service, were not applicable in the SR setting. In this study, the most important eight factors, omitting factors related to the service, were selected and elicited in the context of SR.

3.1.4.2 Difficulty in conducting experiments with real SR users

Although QT is a common concept in the field of DR, it is a totally new concept in the field of SR. For this reason, it was hard to investigate factors that affect the decision for the QT from real users of a SR site; the concept of QT is new to these users since no human or automated triager has been introduced for SR. The best way to evaluate the results of the experiment is to assess them with real SR users. Instead of asking real users of a SR site some survey questions, a collection of question-answer pairs generated as the result of user activities was used to investigate important factors affecting the estimation of relevant answerer candidates to given questions.

3.2 PHASE 1: FACTOR IDENTIFICATION FOR QT FOR SR

This phase of the study sought to answer Research Question 1: what are the elements of automated QT for SR, and how are they different from that of DR? In order to answer this question, the researcher first investigated the elements of DR QT from the literature, and then examined whether those elements could be applied to QT in the context of SR. For this reason, the question was broken down into two sub-questions:

RQ1a. What are the essential elements of QT in the context of DR?

RQ1b. Can those elements be applied to QT in the context of SR?

3.2.1 Identification of the key elements of the QT process

In order to identify key elements of QT, different approaches to QA were explored in chapter 2, and elements involved in the process of QA were identified (Figure 4). Questioner, answerer, question, answer and information source were categorized as core elements in the QA process.

Among the four different approaches to QA, including both human-mediated and automated approaches, the QA approach to DR is preferred in order to develop an automated QT model for SR since the QA process in DR will be almost identical to the QA process in SR, once SR implements the QT in the process.

The QA process of the proposed SR model is illustrated in Figure 7. This study explored QT models studied in the field of DR trying to introduce parallel methods to benefit QT in SR. In the current study, the QT model of DR was used as a tentative model for SR, and online SR sites were examined to verify or modify the models to complete a framework for QT for SR, incorporating these key elements of QT.

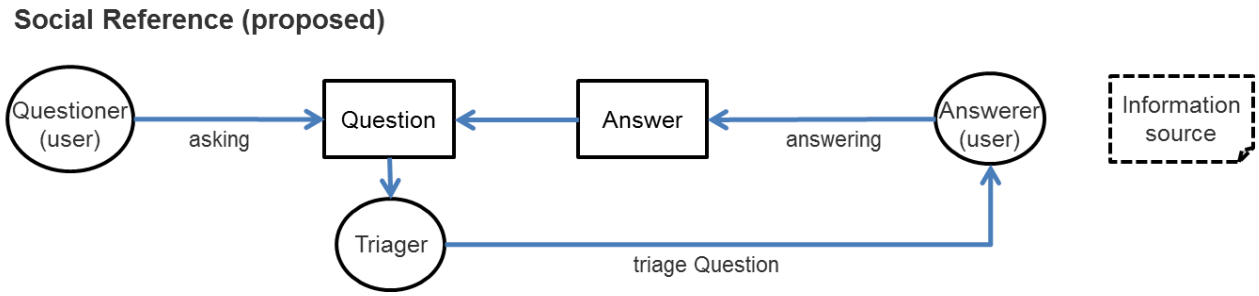

Figure 7. Proposed framework of QA process for SR

Figure 7 illustrates the process of QA in the context of SR, implementing the operation of QT in the process. Since it is not realistic to staff a human triager for the operation of QT for SR, this study focused on automated QT. Thus, questioner, answerer, question and answer are the core elements in this process.

3.2.2 Selection of attributes influencing QT

In the first phase, the researcher sought to investigate attributes of the core elements of QT which could be considered in QT decision making process in the SR. The goal of this phase was to seek to build a foundation for the next phase: developing new models for assessing user expertise.

In order to do that, it is first necessary to identify essential elements to which those attributes belong. Fortunately, the literature provides a comprehensive investigation of the identifying attributes of key elements as factors that affect the decisions of QT. Pomerantz and his colleagues (Pomerantz et al., 2003) determined the fifteen most important factors that affect a triager's decisions in the DR setting, based on a survey of experts about the decision-making process in DR. In the current study, those fifteen factors determined by Pomerantz and his colleagues were adopted and redefined in the context of SR.

Since DR services are usually provided by libraries or institutions, not only the attributes of question and answerer but also the attributes of service were considered.

The question's subject area, language and question type are attributes of the question, with the subject area of the question being the most important attribute being considered for decision making within QT (Pomerantz, 2003, p. 204). In terms of the answerer, only two attributes, the area of subject expertise, experience/skill, were considered for the triage decisions.

Table 4. Fifteen factors affecting QT for DR (Pomerantz, 2003)

Attribute (factor) \ **Element**		**Question**	**Answerer**	**Service**
Subject area		1		
Area of subject expertise			3	2
Level/depth of assistance				4
Number of question per unit for times forwarded	to the service			5
	to the other service			10
Response rate				6
Experience and skill in providing	customer service		7	
	reference service		12	
Past performance in providing correct answers				8
Turnaround time				9
Availability of source				11
Language of the question		13		
Scope of the collection				14
Question type		15		

Note. The number indicates the attributes' importance ranking for that element.

Service was recognized as the most popular element; the other eight factors belong to the service, such as the area of subject expertise, level/depth of assistance, number of question per unit for times, response rate, past performance, turnaround time, availability of source, and the scope of the collection. This seemed to be related to the nature of DR. Since DR services were usually provided and developed by libraries or library affiliated institutions, triagers or reference librarians seemed to pay more attention to the attributes of service than that of the answerer, when making triage decisions.

Table 4 lists the key elements and their attributes that are considered as important factors affecting decision making in QT in the context of DR. In order to use these attributes as factors for decision making for SR, we first identified commonalities between DR and SR. In terms of elements, both question and answerer are common elements, but not service. For this reason, it is possible to directly adopt the attributes of question and answerer elements as factors of QT for SR.

3.2.2.1 Mapping the attributes of the question in SR

Question is one of the core elements that appears in the process of QA in both DR and SR. Thus, the attributes of question are still valid to be considered as decision-making factors of QT in the context of SR. However, the researcher selected only the subject area of question as a factor for investigation in the current study since the researcher is interested in identifying important factors that can be attributed to the answerer, not the question.

Table 5. Mapping the attribute of the question from DR into SR

Rank of importance	Question (DR)	Question (SR)	Note.
1	Subject area of question	Subject area of question	
13	Language of question	Language of question	Applicable, but not selected for this study
15	Type of question	Type of question	Applicable, but not selected for this study

3.2.2.2 Mapping the attributes of the answerer in SR

Like question, answerer is also one of core elements that appear in the QA process in both DR and SR. Thus, the attributes of answerer are still valid to be considered as decision-making factors of QT in the context of SR. However, in terms of the answerer element, the 7th factor "experience and skill in providing customer service" could not directly be applicable to SR since the user in the SR setting does not provide customer service. The attribute "area of subject expertise" can be directly applicable to SR. "Experience and skill in providing reference service" is mapped to "questioning activity" and "answering activity" in the context of SR. Table 6 shows the applicability of those attributes in the context of SR in the current study.

Table 6. Mapping the attribute of the answerer from DR into SR

Rank of importance	**Answerer (DR)**	**Answerer (SR)**	**Note**
3	Area of subject expertise	Area of subject interest	
		Area of topic interest	
7	Experience and skill in providing customer service	N/A	Users provide no customer service in SR.
12	Experience and skill in providing reference service	Questioning activity	
		Answering activity	

3.2.2.3 Mapping the attributes of the service in SR

On the other hand, in the SR setting, the answerer is the single contributor that provides answers to the questioner. For this reason, the attributes of the service cannot be directly used in the SR setting. However, the researcher determined many of those factors attributed from service could still be adopted as the attributes of the answerer by redefining them in the context of SR (Table 7).

The scope of the collection could not directly be adopted as an attribute of the answerer; the answerer does not hold collections in SR settings. However, the service collections could be reinterpreted as representing the service's subject expertise. For this reason, the scope of the answerer's subject expertise was added instead of using the scope of collection directly as an attribute of the answerer. Availability of source was not able to be adopted as an attribute of the

answerer since the answerer does not hold an information source but has access to online information sources, in the context of SR.

As the result, a total of nine attributes were selected that affect the decisions of QT for SR service. One of them belongs to the question and the others belong to the answerer. Table 8 lists these attributes and their applicability to SR service. Further discussion on these attributes is provided in the following sections. In this study, the term ‘subject interest’ was used to refer to the concept of ‘subject expertise’ used in DR, in order to avoid confusion in distinguishing subject interest from the generic term ‘expertise’ used for defining user expertise.

Table 7. Mapping the service attributes onto the answerer

Rank of importance	**Service (DR)**	**Answerer (SR)**	**Note.**
2	Area of subject expertise	N/A.	Redundant attribute. Merge into area of subject expertise of the answerer
4	Level/depth of assistance	Level of subject interest of the answerer	
5	# of questions to the service	Quota	Calculate how many Qs are answered by the answerer.
10	# of questions to other service	N/A.	In SR setting, a user cannot forward a question to another user.
6	Response rate	N/A.	Since there was no question assignment in Ask Metafilter
8	Past performance	Performance	

Table 7 Continue

Rank of importance	Service (DR)	Answerer (SR)	Note.
9	Turnaround time	Response time	It's possible to calculate the response time of the answerer to the question
14	Scope of the collection	Scope of subject interest of the answerer	In the SR setting, the answerer's previous QA pairs can be considered as a collection of the answerer.
11	Availability of source	N/A.	In the SR setting, users do not provide the available source information.

3.2.2.4 Re-definition of the chosen attributes in the SR setting

In this section, the chosen attributes of question, answerer, and service are redefined in the context of SR, and the method to obtain that information from the main collection is addressed.

3.2.2.4.1 Subject area of the question

The question's subject was considered as the single most important factor affecting decision-making in QT in DR (Pomerantz, 2004). In order to make the decision to route a question to relevant answerers, we need to identify the primary matter of the discussion (subject) of the question. For this study, we used two concepts, subject and topic, to represent the primary matter

of the discussion of the question. They were selected as key attributes of the question, in order to investigate factors affecting triage decisions.

Table 8. Mapping the triage factors of DR to SR

Element	**Attribute (in DR)**	**Attribute (in SR)**	**Applicability**
Question	Subject area	Subject area	Category posted
		Topic area	Tag assigned by the questioner
Answerer	Area of subject expertise	Subject area of interest	Category in which answers were provided previously
		Topic area of interest	Tag associated with questions to which the user answered
	Scope of the subject expertise	Scope of subject interest	Generality (subject)
		Scope of topic interest	Specialty (topic)
	Level of subject expertise	Level of subject/topic interest	Calculate score of subject level using some algorithms
	Past performance		Providing good answers
	Experience & skill	Q activity	General experience in posting questions
		A activity	General experience in answering questions
	Turnaround time	Response time	Count the response time
	Number of question per unit for times	Quota	The number of questions to which the answerer provided answers in a day

Note.

a. Scope of subject expertise was used instead of using the service's scope of the collection.

b. Level of subject expertise was added for the purposes of the study.

3.2.2.4.2 Answerer's subject area of interest

The area of subject expertise was originally identified as an attribute of the answerer in DR settings. In this study, the researcher employed two concepts - subject and topic - for addressing the primary matter of the discussion or thought of the answerer. One of the main reasons for employing two concepts for representing the answerers' subject area of interest is that the researcher tried to investigate the difference between subject area of interest and topic area of interest in their effect on correctly choosing relevant answerer candidates to given questions. It was assumed that the answerer's topic of interest, as a specific interest, would be a more important factor than the answerer's subject area of interest, a general interest. In the current study, the answerer's subject area and topic area are regarded as attributes of the answerer. Table 9 shows the conceptual difference between *subject* and *topic*, in terms of the boundary of the answerer's interest.

In order to assess the answerer's subject area of interest using the main collection of this study, the researcher examined the answerer's previous category preference in providing answers, and interpreted the answerer's interest on categories as evidence of the answerer's subject interest.

3.2.2.4.3 Answerer's topic area of interest

The topic area of interest of the answerer was selected as a factor to model the answerer's specific/local interest in a subject area. Unlike subject area of interest, a general interest, the information on the topic area is not contained explicitly in the main collection of this study. For this reason, the researcher tried to capture topic information from user-assigned tags and latent

topics generated from Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) so as to represent the answerer's specific/local interest.

Table 9. Comparison of subject with topic

	Subject area of interest	**Topic area of interest**
Characteristics	Focus on macro interest General interest Global interest	Focus on micro interest Specific interest Local interest
Example	Category	Sub-category Tags in a category Latent topics in a category

3.2.2.4.4 Scope of the subject area of interest of the answerer

The scope of the subject area of interest is defined here as the extent of the area or subject matter that the user showed interest in. There may be differences in the scope of subject area of interest among users. For instance, while some users are only interested in a few subjects, others may have interests in several subject areas at the same time, or the users' subject areas of interest may change over time. Thus, understanding differences in the scope of subject areas of interest among users would help to guide decision-making in the QT process.

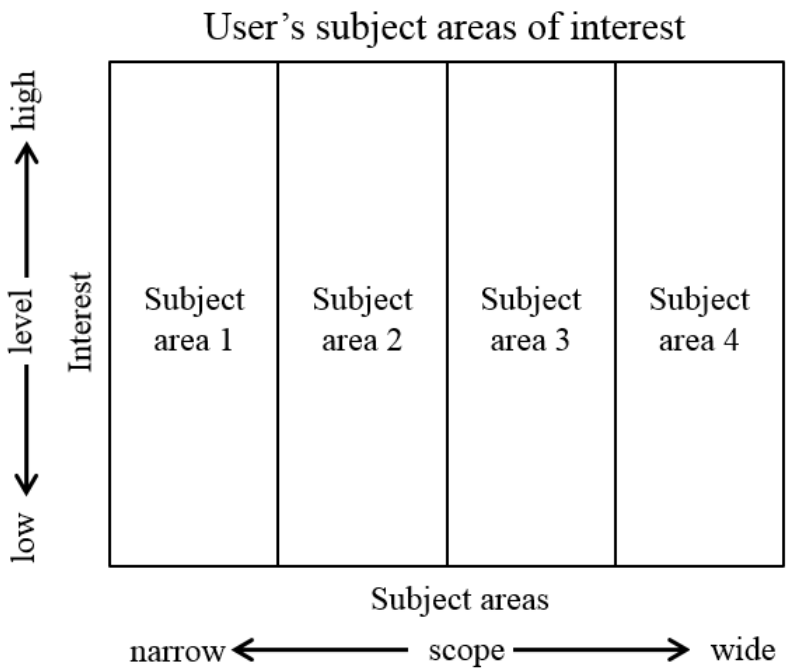

Figure 8. Map of user's subject areas of interest

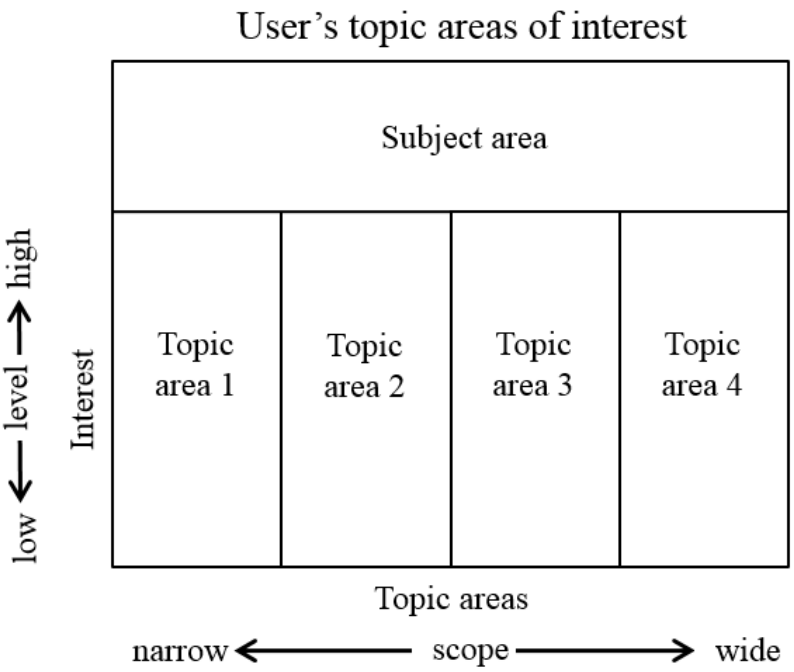

Figure 9. Map of user's topic areas of interest

3.2.2.4.5 Level of subject interest of the answerer

The researcher included this factor for the purpose of the study, to model the answerer's degrees of subject interest. The level of subject interest is defined here as a degree on a hypothetical continuum showing the extent of subject interest. From a preliminary manual examination of question-answer pairs in the main dataset, we observed that there is a difference in the level of subject interest on a specific subject among users. While one may have a relatively high level of interest in a certain subject, another may have a relatively low level of subject interest in the same subject field. Accordingly, the researcher selected this factor to investigate its effect on triage decision making to estimate relevant answerer candidates to given questions.

3.2.2.4.6 Performance

Past performance is defined here as the level of providing good answers which are selected by other users. The answerer's performance in providing good answers can be measured by calculating the number of answers ranked as "favorite" by users.

3.2.2.4.7 Q&A activities

Among users, there may be a difference in number of contributions made to the community. Some users may prefer to ask questions but others may show a higher frequency of providing answers than asking questions. Considering this kind of difference will be useful in making decisions in QT.

3.2.2.4.8 Response time

Instead of using the term ‘turnaround time’, we use ‘response time’ in the context of SR. In the current SR setting, we can infer the answerer’s response time to questions which were posted to the community by looking at the answerer’s previous answering activities. In order to forward a question to appropriate answerers, their response time could be considered as an important factor in triage decision, especially when the question needs urgent care.

3.2.2.4.9 Quota

The number of answers that the user can provide in a day is very limited. Considering the affordable number of answers that the answerer can give is important in the decision of QT; if the number of answers that the answerer provided in a day has reached their limit, the answerer may provide no more answers that day.

3.2.2.5 Summary

For this study, the following attributes was chosen as essential for QT in SR settings:

Question: the question’s subject area and topic area were selected as its attributes, but language of the question and question type were not chosen for the current study.

Answerer: the scope of topic area of interest, the level of subject and topic interest, activities in providing answers and asking questions, response time to answering questions and quota of questions answered by the answerer were chosen.

3.2.3 A framework of expertise: subject interest, performance, and contribution

In the SR setting, users as answerers are the single element that provides answers to the question in the process of QA. In order to sharpen their selection, a framework was built to model user expertise, for this study.

Finding relevant experts or answerers to the question is the key to success in QT. In any expert finding application, a fundamental question is often "What is an expert?" Indeed, agreement on "who or what an expert is" is a highly subjective matter, which may even become controversial. Though expertise is often referred to as an expert's knowledge, skill, and experience (Ericsson, 2007), those concepts need to be defined clearly for further discussion. At first glance, those terms seem synonymous but soon you would realize they are very different concepts. The following three questions are useful to make them clearer:

1) Who knows what? – Knowledge
2) Who can do that? – Skill
3) What has that person done? – Experience

In this research, a framework consisting of subject interest, performance, and contribution is used instead of directly employing the above three generic concepts of expertise to define expertise in the context of SR. To make the framework of expertise used in this research convincing, the following three questions need to be raised:

1) Who is interested in that subject/topic? – Subject interest
2) Who is able to provide 'relevant' answers, as assessed by users, to the question submitted? – Performance
3) Who has contributed to asking questions and providing answers to the community? – Contribution.

While these questions are clear and it's easy to understand what the question seeks, further discussion of its use within the SR setting is needed to refine the framework of expertise for this study.

3.2.3.1 Subject interest (knowledge)

The question "Who knows what?" asks what "knowledge" the person can give as the answer. The Webster's Dictionary defines "knowledge" as "the fact or condition of knowing something with familiarity gained through experience" ("Knowledge," n.d.). By relying on this definition, the concept of knowledge can be differentiated from experience in that experience is a basis of knowledge. Expanding this definition, "knowledge" is referred to as "awareness or familiarity gained through experience of QA by a user." In this research, the above question is rephrased as "Who is interested in the subject and topic of the given question?", and the term *knowledge* is reworded as *subject interest* for this study. In this research we want to locate users who are interested in the question submitted in order to route the question to the users. Thus, the first dimension for judging user expertise is the user interest about the subject and topic of the given question.

3.2.3.2 Performance (skill)

The question "Who can do that?" asks the "skill" of a person to answer. The Webster's Dictionary defines "skill" as "the ability to use one's knowledge effectively and readily in execution or performance" ("Skill," n.d.). This definition contains two key concepts of skill: (1) ability of using knowledge and applying it in a context; (2) efficiency in "execution or

performance". This definition is acceptable for this study since it distinguishes skill from knowledge. While knowledge refers to the theoretical or practical understanding of a subject, skill refers to successfully applying that theory in practice and getting expected results. For instance, in SR services, a user who has preference for a subject category may know something with familiarity gained through the experience of question posting and answer providing in the subject category. Going forward, the user would know more about the subjects or topics that are discussed in the category. All the above can be regarded as the user's knowledge about the subject or topic. Transferring this knowledge to provide a successful answer, satisfying the questioner's need, to the question posted in the subject category is understood as the user's skill, referred to as performance. In this research, the question asking for skill can be rephrased as "Who is able to provide 'relevant' answers to the topic of the question?" in the context of SR service. Thus, the second dimension for judging user expertise is the user's ability to compose good answers, applying knowledge about the subject or topic of the given question.

3.2.3.3 Contribution (experience)

The question "What has the person done?" seeks to measure the "experience" of a person. The Webster's Dictionary defines "experience" as "direct observation of or participation in events as a basis of knowledge" ("Experience," n.d.). Adopting this definition, experience, in this study, refers to the user's participation in question asking and QA as a basis of subject and topic interest. For this research the third question is rephrased as "Who is able to participate in providing answers?" and the term *experience* is reworded as *contribution* for this study.

3.2.3.4 Mapping triage factors into the framework

Employing the above framework of expertise, the factors affecting to the decision of question triager, discussed in the previous section, can be mapped into three dimensions as the aspects of expertise. Table 10 illustrates the dimensions of the framework and related factors affecting QT decisions.

Table 10. The framework of expertise

Dimension	Question	Aspect
Subject interest	Who is interested in the subject and topic of the question submitted?	- Area of the subject and topic interest - Scope of the subject and topic area - Level of subject and topic
Performance	Who is able to provide relevant answers to the question submitted?	- Performance (providing best answers)
Contribution	Who has contributed to asking questions and providing answers to the community?	- Q&A activities (roles) - Turnaround time/availability - Quota

In order to route a question to an expert in the SR setting, we need first to know who know something about the subject or topic of the question. The first question "Who knows about the subject or topic of the question submitted?", seeking knowledge, deals with subject or topic as discussed earlier, and this subject matter has been regarded as one the most important factors in the decision of QT by other researchers (Pomerantz et al., 2004). In fact matching subject

matter between the question and the answer is the key to success in QT, in order to meet the user's need. For this reason, the subject interest of an expert is one of the core aspects of expertise. In addition, since the user's subject interest may vary between people, we also need to consider this variance of subject interest, i.e., the scope of subject and the level of subject interest, as the attribute of knowledge expertise.

3.2.3.4.1 Subject of interest

In order to route a question to an expert, we want to know who has knowledge about the subject or topic of the question. In this research, the user's interest, in terms of answering on a specific subject or topic, is interpreted as one's knowledge about the subject or topic, assuming that users who have provided answers previously on a subject or topic are familiar with the subject or topic so that they possess understanding on the subject or topic. For instance, a user who is ignorant of the subject or topic of the submitted question is not able to provide answers to the question.

3.2.3.4.2 Scope of the subject or topic areas

The scope of the subject of interest is defined here as the extent of the area or subject matter that the user deals with. There are differences among users in terms of the scope of the users' subject interest. For example, while some users are only interested in a few subjects, others may be interested in many subject areas at the same time, or the users' subject interest may change over time.

3.2.3.4.3 Level of subject or topic interest

This represents the level of specialization in the given subject area or category. Since a subject area or category is composed of one or more sub topics, it is assumed that there may be difference among users in the interest in those sub-topics. Some users may be interested in many sub-topics in a subject area, while others may be interested in just a few sub-topics in the subject area. In this study, users who are familiar with most of the topics in a subject area are interpreted as experts who have comprehensive understanding to the subject area, so that they are able to provide better answer to questions which require comprehensive understanding on the subject area.

3.2.3.4.4 Performance in providing good answers

For the second dimension, the performance aspect is considered. The performance factor, i.e. providing the best answers, can be regarded as the answer to the question of "Who is able to provide 'best' answers to the question submitted?", seeking skill as the basis of expertise. In other words, a user who provides the best answers is interpreted as an expert who has knowledge of the topic of the question and is good at utilizing that knowledge to provide answers to the question submitted.

3.2.3.4.5 Activities in providing questions and answers, turnaround time and quota

For the third dimension, contribution aspect is considered. The question "Who is able to participate in providing answers to the question submitted?" is related to contribution. The other

factors, such as activities, turnaround time/availability, and quota, can be evidence of the answer to this question; they are mapped into the contribution dimension as a basis of expertise.

3.2.4 Research steps in Phase 1

Detailed research steps in the first research design phase are listed in Table 11.

Table 11. Research steps in phase 1

Step	Step in Method	Description	Outcome
1	Literature review	Literature review on QT for DR	Understanding of research work on QT
2	Identification of models	Identify existing QT models, both conceptual and practical	List of models of QT in DR
3	Identification of key elements of QT and its attributes for DR	Core elements of QT and its attributes are identified from the literature review	A framework for QT in DR setting
4	SR service selection	A Web directory service is used to select SR service available on the Web	List of exemplary SR services
5	Analytic induction	Testing existing models of QT from DR in SR setting	Commonality and differences are identified
6	Identification of key elements and their relations in the process of QA in SR	Core elements of QA and their relations in SR services are identified	A framework of QA in SR

Table 11 continued

Step	Step in Method	Description	Outcome
7	Identification of new factors and requirements of QT for SR	Based on the framework of QA processes in SR, new factors and requirements are selected by doing analysis on the framework and comparison with that of DR	Factors and requirements for QT for SR
8	Developing a new framework	A new framework of QT for SR is developed	A revised framework of QT for SR
9	Identification of possible factors affecting QT	Possible factors that affect QT for SR is identified.	List of possible factors to QT
10	Limitation of factors	For the experiment, core factors are selected: subject expertise, skill and performance.	Factors considered: subject interest, performance, and contribution

3.3 PHASE 2: EXPERIMENT AND EVALUATION

The goal of this phase was to answer RQ2: What are the factors that affect the performance of automated QT for SR? In order to answer this question, this phase employed experimental methods in order to design and test new models for automated QT for SR (the research objective 3). As the result of Phase 1, factors that would affect the automated QT for SR were determined.

In this phase, the impact of each factor (IV) to estimate a user to be a relevant answerer to a given question (DV) was evaluated using logistic regression analysis for the purpose of developing an automated QT model for SR. Methodological steps in this phase are listed in Table 13.

3.3.1 Logistic regression analysis

In this research, a logistic regression analysis was completed for the experiment. Linear regression analysis is a statistical analysis technique that assesses the impact of a predictor variable (the independent variable) on a criterion variable (a dependent variable). In terms of linear regression, there are three primary assumptions: outlier, linearity, and constant variance. Linear regression is very sensitive to outliers, tied to linear model, and useful to estimate constant variance, i.e., numbers. The dataset we have may contain many outliers, which may harm the prediction, and since we are trying to test several independent variables, we avoided the linear regression method. Also, since our goal for the evaluation is to estimate whether or not a candidate is one of the best possible answerers, logistic regression analysis was selected for this study. Logistic regression, which determines the impact of multiple independent variables presented simultaneously to predict membership of one or other of the two dependent variable categories. While logistic regression gives each predictor (IV) a coefficient 'β' which measures its independent contribution to variations in the dependent variable, the dependent variable can only take on one of the two values: 0 or 1 (Burns & Burns, 2008). What we want to predict from knowledge of relevant independent variables and coefficients is the probability (P) that it is 1 rather than 0 (belonging to one group rather than the other). For this study, the probability can be interpreted as the probability of a candidate to be a relevant answerer to the question assigned. In this study logistic regression is used to determine if the factors selected can be used to predict whether or not a candidate is a relevant answerer to the question assigned.

3.3.1.1 Variables

In order to model a logistic regression model for the experiment, selected factors for automated QT for SR were formulated as independent variables (IVs), and categorical information about a user's relevance information to a given question was defined as a dependent variable (DV) (Table 12).

3.3.1.2 Logistic regression model

For the experiment, the following expertise model was built:

$$\text{Expertise_Score} = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5 + \beta_6 \cdot x_6 + \beta_7 \cdot x_7 + \beta_8 \cdot x_8$$

where $x_1 \ldots x_8$ are the values of each factor, $\beta_1 \ldots \beta_8$ are the regression coefficients of factors, and β_0 is the intercept from the linear regression equation. $\beta_0 \ldots \beta_8$ were calculated by training on our chosen dataset.

Since the experimental design conducted in this phase needs more space for discussion, it is addressed in chapter 4.

Table 12. Variables in the logistic regression analysis

Variables	Factor	Notation
DV	N/A	Expertise_Score
IV	Topic of interest	x_1
	Scope of subject area	x_2
	Scope of topic area	x_3
	Answer providing activity	x_4
	Question providing activity	x_5
	Providing good answers	x_6
	Response time	x_7
	Quota	x_8

3.3.2 Research steps in Phase 2

Detailed research steps completed in the second phase of the research design are listed in Table 13.

Table 13. Research step in phase 2: building new models for QT in SR

	Step in Method	**Description**	**Outcome**
11	Experimental design	An experimental design is developed for automated QT for a SR service. The main task of the experiment is to find answerers to a given questions based on the answerers' expertise (subject expertise and experience).Develop a tentative expertise model for automated QT	A framework for the experiment Tentative expertise model for automated QT
12	Experiment	Do logistic regression analysis with the training dataset	Regression coefficient for factors
13	Evaluation	Test the hypotheses with testing dataset Measure statistical differences regarding performance of automated QT	

4 EXPERIMENTAL DESIGN

The purpose of the experiment was to answer the RQ2: what are the factors that affect the performance of automated QT for SR? In an earlier phase, we selected eight possible factors in the context of SR based on the factors that were regarded as important for QT in DR within the field of library science. Since we selected eight possible factors, the RQ2 was divided into eight sub-questions that ask the effects of each factor as the following:

2a. Does the level of topic of interest of the answerer affect the performance of automated QT?

2b. Does the scope of subject area of interest of the answerer affect the performance of automated QT?

2c. Does the scope of topic area of interest of the answerer affect the performance of automated QT?

2d. Does the answerer's contribution to providing answers affect the performance of automated QT?

2e. Does the answerer's contribution to submitting questions affect the performance of automated QT?

2f. Does the answerer's performance in providing relevant answers affect the performance of automated QT?

2g. Does the answerer's response time of providing answer affect the performance of automated QT?

2h. Does the users' per-day question quota for answering affect the performance of automated QT?

4.1 HYPOTHESES

In order to answer the above eight research questions, the following hypotheses were defined in order to test the regression coefficients of those factors. We expected that finding a statistically important factor would increase the performance of the automated QT algorithm.

4.1.1 Test level of topic interest (x_1)

In this test, we wished to evaluate:

$H_{1\text{-}0}$: User's level of topic interest (x_1) will not affect the performance of automated QT ($\beta_1 = 0$, $P < 0.05$)

$H_{1\text{-}A}$: User's level of topic interest (x_1) will affect the performance of automated QT ($\beta_1 \neq 0$, $P < 0.05$)

4.1.2 Test the scope of subject area of interest (x_2)

In this test, we wished to evaluate:

$H_{2\text{-}0}$: User's scope of subject area (x_2) will not affect the performance of automated QT ($\beta_2 = 0$, $P < 0.05$)

$H_{2\text{-}A}$: User's scope of subject area (x_2) will affect the performance of automated QT ($\beta_2 \neq 0$, $P < 0.05$)

4.1.3 Test the scope of topic area of interest (x_3)

In this test, we wished to evaluate:

$H_{3\text{-}0}$: User's scope of topic area of interest (x_3) will not affect the performance of automated QT ($\beta_3 = 0$, $P < 0.05$)

$H_{3\text{-}A}$: User's scope of topic area of interest (x_3) will affect the performance of automated QT ($\beta_3 \neq 0$, $P < 0.05$)

4.1.4 Test contribution to answer providing (x_4)

In this test, we wished to evaluate:

$H_{4\text{-}0}$: User's contribution to providing answers (x_4) will not affect the performance of automated QT ($\beta_4 = 0$, $P < 0.05$)

$H_{4\text{-}A}$: User's contribution to providing answers (x_4) will affect the performance of automated QT ($\beta_4 \neq 0$, $P < 0.05$)

4.1.5 Test Contribution to Question Submitting (x_5)

In this test, we wished to evaluate:

$H_{5\text{-}0}$: User's contribution to submitting questions (x_5) will not affect the performance of automated QT ($\beta_5 = 0$, $P < 0.05$)

$H_{5\text{-}A}$: User's contribution to submitting questions (x_5) will affect the performance of automated QT ($\beta_5 \neq 0$, $P < 0.05$)

4.1.6 Test performance to providing relevant answers (x_6)

In this test, we wished to evaluate:

$H_{6\text{-}0}$: User's performance in providing relevant answers (x_6) will not affect the performance of automated QT ($\beta_6 = 0$, $P < 0.05$)

$H_{6\text{-}A}$: User's performance in providing relevant answers (x_6) will affect the performance of automated QT ($\beta_6 \neq 0$, $P < 0.05$)

4.1.7 Test Response Time of Answer Providing (x_7)

In this test, we wished to evaluate:

$H_{7\text{-}0}$: User's response time in providing answers (x_7) will not affect the performance of automated QT ($\beta_7 = 0$, $P < 0.05$)

$H_{7\text{-}A}$: User's response time in providing answers (x_7) will affect the performance of automated QT ($\beta_7 \neq 0$, $P < 0.05$)

4.1.8 Test quota per day (x_8)

In this test, we wished to evaluate:

$H_{8\text{-}0}$: User's daily quota in providing answers (x_8) will not affect the performance of automated QT ($\beta_8 = 0$, $P < 0.05$)

$H_{8\text{-}A}$: User's daily quota in providing answers (x_8) will affect the performance of automated QT ($\beta_8 \neq 0$, $P < 0.05$)

With the hypotheses defined above, we tested the performance of automated QT algorithm. Detailed evaluation method is discussed in the section of evaluation below.

4.2 A FRAMEWORK FOR AUTOMATED QT FOR SR

For this study, a process framework for automated QT for SR was developed as shown in Figure 10.

4.2.1 Question analysis module

The question analysis module is in charge of:

- converting the submitted question into a query for document retrieval
- identification of the topic of the question
- identification of the subject area of the question
- question filtering

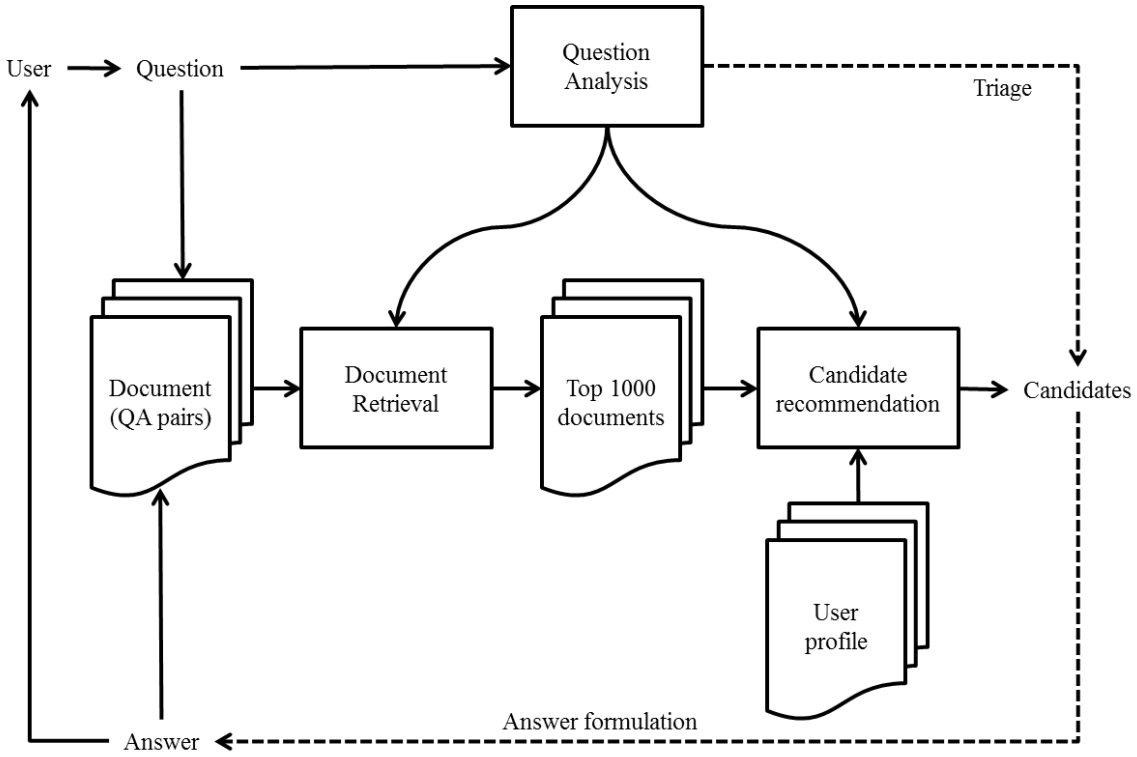

Figure 10. A process framework for automated question triage for social reference

4.2.2 Document retrieval module

This module retrieves documents that are expected to be relevant to the submitted query from the corpus. In this experiment, we selected the top 1000 documents from the search result for doing candidate recommendation.

4.2.3 Candidate recommendation module

This module retrieves possible candidates based on the documents selected. In this experiment, we considered the rank of each retrieved document as the rank of the candidate who is associated with the document. In addition to document rank of candidate, we also considered the frequency

of finding the candidate's document in the selected document set in order to promote a candidate who has more documents that are expected to be similar to the question submitted.

4.3 PROCEDURES

An overview of the procedure of the experiment is provided in Figure 11.

4.3.1 Data preparation

4.3.1.1 Selection of object

For the experiment, a SR site needed to be selected. For this study, Ask MetaFilter[6] was chosen as one example of SR sites among several online communities whose primary function is to provide a forum for users to post questions and contribute answers, such as Yahoo! Answers, WikiAnswers, etc. Many researchers who have studied online QA services have selected Yahoo Answers because of its advantages in dominating the market of SR service, such as its great number of users and question answer pairs. Yahoo Answers, however, had two limitations in terms of this study: difficulty in collecting a complete data set from the site and lack of tag (user-assigned topic) information. Ask MetaFilter is another one of the SR services. Although Ask MetaFilter is not a library-affiliated reference service, it does to a certain extent fulfill some of

[6] http://ask.metafilter.com

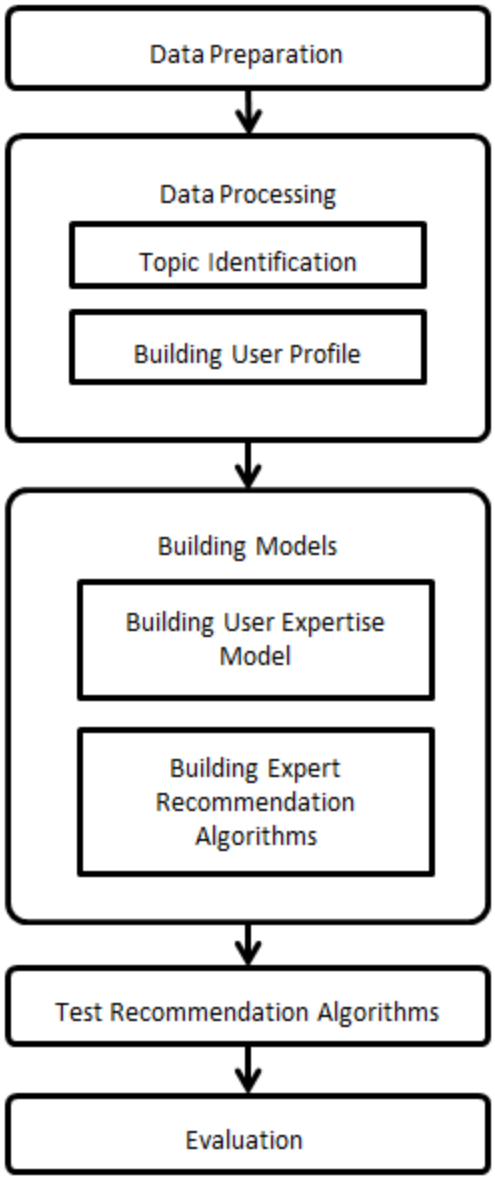

Figure 11. Overview of the experiment procedure

the same functions of human-mediated question-answering, and so provides examples of conversation-like reference transactions (Pomerantz & Stutzman, 2006). Such reference transactions can usually be combined into question-answer pairs for the study. Ask Metafilter has some advantages over Yahoo Answers for the purposes of this study ;

- Accessibility: Access to the archive is not limited, which enables the researcher to build a complete corpus for the archive.
- Tags: Availability of tag information for this study is important. They can be used to assess users' topic preferences as well as the topic of questions and answers in assessing expertise of users.
- User profile: A user's Ask Metafilter profile contains user-assigned tags, questions and answers posted by the user, and a link to the user's blog.
- Long history: Since it began on December 8, 2003, the earliest question found in the archive was posted on December 31, 2003, so the service seems to be mature.

Table 14. Summary of the dataset

	Count
Users	23,375
Questioners	14,448
Answerers	18,514
Questions	95,139
Answers	1,129,284
Categories	20
Data span	Dec. 2003 to Nov. 2010

4.3.1.2 Data collection

Questions and answers are the information artifact of knowledge from users in SR services. In order to do the research, user-associated question answer pairs were required. For the study,

question-answer pairs were collected using a crawler scripted by the researcher. Basic information about the collection is provided in Table 14.

4.3.1.3 Selection of data

For this experiment, only five categories among 20 categories were selected to make this research achievable. The five categories that were selected for this experiment were: (1) computer-Internet, (2) education, (3) health, (4) law-government, and (5) science-nature. Table 15 provides basic information about the datasets of those categories.

Table 15. Summary of categories selected

Category	Computers-internet	Education	Health	Law-government	Science-nature
Questions	17,969	2,713	6,730	2,919	1,831
Answers	128,280	32,424	93,256	32,773	21,697
Questioners	6,570	1,856	3371	2064	1,368
Answerers	10,168	6,695	9,699	5,568	4,904

4.3.1.4 Split dataset

In this study, the dataset was split into two sets of data, the training data and the testing data, following a 70-30 ratio. For the training dataset, the question-answer pairs of the first 70% of the questions in each category were selected. For the testing dataset, the other 30% of questions in each category were reserved for evaluation. The reason for using question-answer pairs for the training dataset is that answers that are associated with the question would also address the same

topics of the question. Thus, combined question-answer pairs were expected to contain more terms that represented topics of the question. However, answers that were associated with the question were not included in the testing dataset, since they would be used for evaluation later. In a real world situation, it is impossible to have answers at the point one is submitting questions to the community.

Table 16. Summary of dataset statistics for training and testing

Size / Category	**Questions**	**Training dataset (70%) - QA pairs**	**Testing dataset (30%) - Q only**
Computers-Internet	17,963	12,578	5,385
Education	2,713	1,899	814
Health	6,730	4,711	2,019
Law-government	2,919	2,043	876
Science-nature	1,831	1,281	550
Total	32,156	22,512	9,664

4.3.2 Data processing

4.3.2.1 Creating user documents

In this study, we employed an expert finding technique that uses document-relevance as evidence of user expertise. For this reason, a question-answer pair, in this study, is treated as a document. Next, each question-answer pair was associated with users who asked or answered the question. By doing this, we could generate a collection of question-answer pairs (similar to document

collections) for each user. The literature proposes two popular approaches to generate this user collection: (1) a user-centered approach and (2) a document-centered approach (Balog, 2006). In the user-centered approach, the collection of all documents that are associated with the user is regarded as a document. Thus, the user-centered approach aggregates all the term information from all documents associated with the user and uses this term to describe that user. On the other hand, we can build a user profile by selecting some documents that are associated with the candidate, assuming there is a conditional independence between the query and the user, rather than directly creating a user collection using all documents associated with the user. For this study, we adopted the document-centered approach in order to build user documents.

4.3.2.2 Creating an index of user documents for search engine retrieval

In this experiment, we used the Indri, a language model based search engine, in order to retrieve documents from the collection. In this study, the main collection for the search engine is previously archived question-answer pairs; each question-answer pair is considered to be a document. For the experiment, we combined a question and all answers to the question as a document.

4.3.2.3 Topic identification

In the SR setting, category information that is associated with questions can be regarded as the subject matter of the questions. In order to model users' subject interest in detail, it is required to identify topics of the subject categories as the subset of the subject. Since there was no explicit

topic information of each category within the given dataset, two approaches to topic detection were employed: (a) a tag-based approach and (b) a topic clustering approach.

4.3.2.3.1 Tag-based approach

One of the reasons to select Ask Metafilter as a subject of this study was that they allow questioners to associate freely determined keywords, called tags, to their questions. These user-determined keywords could be used as an instance of topics of the questions associated with them. In this stage, these tags were used as topics in order to model users' topic of interest as a portion of subject interest. Interesting observation about users' tag usage was that users often used tags for indicating their information need was solved, such as tag "resolved". Since we were interested in tags for using them as keywords that contained some subject meaning, tag "resolved" was ignored in this study.

Table 17. Most frequent top 10 tags in 5 categories chosen.

Category	Computers-internet	Education	Health	Law-government	Science-nature
Tags	*resolved^{1779} mac^{1372} computer^{1044} windows^{1013} software^{919} internet^{825} osx^{725} computers^{663} web^{613} laptop^{579}	education^{444} college^{334} *resolved^{301} school ^{294} university^{162} gradschool^{155} teaching^{141} learning^{116} math^{102}	Health^{771} *resolved^{593} exercise^{374} depression^{242} pain^{239} anxiety^{203} diet^{197} sleep^{194} fitness^{183} medical^{175}	Law ^{391} *resolved^{285} legal^{227} politics^{178} copyright^{109} immigration^{100} lawyer^{97} government^{93} landlord^{84} taxes^{68}	*resolved^{234} science^{177} physics^{110} statistics^{74} math^{72} biology^{65} chemistry^{61} evolution^{47} psychology^{46} astronomy^{45}

Note. Superscripted number is the frequency of questions associated with the tag.

4.3.2.3.2 Topic-clustering approach

Another approach to identifying sub-topics of each subject category employed in this stage was topic clustering. This approach assumes that there are latent topics in a set of documents and they can be captured by doing document clustering based on some rules can capture them. In this study, the Stanford Topic Modeling Toolbox, developed by the Stanford NLP group, was used in order to detect latent topics of documents (questions and answers). One of the advantages of using this tool is that it trains topic models using Latent Dirichlet Allocation (LDA) model that views each document as a mixture of various topics. In practice this results in more reasonable mixture of topics in a document (Blei et al., 2003). Table 19 is an example of clustered latent sub-topics of computer-internet category. Each topic is represented with keywords that appeared in the cluster. For each categories, all question and answer pairs were clustered using the Stanford Topic Modeling Toolbox in order to identify sub-topics of the category.

In order to identify user's topic of interest, we investigated user's topic distribution using the Stanford Topic Modeling Toolbox.

4.3.2.3.2.1 Setting the number of sub-topics

In order to use LDA topic modeling technique, it is required to set the number of topics as a parameter for training; those topics are hidden, so we don't know the exact number of topics in the documents. In order to set the number of topics for each category, perplexity score calculated by the Stanford Topic Modeling Toolbox was use. Perplexity is a measurement of surprise by equiprobable words that represent topics. In order to calculate perplexity score, the toolbox "splits a document into two subsets: one used for training models, the other used for evaluating

their perplexity on unseen data. Perplexity is scored on the evaluation documents by first splitting each document in half. The per-document topic distribution is estimated on the first half of the words. The toolbox then computes an average of how surprised it was by the words in the second half of the document, where surprise is measured in the number of equiprobable word choices, on average" (Markgren, 2006). Thus, a lower perplexity score indicates better generalization performance.

In order to identify the best number of topics for a category, we clustered each category with the number of topics between 10 and 45, and then selected the number of topics which perform the best (the smallest score of perplexity). Figure 12 shows how the perplexity scores are fluctuated while the number of topic is changing. Once we got these perplexity scores then we selected the number of topics that were associated with the lowest perplexity score.

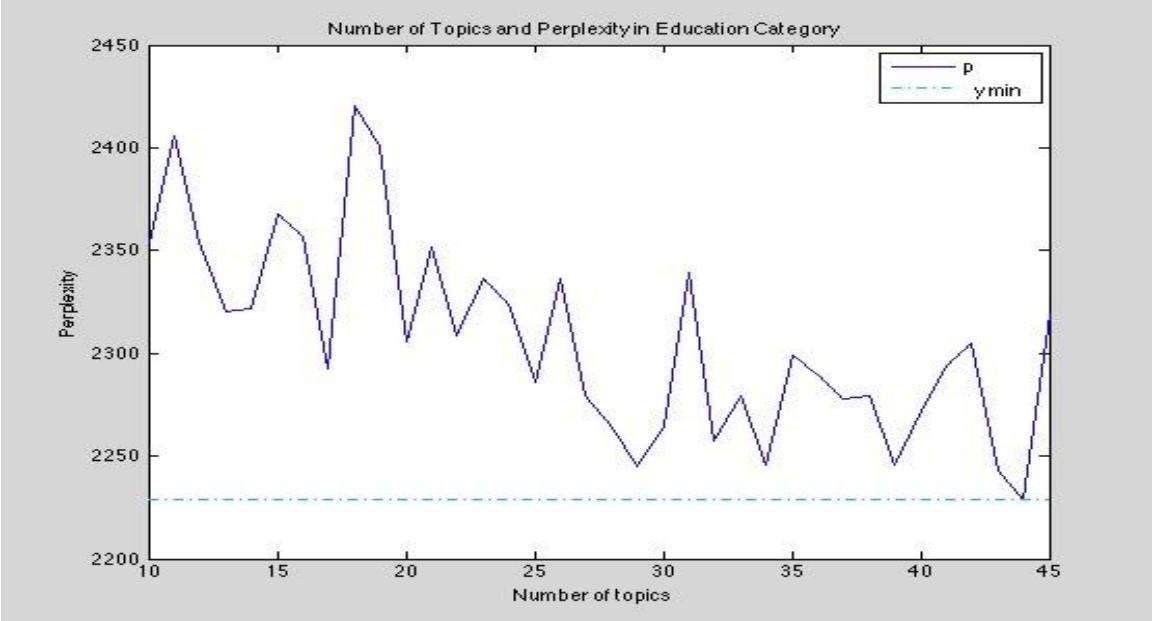

Figure 12. Number of topics and perplexity in education category

Table 18. Selected number of topics for each category

Category	# of topics
Computers-internet	44
Education	44
Health	43
Law-government	39
Science-nature	30

4.3.2.3.2.2 Document-topic distribution

Once we do clustering using LDA algorithm, we can get topic distribution of each document. Our goal in this step for this study was to select candidate answerers based on the document (question-answer pairs associated with answerers). For the reason, we used topic distribution of each document as the evidence of users, who are associated with those documents, would be a member of the topics. Figure 13 illustrates topic distribution of question 3905 and question 3947 posted in education category. Question 3947 has probability of being a member with topic 2, 5, 14, 18, 20, and 24, while question 3905 has probability of being a member with topic 4, 8, 9, 13, 20, 29, 33, and 37.

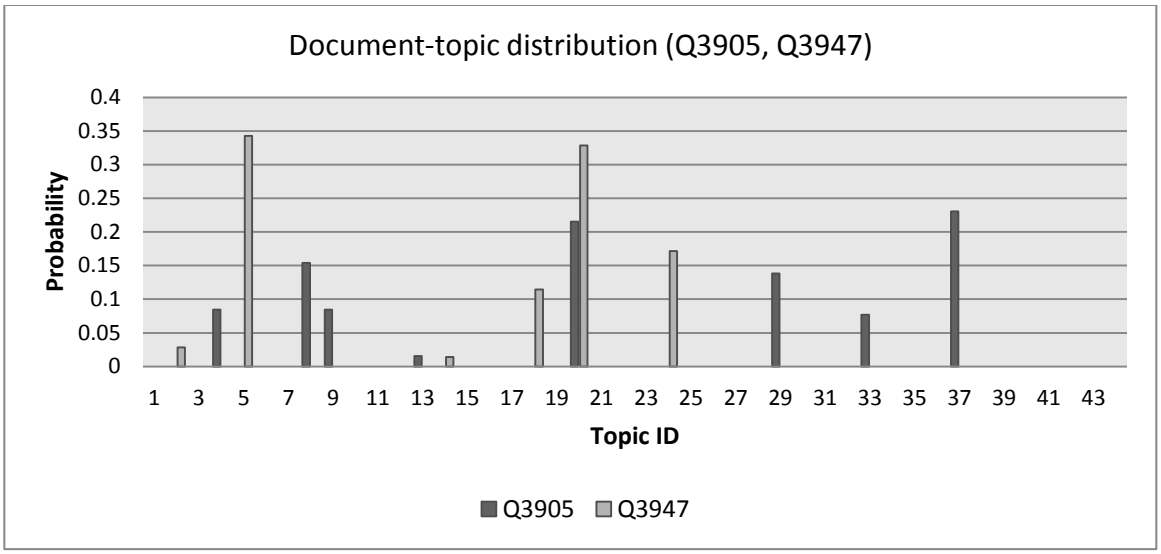

Figure 13. Example of document-topic distribution: question 3905, question 3947

4.3.2.3.2.3 User-topic distribution

In the next step, we tried to get user-topic distribution based on the document-topic distribution. The purpose of this step was to model users' topic of interest. The membership scores were calculated by the sum of documents that are associated with the user to each topic. Figure 14 shows user-topic distribution of two users (user 14648 and 1104) to each topic. This graph shows not only the user-topic distribution but also users' contribution to the topic, since the membership score is the sum of documents' topic distribution. User 14648 demonstrated a lot of contribution to the topic 18, and user 1104 contributed more to topic 5 and 7 than user 14648 did to those topics.

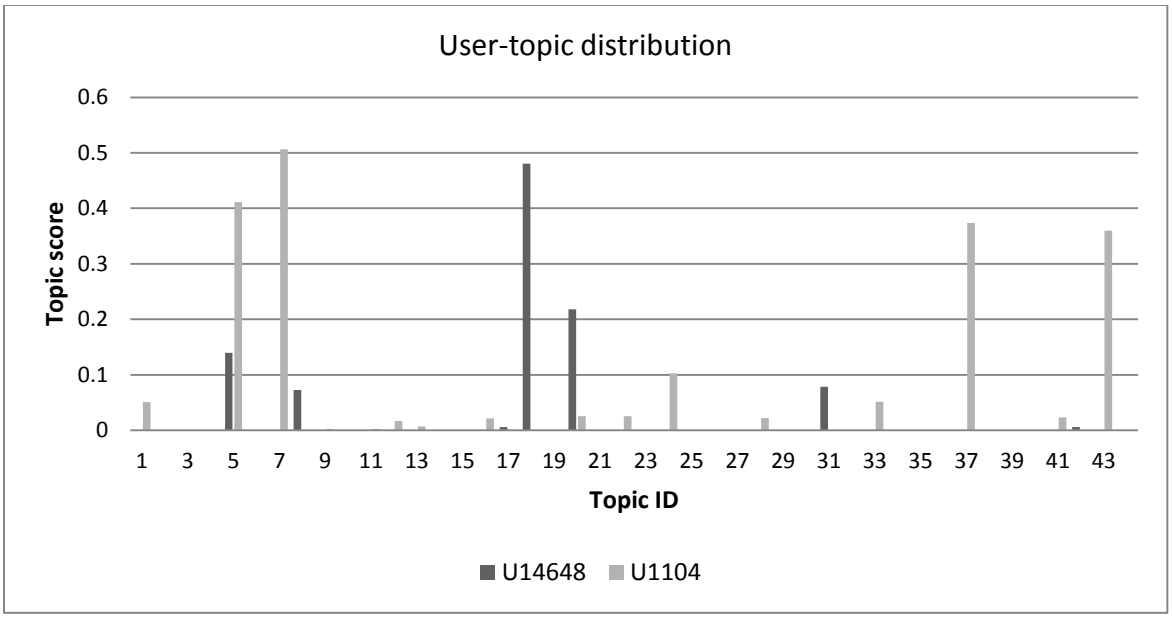

Figure 14. User-topic distribution (user 14648 and 1104)

4.3.2.4 Creating user expertise profile

A prerequisite for developing QT systems that recommends candidate answerers to the submitted questions is user expertise profile. User expertise profile, here, is a representation of expertise of any individual candidates. Roughly, an expert profile is a structured representation of the candidate's expertise through which a QT system should able to locate relevant candidate answerers to the given question based on that profile applying some algorithms.

Table 19. Example of the first four topics of computer-internet category using LDA model.

	Topic 1	**Topic 2**	**Topic 3**	**Topic 4**
Keywords	word^{3078} text^{2566} document^{1410} office^{815} character^{707} [oov]^{675} editor^{616} format^{478} type^{449} microsoft^{435} write^{425} save^{424} format^{413} doc^{404} version^{376}	server^{4797} network^{1559} share^{1298} machine^{1271} client^{1268} access^{1204} remote^{1193} port^{1031} set^{988} ftp^{805} firewall^{731} home^{721} runn^{711} ssh^{696} run^{672}	school^{748} student^{624} kid^{562} presentation^{410} year^{386} slide^{373} powerpoint^{351} parent^{301} people^{280} univers^{278} online^{271} college^{269} friend^{260} fami^{255} learn ^{249}	learn^{2215} code^{2016} language^{1943} programm^{1593} java^{1280} book^{1254} write^{852} python^{826} php^{704} basic^{677} program^{671} web^{633} net^{622} start^{593} ruby^{570}

Note. Super-scripted number is the frequency of the term appeared in the collection.

4.3.2.4.1 User Expertise Modeling

An expertise modeling process requires two steps which constitutes the expert profile modeling methodology:

1) What has to be represented, that is which information pertaining to the candidate has to represented, and
2) How this information is effectively represented.

Since the goal of this step is to locate relevant candidate answerers for QT systems, it is quite clear that in an expertise model we have to represent the candidate's expertise.

In the current study, we defined a framework of user expertise in chapter 3 using three concepts: knowledge, performance and contribution.

4.3.2.4.1.1 A framework of user expertise for SR

In any expert finding application, a fundamental question is often “who is an expert?” This question again raises another question “what is expertise?” In order to define expertise, a framework of user expertise in the context of SR is proposed earlier (section 3.2.3). In this section, the approaches to model user expertise, employing the framework of expertise proposed earlier, is addressed.

4.3.2.4.1.1.1 Subject interest

In this research we wanted to locate experts who could give answers to the question in order to route the question to the experts. Thus, the first dimension for judging user expertise is the knowledge about the subject or topic of the given question.

We observed in a preliminary manual examination of question-answer pairs in the main dataset that the users’ subject interest is different from one another. Furthermore, the users’ topic of interest in a subject is different from one another. While some users show interests on diverse subjects or topics in providing answers, another group of users are interested in only small number of subjects or topics in asking questions and providing answers. Thus we need to model this diverse subject interest and topic interest of the user in the profile for QT. In this study, a user’s subject interest K is represented as the following:

- A user's subject interest *K* is composed of one or more subject areas *S*. In this study, a category to which a question is posted is regarded as a subject area of the question.

$$K = \{S_1, S_2, \dots, S_m\} \tag{1}$$

Employing this model, the scope of subject area can be inferred from calculating the number of subject areas (|*K*|)

- A subject area *SA* is composed of one or more topic areas *TA*.

$$\mathrm{S} = \{\mathrm{T}_1, \mathrm{T}_2, \dots, \mathrm{T}_\mathrm{n}\} \tag{2}$$

As the same way, the scope of topic area can be inferred from calculating the number of topic areas (|*S*|)

4.3.2.4.1.1.2 Performance

We observed different performance in QA among users in a preliminary manual examination of question-answer pairs in the main dataset. In this study, user performance can be measured with the amount of good answers selected by the questioner and other users. Thus, a user's performance *P* can be represented as the following:

- A user performance *P* is a combination of good answers of the user.

$$P = \{a_{b1}, a_{b2}, \dots, a_{bn}\} \tag{3}$$

- The user's performance in providing good answers can be inferred from counting the number of good answers of the user (|P|).

4.3.2.4.1.1.3 Contribution

The researcher observed that user contribution to the community was made in two distinct activities - questioning and answering - in SR, and there is difference in user's preference to those activities. Some users showed preference to question asking, while another group of users showed preference to QA. Furthermore, there is difference in users' both activities in each subject areas. In this study, a user's activity *A* is represented as the following:

- A user activity *V* is composed of two roles, questioning activity *Q* and answering activity *A*.

$$V = \{Q, A\} \tag{4}$$

- An activity is the combination of user activities made in each subject areas.

$$Q = \{q_1, q_2, \dots, q_m\} \tag{5}$$

, where q_m means the user's question activities *Q* in a subject area SA_m.

$$A = \{a_1, a_2, \dots, a_m\} \tag{6}$$

, where a_m means the user's answering activities *A* in a subject area SA_m.

In addition to questioning and answering activities *V*, response time for answering *R*, and number of questions answered by the user per day *N* are included in experience category. Thus, user contribution *C* can be modeled as the following:

$$C = \{V, R, N\} \tag{7}$$

All in all, an expert profile *U* can be represented as the combination of subject interest *K*, performance *P*, and contribution *C* as the following:

$$U = \{K, P, C\} \tag{8}$$

In this study, a relevant expert who can provide answers to the given question was expected to possess relevant subject interest on the subject or topic of the question. Also, relevant expert was expected that s/he usually contribute to the community providing answers rather than asking questions. Additionally, the number of questions answered in the day must not exceed the maximum number of questions that the user can provide answers per day. If the quota is already exceed the limit, it would result in the failure of getting answers from the expert or extension of turnaround time for answer without estimation.

4.3.2.4.1.2 Calculate the Attributes of User Expertise

4.3.2.4.1.2.1 Scoring user's level of topic knowledge

In tag-based approach, user's topic knowledge score was calculated just simply counting the frequency of the user-answered-questions that are associated with tags of the question given. While, in topic-clustering approach, user's topic score was calculated using the user-topic distribution created in the previous stage.

4.3.2.4.1.2.2 Scoring user's scope of subject area

The scope of user's subject area was simply counted by the number of subjects which were associated with the user.

4.3.2.4.1.2.3 Scoring user's scope of topic area

The scope of user's topic area was simply counted by the number of topics which were associated with the user. In tag-based approach, the number of topics is actually the number of

tags that are associated with the user. In topic-clustering approach, the number of topics was calculated by counting topics that has probability larger than certain threshold from the topic distribution of the user.

4.3.2.4.1.2.4 Scoring user's contribution

User's contribution to providing answers was simply counted by the number of answers provided by each user. User's contribution to submitting questions was simply counted by the number of questions submitted by each user.

4.3.2.4.1.2.5 Scoring user's performance

User's performance to providing good answers was calculated by the number of answers which were selected by other users as "favorite answer". Response time is simply calculated by the mean of response time by each user. Ability to handle questions per day was counted by the mean of the number of questions which were answered by the user.

4.3.3 Candidate recommendation

The goal of this stage was to locate relevant answerers assessing their expertise, and then recommend them in an order of relevance. Thus, this stage was composed of three steps: (1) topic detection of the question submitted, (2) select pool of candidates with possible answerers and (3) ranking them in an order of relevance.

4.3.3.1 Topic identification of the question

Once a question is given, we first identified the subject and topics of the question. Since a question is submitted to a subject category in real world, we simply used that category information that is associated with the question as a subject area assigned to the question. Next, sub-topics of the question were investigated using two different approaches that were used for topic identification.

In tag-based approach, the tags associated with the question were simply regarded as sub-topics of the question. On the other hand, the topic detection of the question in topic-clustering approach is more complicated. In topic-clustering approach, sub-topics of the question were investigated using LDA topic model. In order to determine topics of the question, we set a threshold to limit the number of topics less than 5. For example, the question 3947, shown in Figure 13, has positive probability to be a member with 6 topics among 44 topics in education subject area. But, the probabilities to topic 2 and 14 are relatively very low compared to topic 5, 18, and 24. We tried to cut off those topics by setting threshold. For the experiment, we set the threshold of the probability as 0.15. This threshold reduced the number of users' topic area of interest to lesser than 5. Thus, the topics of question 3947 were determined as topic 5, 18, and 24.

```
Determine_Topics(Question Q)
Identify SubjectArea SA of the Q
Get LDA Topic Model TM_train,which was trained with training dataset, for the SA
Calculate topic distribution of Q using TM_train.
Using the topic distribution,
Select Topic T its probability is higher than the threshold.
Return selected topics Ts.
```

Figure 15. Pseudo code for determining topics of the question (topic-clustering approach)

4.3.3.2 Selecting pool of candidates

The researcher observed in a preliminary manual examination of question-answer pairs in the main dataset that there were a lot of candidates who were associated with a subject category. For instance, in the category of education, there are 6,695 potential candidates (answerers), and this number is too large to recommend them to a question submitted. The goal of this step was to narrow down the size of candidate pool, including relevant answerers as candidates, as possible as we can. In order to locate relevant candidates, narrowing down its size, the researcher used *document-based approach*. In order to obtain feasible document to the given question, we used Indri, an language model search engine, as a search tool.

Figure 16 shows a pseudo code for the document-based approach to selecting candidate experts using candidate-centered model.

Relevance score (*Rel*) for each candidate was calculated as the following:

Equation 1. Relevance score (Rel)

$$Rel_u = \left(\sum_{d=1}^{n} (N - rank + 1) * a(d,u) \right) * F$$

, where *a(d,u)* is the association between the document and the user, and *F* is the frequency of documents which is associated with the user.

This association was calculated as the following:

$$a(d,u) = \begin{cases} 0, & \text{if } d \text{ is not associated with } u \\ 1, & \text{if } d \text{ is associated with } u \end{cases}$$

Frequency score (*F*) of the user's document in the retrieved result set was normalized as the following:

Equation 2. Frequency score (F)

$$F = \frac{\sum_{d=1}^{n} a(d,u)}{N}$$

Using document-based approach using search technique, we could get another model for user topic interest. An assumption here is that retrieved documents and the query terms share a similarity of topics. In this approach, we used the question as a query term to the search engine, and then select top n documents from the result. Finally, we retrieve all candidates who are associated with the documents.

```
Find_Candidate_Group(Question Q)

    For the given Q:

        Retrieve top n documents d with the Q as the query, where d is all question-
    answer pairs that are associated with a candidate.

        Retrieve all candidates Cs who are associated with d.

        For each candidate C in Cs:

            Calculate relevance score

        Return Cs in a descending order of relevance score of d associated with each C.
```

Figure 16. Pseudo code for candidate selection - document-based approach

4.3.4 Generating dataset for logistic regression analysis

In order to investigate important factors affecting to decisions of QT for SR, it needed to generate training dataset that contains relevant cases and non-relevant cases to each training question, using the user expertise profile created earlier. Total 22,512 questions were used to generate training dataset and 9,664 questions were used to generate testing dataset

4.3.4.1 Conditions to generating cases for training

4.3.4.1.1 Topic modeling approach: the tag-based vs. the topic-clustering-based

In this study, we tested two different approaches to model user's topic of interest: tag-based approach and topic-clustering-based approach. This resulted in different set of samples for the training.

4.3.4.1.2 Balancing

Examining the generated samples for training, we observed that there was big difference in the number of cases between the relevant cases and the non-relevant cases. In order to make the number of cases between them even, we tried to append randomly selected relevant cases to the sample dataset.

4.3.4.1.3 Limit the number of candidates (N)

In our collection, it was observed that average number of candidates in a subject area is about 7,000. Thus, if we generate the training dataset by including all relevance information to all 7,000 candidates for each question in training question, it would resulted in a lot of cases with a lot of noised in the dataset. In order to generate good samples for the training, we selected top N number of candidates for each question.

4.3.4.1.4 Precision-filtering

We also tried to generate good samples by using good questions selected based on precision. Precision-filtered questions had a least one relevant answerers in the candidate top N candidates. By applying above conditions, we generated different set of datasets for the evaluation.

4.3.4.2 Normalize the score

In order to understand the characteristic of our dataset in user profile, we examined the distribution of each attribute, and found that it was hard to investigate the difference between relevant answerer and non-relevant answerer using the raw score of each attribute (Figure 17). For the purpose of understanding the characteristics of the dataset, we transformed raw dataset to z-score. Z-score transformed the raw data based on the normal distribution of the dataset, it provides relative scores to the distribution.

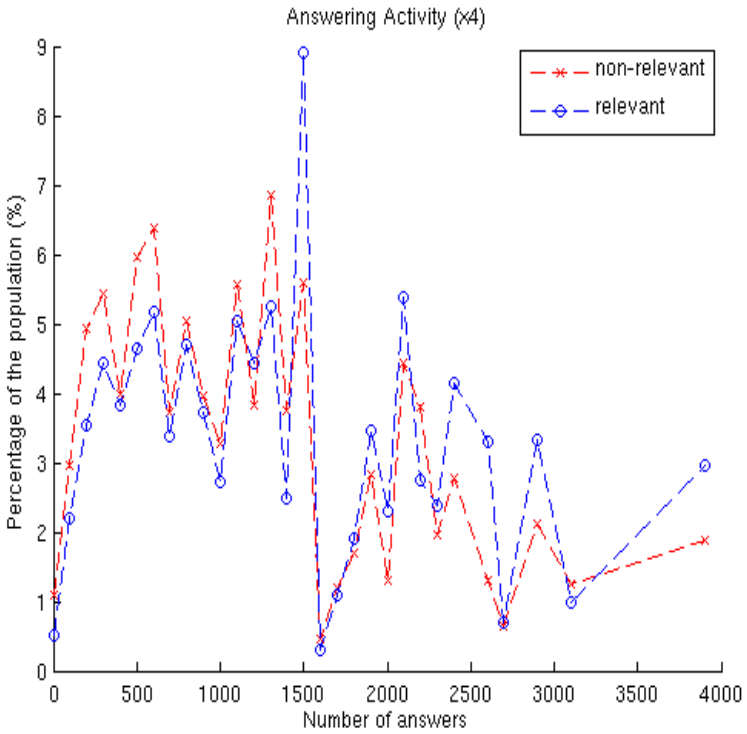

Figure 17. Raw distribution of answering activity

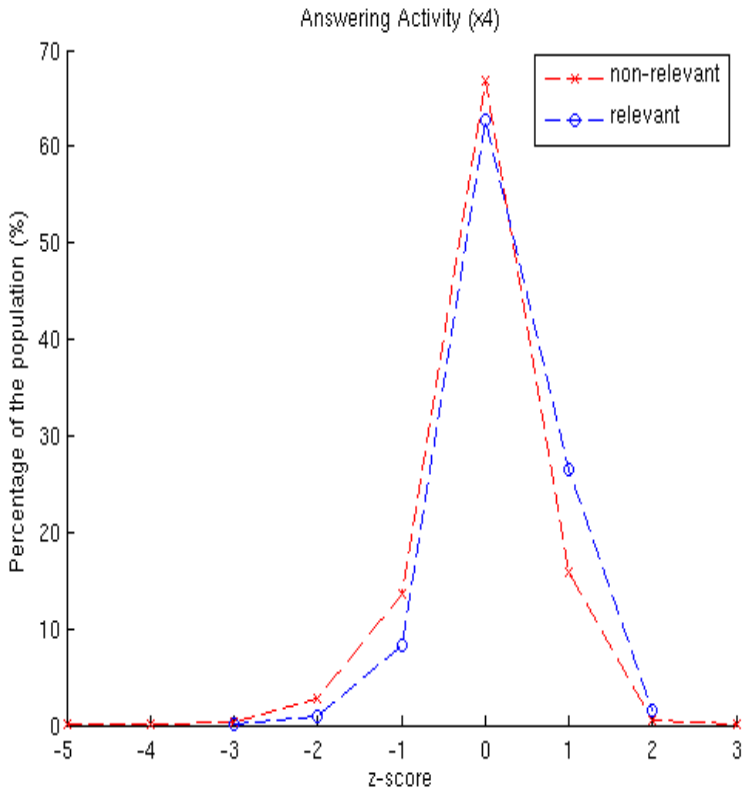

Figure 18. Z-distribution of answering activity

Z-scored distribution provides further understanding of the characteristics of the dataset (Figure 18). In this case, in the group of answerers who were in top 16% population in terms of the number of answers, the percentage of relevant answerer was increased while the percentage of non-relevant answerers was decreased.

4.3.5 Evaluation

4.3.5.1 Measures for performance evaluation

For the evaluation, we used precision and Mean Average Precision (MAP) in order to measure the performance of the algorithm. Precision (P) is the fraction of the retrieved candidates (the set A) which is retrieved i.e.,

Equation 3. Precision (P)

$$P = \frac{|Ra|}{|A|}$$

, where |*Ra*| is the number of relevant candidates in the candidates (set A).

Average Precision is the average of the precision value obtained for the set of top k documents existing after each relevant document is retrieved, and then this value is averaged over information need (relevant documents)

Equation 4. Average Precision (AveP)

$$\text{AveP} = \frac{\sum_{k=1}^{n}(P(k) \times rel(k))}{|Ra|}$$

, where rel(k) is an indicator function:

$$rel(k) = \begin{cases} 0, & \text{if candidate at rank } k \text{ is not relevant} \\ 1, & \text{if candidate at rank } k \text{ is relevant} \end{cases}$$

MAP is the mean of the average precision scores for each query for a set of queries.

Equation 5. Mean Average Precision (MAP)

$$\text{MAP} = \frac{\sum_{q=1}^{|Q|} AveP(q)}{|Q|}$$

, where Q is the number of queries.

4.3.5.2 Relevance judgment

In order to judge relevance of a candidate recommended, we needed ground truth that is a basis of the relevance decision. Since our training dataset contained information about the association between question and user, a real answerer to the given question was interpreted as a somewhat relevant answerer to the question. Among them, some answerers may be marked as 'favorited' answerer to a question; they are interpreted as highly relevant answerers to the question.

4.3.5.3 Evaluating the impact of factors on the performance

In order to evaluate the impact of factors on the performance of the automated QT system, logistic regression analysis was used. Once we obtained the best logistic model in terms of performance, then we could do analysis for the evaluation using the model.

Logistic regression gives each predictor (IV, a factor) a coefficient 'β' which measures its independent contribution to variations in the dependent variable (relevance of the user to the question). What we wanted to predict from knowledge of relevant independent variables and coefficients is the probability (P) that it is 1 rather than 0 (belonging to one group rather than the other). For this study, the probability of the logistic regression model can be interpreted as the probability of a candidate to be a relevant answerer to the question assigned. In this study logistic

regression was used to determine if the factors selected can be used to predict whether or not a candidate is relevant answerer to the question assigned.

4.3.5.4 Variables

The variables used for the evaluation are listed in Table 20.

Table 20. Variables used in the evaluation

	IV	**DV**
Relevance		×
Level of topic interest (x_1)	×	
Scope of subject area (x_2)	×	
Scope of topic area (x_3)	×	
Contribution to providing answers (x_4)	×	
Contribution to asking questions (x_5)	×	
Performance to providing good answer (x_6)	×	
Response time (x_7)	×	
Quota per day (x_8)	×	

4.3.5.5 Baseline

For the evaluation, baseline was modeled using a relevance score that combines the candidate's ranking scores and frequency of the candidate's documents in the search engine result set to the questions (Equation 1).

4.3.5.6 Logistic regression models

Since we tried to investigate the impact of the factors selected earlier on the performance of automated QT for SR, we defined the following logistic regression model for the evaluation.

Equation 6. The logistic regression model (full model)

$$\text{Predicted logit} = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5 + \beta_6 \cdot x_6 + \beta_7 \cdot x_7 + \beta_8 \cdot x_8 + \beta_9 \cdot x_9$$

, where $x_1 \ldots x_9$ are the values of each factors; $\beta_1 \ldots \beta_9$ are the regression coefficients of factors, β_0 is the intercept from the linear regression equation.

In the above model, we added relevance score (x_9) by Equation 1 in the model since we tried to estimate relevant answerers among candidates who are recommended by the search engine result using Equation 1. Thus, the last factor, relevance score, can be understood as a default factor. For this reason, the baseline can be represented using the logistic regression model as the following:

Equation 7. The logistic regression model for baseline

Predicted logit

$$= \beta_0 + 0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4 + 0 \cdot x_5 + 0 \cdot x_6 + 0 \cdot x_7 + 0 \cdot x_8 + \beta_9 \cdot x_9$$

$$= \beta_0 + \beta_9 \cdot x_9 \approx x_9$$

From the logistic regression analysis of relevance score (x_9), we observed that the values of β_9 are always positive. Since β_9 is always positive and β_0 is constant value, the predicted logit score for baseline is proportional to x_9 unless $\beta_9 < 0$.

5 RESULTS

5.1 PHASE 1: FACTOR IDENTIFICATION FOR QT FOR SR

5.1.1 Commonality and differences between DR and SR

In order to study QT for SR, it needs to have a framework to understand the process of QA in the SR setting. Before presenting the idea of QT for SR, we first should explore QT for DR in the field of libraries, since there are commonalities and differences in the QA process between DR and SR.

5.1.1.1 Core elements and the QA process

From our investigation, we identified four core elements of the QA process: the questioner, the answerer, the question and the answers. We also identified subordinate elements, such as triager and system, involved in the QA process. An overview of these approaches is illustrated in Figure 5. In the overview, it was shown that the QA process of DR is very close to that of SR except that the triage and information source were not involved in the QA process of SR (Figure 5). This finding supported the goals of this study, including developing an automated QA process for social reference based on DR.

5.1.1.2 User as both questioner and answerer

One of the key elements that are involved in the process of QA is the user or expert. In this study, it was found that the role of the user in SR was equivocal. We observed that the user was involved in the QA process as both the questioner and the answerer. This is different from the user's role in DR. In library settings, the user can be clearly distinguished from the expert, the reference librarian or the answerer since the user only asks questions to the service.

5.1.1.3 Question and answer

As the result of the user's equivocal roles in SR setting, both questions and answers are produced by the user. Thus, the user shows two distinct activities, questioning and answering, in the community. In order to assess the user's expertise, the model needs to include both the questioning and answering activities of the user.

5.1.2 Selecting attributes affecting to triage decision

In order to investigate important factors affecting the decision making of QT for SR, we explored fifteen of the most important factors of QT in DR. Since DR services are often provided by libraries or library-affiliated institutions, the majority of these factors were attributes of a library reference service. Since the answerer is the only actor who provides answers to the question, the SR model needed to define new attributes of the answerer by adopting and redefining important triage factors found in DR to the context of SR.

For this study, we defined useful attributes of the question and the answerer as the following:

1. The question's subject area and topic area
2. The subject and topic areas of the answerer's interest,
3. The scope of subject and topic areas of the answerer's interest,
4. The answerer's level of subject and topic interest,
5. The answerer's activities in providing answers and asking questions,
6. The answerer's response time to answering questions and
7. The number of questions answered previously by the answerer (quota).

5.1.3 Framework of user expertise

In order to use the attributes of the answerer for decision making of QT for SR, they needed to be arranged in a consolidated manner. In this study, a framework of expertise using three concepts—knowledge, performance, and contribution—was proposed in section 3.2.3. This framework can provide a generic perspective to assess the user expertise for QT within the SR setting.

5.2 PHASE 2: EXPERIMENT AND EVALUATION

5.2.1 The logistic regression model

In the previous chapter, we defined a logistic regression model that includes nine factors that affect estimation of relevant answerers to a given question (Equation 6), and a baseline model that includes only the ninth factor generated as default by the search engine (Equation 7).

5.2.1.1 The full model

In order to choose the best performing models, we first performed a logistic regression analysis using the full model with a different set of training dataset. Next, the resulted regression coefficient values (weight of each factor) were applied using equation 6 in order to estimate relevant answerers to the given questions. This application resulted in changes in the order of candidate expertise. Finally, we calculated their performance using Mean Average Precision (MAP).

5.2.1.2 Baseline

In the previous chapter we developed a logit function for baseline as the following under the condition of $\beta_9 > 0$.

$$\text{Predicted logit} = \beta_0 + \beta_9 \cdot x_9 \approx x_9$$

In order to prove this equation, we performed a logistic regression analysis using the single factor (x_9) with our training dataset, and we observed that the value of β_9 was always

positive. Since we were interested in the order of candidates in the given candidate pool, the constant value (e.g., -4.465 in Table 21) and positive coefficient of x_9 are meaningless since it doesn't affect the order of the ranking. For this reason, we simply used the order of x_9 as the evidence of user expertise as baseline in order to evaluate the proposed logistic regression.

Table 21. Logistic regression analysis of TAG-B0F1N20 dataset using a single factor for baseline

Predictor	β	SE β	Wald's X^2	df	Sig.(p)	Exp(β)
Constant	-4.465	.042	11190.521	1	.000	.012
Indri score (x_9)	.565	.021	755.408	1	.000	1.760

5.2.2 Imbalanced dataset

During the evaluation, we observed there was a lot of noise in our raw dataset regardless of topic identification approaches (whether using tag-based or topic-clustering approaches). For example, in order to prepare a set of cases for a given question in the education category, we had to create 6,695 cases of user profiles, and only about 10 cases among them were considered as relevant cases. This characteristic of the dataset results in a highly unbalanced training dataset. This problem can sharply underestimate the probability of rare events (relevant cases) (King & Zeng, 2001). In order to reduce this risk using our dataset, we collected data with the top N number of candidates to the given questions.

Table 22. Ratio between relevant cases and non-relevant case in the training dataset

		Top 20 (N20)	**Top 30 (N30)**	**Top 40 (N40)**	**Top 50 (N50)**
TAG	**Relevant**	3448	4310	4997	5559
	Non-relevant	114232	169450	223723	276891
	Ratio	.030:1	.025	.022:1	.020:1
TC	**Relevant**	3839	4850	5659	6432
	Non-relevant	126201	190210	254421	318668
	Ratio	.030:1	.026:1	. 022:1	.020:1

Note. The ratio expresses the relationship between the number of relevant cases and the number of non-relevant cases.

5.2.3 Data Normalization

During the experiment, a lot of inconsistencies in the scale of the values in the dataset were found. This problem was due to the characteristics of our dataset. In order to build a training dataset for the analysis, we collected the top N user profiles for answerers to each question in the main collection. In this situation, generating a set of user profiles to the given question can be regarded as a way of doing an independent experiment. That is, two different sets of user profiles of two different questions are independent from each other, thus they have different distribution of values for each factor. For this reason, we needed to normalize those values in the dataset. We used log and z-score to transform those values. Table 23 shows the difference among three different data types.

Table 23. Comparison of data type: raw, log, and z-transformation

	Raw data	**Log data**	**Z score data**
Level of topic of interest (x_1)	39.63	1.5980	1.041
Scope of subject area of interest (x_2)	17.0	1.2304	-.441
Scope of topic area of interest (x_3)	44	1.6435	.690
Answering activity (x_4)	446	2.6493	.069
Questioning activity (x_5)	11	1.0414	.0133
Relevant answering (x_6)	3	0.4771	-1.489
Response time (x_7)	2.48	0.3945	-.702
Quota (x_8)	2.0	0.3010	.842
Indri score (x_9)	140.46	2.1476	3.415

5.2.4 Training

5.2.4.1 Selecting the best conditions

Throughout the logistic regression analysis, we selected the tag-based approached model to evaluate its performance using the dataset.

Once we obtained the coefficient of each factor for the logit function, we could test the model to evaluate its performance using a real dataset. In order to do the evaluation, we used a full logistic regression model that incorporates all nine factors as a way of re-ranking candidates, expecting relevant answerers to be promoted, based on the logit function in the result set returned by the search engine.

In this experiment, the best model was chosen based on performance. Performance is about how successfully the model estimated relevant answerers among the candidates that were top N users returned by the search engine. In order to measure performance, Mean Average Precision (MAP) was calculated for the given set of questions.

5.2.4.1.1 Selecting the best pool size (N)

As the nature of our full dataset, our study can be considered as a rare event case study in which the number of relevant cases (events) is very small. In order to develop a realistic regression model, we had to devise different sets of training data using the top N number of user profiles to the given questions. As discussed earlier, we selected the top 20, 30, 40, and 50 candidates in order to prepare the training dataset, and the maximum percentage of relevant cases of the dataset is about 0.3 percent (Table 22). In order to choose the best model we had to decide the pool size (N) first.

Figure 19 shows that the MAP of baseline decreased as the candidate pool size (N) was increasing. This seemed to be the result of the noise in the candidate pool retrieved by the search engine as default. In fact, the number of relevant answerers to a question in the training dataset is very small. As a result, the precision decreased as the candidate pool size was increasing. Table 24 also shows this trend in the trained models. For this reason, we selected N20, the smallest candidate pool size (N=20) in order to test the hypotheses since it performed as the best among the test models in terms of MAP

.

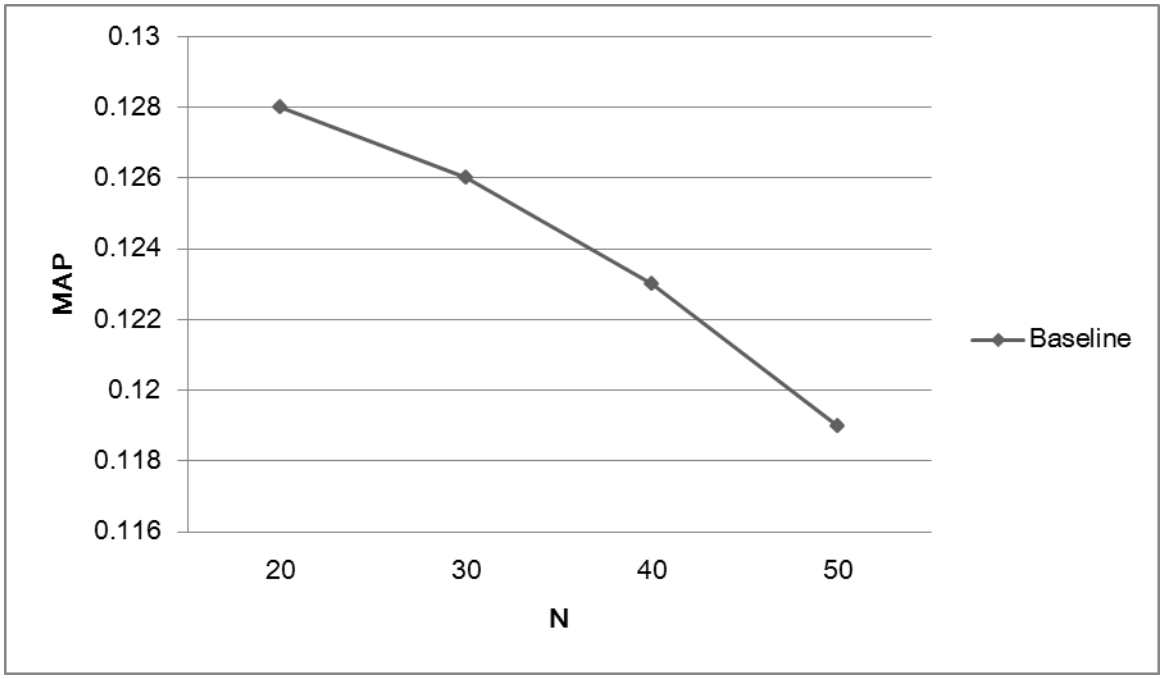

Figure 19. Changes in MAP as candidate pool size (N) is increasing (baseline)

5.2.4.1.2 Balancing Cases: No-balancing (B0) vs. balancing (B1)

For each paired model, a paired-samples t-test was conducted to compare the precision of the models in balanced (B0) and not-balanced (B1) conditions. Table 24 shows the difference in performance between the unbalanced model (B0) and balanced model (B1) for each paired model.

Table 24. Performance comparison (MAP) of each model that used different set of training data with different conditions.

Model \ Dataset		Train			
		N20	**N30**	**N40**	**N50**
TAG-B0F0	**MAP**	.141*a)	.141*	.138*	. 134*
TAG-B1F0	**MAP**	.135	.136	.133	. 128
TAG-B0F1	**MAP**	.366*b)	.318*	.287*	.267*
TAG-B1F1	**MAP**	.351	.302	.278	.256
TC-B0F0	**MAP**	.134*c)	. 131*	.128*	. 123*
TC-B1F0	**MAP**	.131	.125	.122	.117
TC-B0F1	**MAP**	.336*d)	.287*	.262*	.238*
TC-B1F1	**MAP**	.327	.272	.247	.226

Note:
* indicates significant statistical difference between B0 and B1.

Throughout the paired-sample t-test, we observed that:

a) There was a significant difference in MAP between TAG-B0F0N20 (M=0.141, SD=0.252) and TAG-B1F0N20 (M=0.135, SD=0.246) conditions; t(5872)=4.632, p=0.000.
b) There was a significant difference in MAP between TAG-B0F1N20 (M=0.366, SD=0.297) and TAG-B1F1N20 (M=0.351, SD=0.299) conditions; t(2079)=3.955, p=0.000
c) There was a significant difference in MAP between TC-B0F0N20 (M=0.134, SD=0.244) and TC-B1F0N20 (M=0.131, SD=0.242) conditions; t(5872)=2.351, p=0.019.
d) There was a significant difference in MAP between TC-B0F1N20 (M=0.336, SD=0.291) and TC-B1F1N20 (M=0.327, SD=0.293) conditions; t(2079)=2.715, p=0.007.

5.2.4.1.3 Selecting the best performing topic modeling approach: TAG or Topic clustering (TC) ?

The above two with-in-approach comparisons proved that case balancing between relevant cases and non-relevant cases did not improve the performance regarding MAP. In the next step, a comparison between two different topic modeling approaches, a tag-based approach (TAG) and a topic-clustering-based approach (TC), was performed in order to select the best approach between them. Since the previous result supported that B0 models performed better than B1 models, we tested only B0 models of the two approaches. For this comparison, a paired t-test was performed between the models of the two approaches, and we observed a significant difference between the tag-based approach (TAG) and the topic-clustering-based approach (TC);

- There was significant difference in MAP between TAG-B0F0N20 (M=0.141, SD=0.252) and TC-B0F0N20 (M=0.134, SD=0.244) conditions; t(11744)=1.505, p=0.132.
- There was significant difference in MAP between TAG-B0F1N20 (M=0.366, SD=0.297) and TC-B0F1N20 (M=0.336, SD=0.291) conditions; t(4158)=3.375, p=0.001.

These results supported that the tag-based approach (TAG models) achieved better performance than topic-clustering-based approach (TC models). Based on this result, we rejected TC models in favor of TAG models. All in all, the TAG-B0F0N20 and TAG-B0F1N20 models were selected for the evaluation of the experiment.

5.2.4.1.4 Sampling Dataset: Precision-filtering (F0) vs. No precision-filtering (F1)

A two-sample t-test was performed to compare the performance of a precision-filtered model (F0) and a not-precision-filtered model (F1). While F0 models trained with the top N user profiles to all the questions in the training dataset, F1 models trained with the top N user profiles to only selected questions in the training dataset, based on its precision; those questions were selected only if its precision at N was positive. For this feature of the F1 model, the number of training questions was different from each other. In order to compare the difference in performance between the two models, we evaluated the models with the questions of F1 models since they were a subset of questions of F0.

These results suggest that precision-filtered model (F1) performs better than not-precision-filtered model (F0). Thus, precision-filtered model (F1) was favored for this study.

Table 25. Performance comparison between F0 and F1 models

Model \ Dataset		Train			
		N20	**N30**	**N40**	**N50**
TAG-B0F0	**MAP**	.364	.317	.286	.265
TAG-B0F1	**MAP**	.366	.318	.287	.267*

Note. Paired t-test was performed with the questions of F1 models.

Table 25 shows the MAP score of F0 and F1 models tested with the questions of F1 dataset. Using this result, we performed paired sample t-test. From the result, we observed that:

- There was no significant difference in MAP between TAG-B0F0N20 (M=0.364, SD=0.297) and TAG-B0F1N20 (M=0.366, SD=0.297) conditions; t(2079)=-1.132, p=0.258.
- There was no significant difference in MAP between TAG-B0F0N30 (M=0.317, SD=0.284) and TAG-B0F1N30 (M=0.318, SD=0.284) conditions; t(2346)=-1.023, p=0.306.
- There was no significant difference in MAP between TAG-B0F0N40 (M=0.286, SD=0.272) and TAG-B0F1N40 (M=0.287, SD=0.272) conditions; t(2519)=-1.265, p=0.206.
- There was a significant difference in MAP between TAG-B0F0N50 (M=0.265, SD=0.261) and TAG-B0F1N50 (M=0.267, SD=0.263) conditions; t(2647)=-2.283, p=0.023

Based on this result, we concluded that there is no significant difference in MAP between F0 and F1 models.

All in all, we observed that the TAG-B0F0N20 and TAG-B0F1N20 models performed best of all the models of different training datasets. Since we observed that there was no difference in terms of MAP between TAG-B0F0N20 and TAG-B0F1N20 models except for the quality of questions in the training dataset, we selected TAG-B0F0N20 for testing with the testing dataset.

5.2.4.2 Performance comparison between the selected model (TAG-B0F0N20) and baseline

In order to see if there was any improvement in the performance in estimation of relevant candidates to the given questions, we performed a paired-sample t-test between the best performing model (TAG-B0F0N20) and the baseline model that only used a single factor provided as default by the search engine. Test results showed that there was a significant difference in MAP between baseline (M=0.128, SD=0.241) and TAG-B0F0N20 (M=0.141, SD=0.252) conditions; t(5872)=-6.473, p=0.000.

Table 26. Performance comparison between the proposed model and baseline (training dataset)

Model \ Dataset		Train			
		N20	**N30**	**N40**	**N50**
TAG-B0F0	**MAP**	**.141***	.141*	.138*	.134*
Baseline	**MAP**	**.128**	.126	.123	.119

5.2.4.3 Testing the selected model with the testing dataset

Previously, we observed that the tag-based topic modeling approach with top 20 candidate profiles (TAG-B0F0N20) was the best method to collect dataset for the logistic regression analysis, in order to estimate relevant answerers to the given questions. In order to evaluate the selected model (TAG-B0F0N20) with the testing dataset, we performed a statistical t-test on the

MAP scores of the model. For the test, two set of questions were used; (1) not-precision-filtered questions (F0) and (2) precision-filtered questions (F1).

Table 27. Performance comparison between the proposed model and baseline (testing dataset)

Model \ Dataset		Test			
		N20	**N30**	**N40**	**N50**
TAG-B0F0	**MAP**	.132*	.132*	.130*	.128*
Baseline	**MAP**	.110	.109	.107	.105
TAG-B0F1	**MAP**	.362*	.315*	.279*	.255*
Baseline	**MAP**	.302	.258	.229	.209

Note: For testing TAG-B0F1 model, a set of question that had relevant answerers in the candidate pool was used for the evaluation.

- There was a significant difference in MAP between baseline (M = 0.110, SD = 0.227) and TAG-B0F0N20 (M = 0.132, SD = 0.252) conditions; t(2520) = -6.855, p = 0.000.
- There was a significant difference in MAP between baseline (M = 0.302, SD = 0.288) and TAG-B0F1N20 (M = 0.362, SD = 0.300) conditions; t(1000) = -7.537, p = 0.000.
- There was a significant difference in MAP between baseline (M = 0.109, SD = 0.219) and TAG-B0F0N30 (M = 0.132, SD = 0.243) conditions; t(2481) = -7.432, p = 0.000.
- There was a significant difference in MAP between baseline (M = 0.258, SD = 0.275) and TAG-B0F1N30 (M = 0.315, SD = 0.290) conditions; t(1148) = -8.235, p = 0.000.
- There was a significant difference in MAP between baseline (M = 0.107, SD = 0.212) and TAG-B0F0N40 (M = 0.130, SD = 0.234) conditions; t(2449) = -7.621, p = 0.000.
- There was a significant difference in MAP between baseline (M = 0.229, SD = 0.265) and TAG-B0F1N40 (M = 0.279, SD = 0.276) conditions; t(1248) = -8.109, p = 0.000.
- There was a significant difference in MAP between baseline (M = 0.105, SD = 0.208) and TAG-B0F0N50 (M = 0.128, SD = 0.228) conditions; t(2419) = -7.584, p = 0.000.
- There was significant difference in MAP between baseline (M = 0.209, SD = 0.256) and TAG-B0F1N50 (M = 0.255, SD = 0.266) conditions; t(1308) = -8.042, p = 0.000.

Evaluation with the testing dataset supported that the proposed model (TAG-B0F0N20) was valid to improve the performance of estimating relevant answerers to a given question. This result also suggested that if the baseline system can provide a good relevant candidate pool, the suggested logistic regression model performed better estimating relevant answerers among the given candidates; the absolute value of t-scores of F1 models were larger than that of F0 models, meaning that the more the baseline system provides relevant answerers in top 20 candidate pool, the more the suggested model estimates relevant candidates versus the baseline system.

5.2.5 Selecting the best fit logistic model

Before we propose the best fit logistic regression model for the automated QT for SR, we needed to investigate whether the proposed logistic regression model fits well to our dataset. In the previous chapter, we defined a testing logistic regression model that incorporates nine factors—eight factors from the attributes of users, with relevance scoring by the search engine result as the ninth factor (Equation 6).

In order to choose the best model, we performed two separate sets of analysis, one including all nine factors of interest (the full model) and the other without it (the reduced model) for each topic model approaches: tag-based approach (TAG) and topic-clustering approach (TC).

5.2.5.1 Test the model using the tag-based (TAG) approach

In order to test the full model using the tag-based approach, we performed a binary logistic regression analysis using the training dataset in which topic score was calculated using tags that are associated with the question and the user. Table 28 shows a summary of the result of the full

model. This result supports that topic of interest (x1), the scope of topic area (x3), contribution to answering (x4), the performance of providing good answers (x6), answer quota per day (x8), and Indri relevance score (x9) have an impact on estimating relevant answerers to the given question for automated QT. While the scope of subject (x2), contribution to submitting questions (x5), and response time to answer (x7) seemed to not have an impact on the estimation with condition; their coefficients $\neq 0$, but $P > .05$. In order to make sure whether those factors have an impact on the estimation, we tested a reduced model excluding those three factors.

5.2.5.1.1 Test result of the full logistic regression model

As discussed earlier, the full model is defined as the following:

$$\text{Predicted logit} = \beta_0+\beta_1 \cdot x_1+ \beta_2 \cdot x_2+\beta_3 \cdot x_3+\beta_4 \cdot x_4+\beta_5 \cdot x_5+\beta_6 \cdot x_6+\beta_7 \cdot x_7+\beta_8 \cdot x_8+\beta_9 \cdot x_9$$

$$= \beta_0+\beta_1 \cdot x_1+\beta_3 \cdot x_3+\beta_4 \cdot x_4+\beta_5 \cdot x_5+\beta_6 \cdot x_6+\beta_7 \cdot x_7+\beta_8 \cdot x_8+\beta_9 \cdot x_9$$

5.2.5.1.1.1 Level of topic of interest (x_1)

Earlier, we set the following hypothesis for the factor of topic of interest.

$H_{1\text{-}0}$: User's level of topic interest (x_1) will not affect to the performance of automated QT ($\beta_1 = 0$, $P < .05$)

$H_{1\text{-}A}$: User's level of topic interest (x_1) will affect to the performance of automated QT ($\beta_1 \neq 0$, $P < .05$)

Table 28. Logistic regression analysis of factors affecting to the estimation of relevant answerers (Dataset of TAG-B0F0N20).

Predictor	β	SE β	Wald's X^2	df	Sig.(p)	Exp(β)
Constant	-4.5084	.043	10744.348	1	.000	.021
Topic of interest (x_1)	.194	.015	158.364	1	.000	1.214
Scope of subject interest (x_2)	.044	.028	2.421	1	.120	1.045
Scope of topic interest (x_3)	-1.810	.078	536.797	1	.000	.164
Answering (x_4)	1.742	.100	300.299	1	.000	5.711
Questioning (x_5)	-.031	.019	2.778	1	.096	.969
Relevant answering (x_6)	.287	.038	57.301	1	.000	1.332
Response time (x_7)	.032	.020	2.688	1	.101	1.033
Quota (x_8)	-.222	.028	63.440	1	.000	.801
Indri relevance score (x_9)	.415	.023	316.986	1	.000	1.514

Note. Model summary: -2 Log likelihood = 29573.936

Based on the result above, we reject $H_{1\text{-}0}$ in favor of $H_{1\text{-}A}$ with β1 = 0.194, P = .000. Thus, the user's topic interest affects the performance of automated QT.

The factor of topic interest is ranked as the fourth most important factor. Since the training dataset (TAG-B0F0N20) was collected from the user profiles that use tag-based topic modeling approach, this factor actually means how many tags from the question are associated

with the user. The result shows that a user who had a tag-based topic score that was larger than the standard deviation of the score in the candidate profiles had 1.214 times more chance of being a relevant answerer to the given question compared to users who had tag-based topic score in its standard deviation.

5.2.5.1.1.2 The scope of subject area of interest (x_2)

Earlier, we set the following hypothesis for the factor of subject scope.

$H_{2\text{-}0}$: User's scope of subject area (x_2) will not affect the performance of automated QT ($\beta_2 = 0$, $P < .05$)

$H_{2\text{-}A}$: User's scope of subject area (x_2) will affect the performance of automated QT ($\beta_2 \neq 0$, $P < .05$)

The logistic regression analysis support that $H_{2\text{-}1}$ is rejected in favor of $H_{2\text{-}0}$ with condition; $\beta_2 = .044$, $P = .120$. Thus, the user's subject scope does not affect the performance of automated QT.

5.2.5.1.1.3 The scope of topic area of interest (x_3)

The following hypothesis was made to test the impact of the scope of topic area to estimate relevant answerers for automated QT.

$H_{3\text{-}0}$: User's scope of topic area (x_3) will not affect the performance of automated QT ($\beta_3 = 0$, $P < .05$)

$H_{3\text{-}A}$: User's scope of topic area (x_3) will affect the performance of automated QT ($\beta_3 \neq 0$, $P < .05$)

Based on the result, we reject $H_{3\text{-}0}$ in favor of $H_{3\text{-}A}$ with conditions; $\beta_3 \neq -1.810$, P= .000. Thus, we can conclude that the user's topic scope affects the performance of automated QT.

5.2.5.1.1.4 Contribution to answer providing (x_4)

In order to test the importance of contribution to answer providing, in order to estimate a relevant answerer to the question, the following hypotheses were defined.

$H_{4\text{-}0}$: User's contribution to providing answers (x_4) will not affect the performance of automated QT ($\beta_4 = 0$, $P < .05$)

$H_{4\text{-}A}$: User's contribution to providing answers (x_4) will affect the performance of automated QT ($\beta_4 \neq 0$, $P < .05$)

Based on the result of the logistic regression analysis, we reject $H_{4\text{-}0}$ in favor of $H_{4\text{-}A}$ with conditions; $\beta_4 \neq 1.742$, P= .000. Thus, it is concluded that the user's contribution to providing answers affects the performance of automated QT.

Answer activity is the most important predictor since the exponentiation of the coefficient (Exp(β)), called odd ratio, is the largest one. In this case an odds ratio of 5.71 indicates that a user who contributed to providing answers more than a standard deviation had 5.71 times the chance of being a relevant answerer to the given question compared to users who provided answers within a standard deviation.

5.2.5.1.1.5 Contribution to question submitting (x_5)

In order to test the importance of the contribution to submitting question to estimate relevant answerer to the question, the following hypotheses were defined.

$H_{5\text{-}0}$: User's contribution to submitting questions (x_5) will not affect the performance of automated QT ($\beta_5 = 0$, $P < .05$)

$H_{5\text{-}A}$: User's contribution to submitting questions (x_5) will affect the performance of automated QT ($\beta_5 \neq 0$, $P < .05$)

Based on the result above, we reject $H_{5\text{-}A}$ in favor of $H_{5\text{-}0}$ with conditions; $\beta_5 \neq -.031$, P= .096. Thus, we can conclude that user's contribution to submitting question do not affect to the performance of automated QT.

5.2.5.1.1.6 Comparing performance to providing relevant answers (x_6)

In order to test the importance of the performance to providing good answers, in order to estimate relevant answerer to the question for automated QT system, the following hypotheses were defined.

$H_{6\text{-}0}$: User's performance to providing good answers (x_6) will not affect the performance of automated QT ($\beta_6 = 0$, $P < .05$)

$H_{6\text{-}A}$: User's performance to providing good answers (x_6) will affect the performance of automated QT ($\beta_6 \neq 0$, $P < .05$)

Based on the result above, we reject $H_{6\text{-}0}$ in favor of $H_{6\text{-}A}$ with conditions; $\beta_6 \neq .287$, P= .000. Thus, we can conclude that user's performance to providing good answers affects the performance of automated QT.

5.2.5.1.1.7 Response time of answer providing (x_7)

In order to test the importance of the response time of answer providing, in order to estimate relevant answerer to the question for automated QT system, the following hypotheses were defined.

$H_{7\text{-}0}$: User's response time in providing answers (x_7) will not affect the performance of automated QT ($\beta_7 = 0$, $P < .05$)

$H_{7\text{-}A}$: User's response time in providing answers (x_7) will affect the performance of automated QT ($\beta_7 \neq 0$, $P < .05$)

Based on the result, we reject $H_{7\text{-}A}$ in favor of $H_{7\text{-}0}$ with conditions; $\beta_7 \neq .032$, $P=.101$. Thus, we can conclude that user's response time in providing answers do not affect the performance of automated QT.

5.2.5.1.1.8 Quota per day (x_8)

In order to test the importance of answer quota per day to estimate relevant answerers to the question for an automated QT system, the following hypotheses were defined.

$H_{8\text{-}0}$: The user's daily quota in providing answers (x_8) will not affect the performance of automated QT ($\beta_8 = 0$, $P < .05$)

$H_{8\text{-}A}$: The user's daily quota in providing answers (x_8) will affect to the performance of automated QT ($\beta_8 \neq 0$, $P < .05$)

Based on the result, we reject $H_{8\text{-}0}$ in favor of $H_{8\text{-}A}$ with conditions; $\beta_8 \neq .222$, $P= .000$. Thus, we can conclude that the user's daily quota in providing answers affect the performance of automated QT.

5.2.5.1.1.9 Indri relevance score (x_9)

The Indri relevance score is the second most important predictor for estimating relevant answerers to the question. In this study, this factor is considered as a default factor for the baseline (Equation 7).

5.2.5.1.2 Test result of the reduced model

Based on the result of the logistic regression analysis for the full model, we could reject $H_{2\text{-}A}$, $H_{5\text{-}A}$, $H_{7\text{-}A}$ in favor of $H_{2\text{-}0}$, $H_{5\text{-}0}$, $H_{7\text{-}0}$. In order to confirm this decision, we tested three reduced models as the following:

Reduced model 1: $\beta_2=0$

Predicted logit = $\beta_0+\beta_1\cdot x_1+ 0\cdot x_2+\beta_3\cdot x_3+\beta_4\cdot x_4+\beta_5\cdot x_5+\beta_6\cdot x_6+\beta_7\cdot x_7+\beta_8\cdot x_8+\beta_9\cdot x_9$

$= \beta_0+\beta_1\cdot x_1+ \beta_3\cdot x_3+\beta_4\cdot x_4+\beta_5\cdot x_5+\beta_6\cdot x_6+\beta_7\cdot x_7+\beta_8\cdot x_8+\beta_9\cdot x_9$

Reduced model 2: $\beta_5=0$

Predicted logit = $\beta_0+\beta_1\cdot x_1+ \beta_2\cdot x_2+\beta_3\cdot x_3+\beta_4\cdot x_4+0\cdot x_5+\beta_6\cdot x_6+\beta_7\cdot x_7+\beta_8\cdot x_8+\beta_9\cdot x_9$

$= \beta_0+\beta_1\cdot x_1+ \beta_2\cdot x_2+\beta_3\cdot x_3+\beta_4\cdot x_4+\beta_6\cdot x_6+\beta_7\cdot x_7+\beta_8\cdot x_8+\beta_9\cdot x_9$

Reduced model 3: $\beta_7=0$

Predicted logit = $\beta_0+\beta_1\cdot x_1+ \beta_2\cdot x_2+\beta_3\cdot x_3+\beta_4\cdot x_4+\beta_5\cdot x_5+\beta_6\cdot x_6+0\cdot x_7+\beta_8\cdot x_8+\beta_9\cdot x_9$

$= \beta_0+\beta_1\cdot x_1+ \beta_2\cdot x_2+\beta_3\cdot x_3+\beta_4\cdot x_4+\beta_5\cdot x_5+\beta_6\cdot x_6+ \beta_8\cdot x_8$

In order to test whether if there is significant difference between the full model and reduced models, a log likelihood ratio test was performed to compare each reduced model with the full model.

The log likelihood ratio test suggested that there is no significant difference between the full model and the reduced model 1 with the condition; LR chi2(1) = 2.53, P = .112.

The test also suggested that there is no significant difference between the full model and the reduced model 2 with the condition; LR chi2(1) = 2.77, P = .096.

The test also supported that there was no significant difference between the full model and the reduced model 3 with the condition; LR chi2(1) = 2.69, P=.101.

Thus, we can conclude that the scope of subject areas (*x2*), contribution to submitting questions (*x5*), and response time (*x7*) do not affect the performance of automated QT.

6 DISCUSSION AND IMPLICATIONS

6.1 INTRODUCTION

One of the goals of the current study was to introduce QT from the field of DR into the field of SR, hoping that QT might improve user satisfaction by increasing the success or speed of answering questions.

In the current study, the problem of assigning a question to appropriate answerers in the setting of SR was approached as a problem of finding experts, since the kernel of QT is to find experts who have ability to provide an answer to the question.

The major focus of the current study was on answerers who share their expertise with other users by providing answers to submitted questions. Their expertise can be considered an important source of information in SR services. In this study, the answerer's expertise was described as a combination of subject interest, performance, and contribution.

The current study was an exploratory investigation into factors that affect the identification of relevant answerers, for automated QT for a SR service. From the literature review on DR services, eight important attributes of answerers, sub-divided into nine factors, were identified and evaluated in order to route a submitted question to appropriate answerers who have the ability to answer the question assessing their expertise in the context of SR.

6.2 THE AUTOMATED QT PROCESS

The main focus of the current study was the automated QT. The process of QT can be done either by a human or an automated process in libraries. However, human-mediated QT seems to be impossible in the context of SR; a huge number of questions, for example, an average of 88,122 questions per day in Yahoo Answers, are posted every day to an online community.

One of the preliminary findings of the current study is the description of a framework of the automated QT process for SR, as shown in Figure 7. This framework provides an overview to researchers trying to build an automated QT system using archived users' question-answer pairs.

6.3 REPRESENTATION OF USER EXPERTISE

Another preliminary finding of the current study is the description of a framework for user expertise within SR (see, section 3.2.3). In the current study, user profiles representing user expertise were prepared and used in order to assign a question to appropriate answerers, after assessing their expertise from user profiles. In order to model user expertise, a framework of user expertise based on subject interest, performance and contribution, was proposed for SR services.

6.4 ATTRIBUTES THAT AFFECT THE TRIAGE PROCESS

One of the main findings of the current study is the report of factors that affect the performance of automated QT. In this study, attributes of the question and the answerers were identified and

evaluated to investigate important factors affecting the performance of the automated QT within SR.

6.4.1 Attributes of the question

In this study, subject, topic and key terms of the question were evaluated in order to route the question to appropriate answerers. In this study, the user-assigned category for the question was regarded as the subject area. In our dataset, the researcher observed that the average number of answerers for each category is larger than 7,000. For this reason, the concept of topic, as a sub-class of a subject area, was identified and evaluated in order to recommend appropriate answerers in the subject area. In order to identify the topic area of a question, two different approaches were used: tag-based approach and topic-clustering-based approach. In the current study, the researcher observed that the tag-based approach, in which a tag associated with the question is considered as a topic area, was more useful than the topic-clustering-based approach. There are two main reasons for this observation: (1) the poor performance of the topic-clustering tool and (2) the high-quality match between tags and question topic. The performance of the topic-clustering tool, the Stanford Topic Modeling Toolbox, may vary as the number of topic areas is changed. With a real dataset, it is almost impossible to define the best-performed number of topics for each subject area or category. Another explanation of this observation is that user assigned tags to questions represent the topic area of the question better. Indeed, most of the tags associated with the question were keyword terms appearing in the question. Another possible explanation is that explicitly expressed tags attract more answerers than more vague topic area postings, where the specific question is hidden.

6.4.2 Attributes of the answerer

The main focus of the current research was to investigate important attributes of the answerer as factors affecting the performance of automated QT for SR.

6.4.2.1 Subject and topic areas of interest

In DR settings, the area of subject expertise was considered as the second most important factor, following the subject area of the question. In the current logistic regression analysis, the subject area of interest was not implemented in the suggested user expertise model, but this factor was used previous to the estimation created by the proposed logistic regression model; in the current approach, we first selected candidates who showed interest on the subject area or topic area of the question, then applied the proposed logistic regression model to the selected candidates to select the most relevant answerer candidates.

This study seeks to assign user-submitted questions to appropriate answerers rather than let the user post their questions to a category in the community, thus more detailed or specific information than category information needs to be captured in order to reduce the number of candidates to a number smaller than the large pool of users who showed interested in that category. For this reason, the researcher preferred topic area of interest (sub-category) to subject area of interest (category).

6.4.2.2 Level of topic interest

In the estimation of relevant answerers to the question, the user's level of topic interest is one of the important factors in the decision. Since a language model search engine was used in order to generate a training dataset for the evaluation, an Indri relevant score was provided as the default for each candidate to a question. This score was categorized into the topic of interest since this score represents similarity between the question and the question-answer pairs of the user.

From the logistic regression analysis, the user's topic of interest showed positive tendency to performance for estimating relevant answerers to the question. The Indri relevance score showed a moderate tendency (β=.415) to estimate relevant answerers to the question. In the logistic regression model, the Indri relevance score was included as the ninth factor and this was regarded as a single factor that affected the performance of the baseline. This means that both the baseline and the proposed full logistic regression model are dependent upon the performance of the search engine. In this study, the researcher also observed that explicitly expressed tag information, assigned by the user, is more useful than implicitly expressed topic information generated by the topic-clustering approach, to estimate relevant answerers to the question. Tag-based topic information showed a poor correspondence to the estimation (β=.194). In the analysis, a weak correlation between tag-based topic of interest (x1) and the Indri relevance score (x9) was observed (r = -.385). An explanation of this observation is that both factors model the user's topic of interest in somewhat in different ways.

6.4.2.3 Scope of the subject/topic area

The researcher was interested in the scope of the subject area and the scope of the topic area. The scope of subject area is a measurement of the user's breadth of general interest, and the scope of topic area is a measurement of the user's breadth of specific interest within a narrower band of the subject area. In this study, the scope of subject area (x2) and the scope of topic area (x3) were included as factors to estimate relevant answerers to the question.

In this study, the researcher observed that the user's scope of subject area is not an important factor (β=.044, P>0.05), but the user's scope of topic area was identified as one important factor, when trying to estimate relevant answerers to the question given (β=-1.810, P=.000). The relation between the estimation and the scope of topic area showed a negative but strong tendency. This means that an answerer who was interested in a relatively specific topic areas (narrow scope of the subject area) had more chance to be a relevant answerer to the question.

6.4.2.4 Activities in answering and questioning

In this study, user activities in providing answers and submitting questions were interesting factors, computing user contribution for the estimation. In this study, the researcher observed that there is a strong positive relationship between answering activity and the estimation (β=1.742, P=.000) but no relationship between questioning activity and the estimation (β=-.031, P>.05). Thus, the user's contribution to answers provided was one of the important factors for estimating relevant answerers to the question, and answer activity is revealed as the most important factors among them. It showed the strongest impact on the estimation (Exp(β)=5.711). This means that

if an answerer is outstanding in terms of the number of answer provided to the community among the candidates recommended by the Indri relevance score, the answerer has 5.711 times more chance to be a relevant answerer than that of substandard answerers.

6.4.2.5 Performance in providing relevant answers

In the current study, we are also interested in the user's performance in providing good answers. From the analysis, a positive relationship between the user's performance and the estimation had a weak tendency (β=.287, P<.05, Exp(β)=1.322). This means that a user who had an outstanding number of good answers provided to the community had 1.322 times more chance to be a relevant answerer than other substandard answerers.

6.4.2.6 Response time

The factor of response time was also regarded as an important factor affecting the pathway of QT in DR. Indeed, sometimes a question is submitted to reference librarian with a time limit in DR. In this sense, the response time seems to be an important factor on decision making for QT in DR. However, this scenario is not applicable to the SR setting since there is no one who is responsible for meeting the deadline, as is the nature of online SR services.

In this study, the factor of response time was evaluated and it was revealed that there was no meaningful impact on the estimation (β=.032, P>.05).

6.4.2.7 Quota

The maximum number of questions that were answered by each answerer was considered as a factor of quota in order to investigate whether if there was any relationship between this factor and the estimation. From the logistic regression analysis, a poor negative relationship between them was observed (β=-.222, P=.000, Exp(β)=.801). This means that a user who answered an outstanding number of questions per day had 0.801 times less chance to be a relevant answerer than other answerers who were not. In other words, a relevant answerer provided answers to a smaller number of questions.

6.4.3 Summary

From the logistic regression analysis of nine factors from eight attributes of the answerer, the researcher reports that:

1. Among the nine factors expected to affect the decision of automated QT, the user's topic of interest (x1 and x9), scope of topic area (x3), answering activity (x4), performance in providing good answers (x6), and quota (x8) were observed as important factors that should affect automated QT decisions within SR.
2. Among them, the user's answering activity was identified as the most important factor to affect the automated QT decisions within SR, followed by topic of interest (the Indri relevant score and tag-based topic of interest).
3. In terms of user contribution, the user's answering activity was a more important factor than their questioning activity, in the automated QT decisions within SR.

4. Subject scope, questioning activity and response time were not important factors affecting automated QT decisions within SR.
5. User assigned tag information was good evidence of the user's topic of interest.

Area of subject expertise of the answerer, which is considered as the third most important factor affecting QT decisions within DR, was not included in the proposed user expertise model for SR. In fact, it was used to select appropriate answerers based on the answerer's subject area of interest previous to the estimation. Thus, this factor still has the strong impact on the triage decision.

Although quota showed small impact on the estimation, it can be used for secondary decisions. For example, if the quota of the answerer has been reached, the triager did not forward the question to the answerer.

6.5 IMPLICATIONS FOR RESEARCH DESIGN

6.5.1 Theoretical Implications

In these days, we are facing the advent of new forms of user-participated QA, referred to as SR in this study. This study suggested expanding the use of QT, which has been well-developed in the field of DR, to SR in order to increase the quality of these services, by assigning the question to appropriate answerers in the community instead of letting it be done randomly.

The current study provides a framework of user expertise for SR. This framework provides guidelines to model or assess user expertise for SR. In order to route a question to a relevant answerer, a user profile representing user expertise is required to be assessed by the

triager. In the current framework, the user's subject interest, contribution and performance were modeled. This framework can be expanded as we identify meaningful factors affecting the decision of automated QT for SR.

This approach to model user expertise for SR can be also employed for user-participated reference in the context of DR. With the rise of library 2.0, many libraries try to encourage user's participation in their online library services. In the future we also can use user expertise in order to provide answers to the question submitted by the user (although there might be some limitations). To do that, we need to model the library users' expertise and to investigate important factors affecting the decisions of QT.

6.5.2 Practical implications

One of the main focuses of this study is automated QT. In the field of library science, many librarians and researchers proclaimed the need for research on automated QT. This study investigated important factors affecting the decision of automated QT for SR by assessing user expertise. The research method employed in this study enabled the researcher to investigate the weight of each factor for the estimation of relevant answerers to the question given. This method can be applied to investigate other important factors to develop a better model in the future.

7 CONCLUSION

7.1 REFLECTIONS ON ACCOMPLISHING THE RESEARCH GOALS

7.1.1 RQ1. What are the elements of automated QT for SR, and are they different from that of DR?

This question was raised in order to investigate key elements of automated QT and to explore new factors and requirements for automated QT in the context of SR services. In order to answer this question, the process of QA by DRs was explored. From the exploration, it was found that key elements involved in the process of QA in both DR and SR were identical. Because of the similarity between them, it was possible to adopt findings in the field of DR to SR. Four core elements of the QA process—question, answer, questioner, and answerer—were also observed in the QA process of SR. The element 'triager' observed in DR was missing in current online SR services. When the operation of QT was modeled in the QA process in SR, the flow of the question and the answer are almost identical between DR and SR.

The practice of QT has been well developed in the field of DR; the generic QT model could be used to develop QT function for SR. Researchers in the field found important factors affecting the QT decisions based on the attributes of the question, the answerer, and the service.

However, the attributes of the service could not be used for QT for SR since any services or institutions providing answers were not available in the context of SR.

Before investigating important factors affecting to the performance of automated QT for SR, we explored the fifteen most important factors of QT in DR settings, but they could not all be directly implemented for QT within SR since many of those factors were attributes of only the library-based services. However, they could be successfully modeled as attributes of the answerer in the SR settings by redefining them.

As the result, eight attributes of the answerer and one attribute of the question were chosen and evaluated in order to find important factors among them.

7.1.2 RQ2. What are the factors that affect the performance of automated QT for SR?

This question was asked in order to investigate important factors affecting QT decisions in the context of existing SR in order to develop guidelines for automated QT.

In the current study, it was found that the answerer's level of topic interest, contribution to providing answers, the scope of topic area, performance in providing good answers and quota were identified as important factors. The answerer's contribution to providing answers to the community, followed by user's topic of interest, was the most important factor affecting the performance of automated QT for SR. Although the user's topic of interest was regarded as the most important single factor affecting the decision of QT in the context of DR, it was ranked as the second important factor among the factors selected. This seems to be caused by the nature of SR. In the context of SR, an answer is provided by a user in the community, and answerers are not explicitly distinguished among users; a user's role, such as questioner or answerer, is

constantly changing as a question is given. In this sense, the factor of user answering activity was seemed to be boosted to distinguish answerers from questioners.

On the other hand, the answerer's scope of subject area, contribution to submitting questions and response time were discovered to not be important factors affecting the performance of automated QT for SR. Although these factors were not identified as important influences on automated QT decisions for SR, some of them still can be used as secondary factors for the decision. For example, the maximum number of answers provided per day by the user can be used to make a decision to not assign the question to the recommended answerer if the number of questions forwarded to the answerer has reached the quota.

7.1.3 Summary

RQ1 and RQ2 were answered through this study. In the perspective of question answer process, DR and SR are very similar except that QT is missing and the information source is represented implicitly in SR. Once QT is implemented in SR, the process of QA between DR and SR will become almost identical. In future research, we can count the user's social bookmarks or social citations as information sources or collections held by the user. From this viewpoint, the framework of SR is identical to that of DR. Then, we can introduce the findings of automated QT for SR to the field of digital libraries in order to develop user-participated reference services by utilizing the library user's expertise.

7.2 LIMITATIONS OF THE STUDY

The main focus of this study was to investigate important factors affecting the performance of automated QT for SR. In the current study, the researcher evaluated only nine factors that were captured from eight attributes of the answerer. However, there are more factors that can be captured from the attributes of the question, the questioner, and the answerer, such as type of question, questioner's topic of interest, user's geographic location, etc.

For training the suggested model for assessing user expertise, the training dataset was generated from a collection of five categories among twenty categories. For this reason, this study may not represent the characteristics of the main collection. Thus, the result of the evaluation cannot be generalized to represent the characteristics of the main collection.

Another limitation of this study is that the examined factors were selected from literature review on DR; they were not captured from real users who provide answers. In this sense, this study is a data-focused study rather than a subject-focused study. Since SR service users may have their own criteria to choose which question to answer, this study has the limitation of not representing all existing factors that affect a real users' decision-making process for selection questions to answer.

7.3 REMAINING QUESTIONS

QT is not about assigning a question to an answerer in itself; rather it also includes the filtering of repeated questions, out-of-scope questions, or unanswerable questions. In order to develop an automated QT system, a question-filtering module needs to be developed and implemented in

order to augment the performance of the system. In the current study, it was observed that questions which have better precision in generating more of the first top N selected candidates resulted in better performance in estimating relevant answerers to the given question. Thus, if the system can estimate an answerable question by examining the quality of the first result of top N candidates returned by the search engine (baseline), it would be possible to increase the performance of the automated QT system.

Some researchers focused on the type of questions in order to provide answers or answerers to the question. Another remaining question is whether there any relationship between question type and the answerer in the process of QT.

7.4 FUTURE WORK

7.4.1 Selection of identifying additional factors affecting QT for SR

In this study, the researcher borrowed factors affecting the decision-making of QT from DR and tested them for QT in the context of SR. For this reason, the factors tested in this study mainly originated from DR so that has some limitation in its applicability to SR. In terms of SR, it is necessary to investigate other factors that are inherent from SR, such as social relation among users.

In utilizing the decision-making factors of QT in DR, question type and language were not tested, in order to accomplish this study in a limited time. Furthermore, we could select and test additional factors from the attribute of the questioner, such as the questioner's topic of interest. In the future, these factors need to be considered and tested for SR.

7.4.2 Question recommendation for SR sites

The initial motivation of this study was to route a question to the relevant answerer candidates for a SR service. The research methods employed in this study could also be adopted by SR services, such as Ask Metafilter or Yahoo Answers, for question recommendation, as a type of QT, to help user to locate interesting questions. Once a question recommendation service is provided, we can test more factors from additional attributes captured from the new question recommendation service, for example, response rate.

7.4.3 Study on automated question filtering

As discussed earlier, question filtering is also an important part of QT. In order to develop fully working automated QT system, an automated question filtering module needs to be developed. The goal of this work is to estimate which user submitted questions are answerable by examining not only the quality and subject of the question but also the answerers' expertise. Adopting the current research design, it is possible to investigate important factors to use when filtering out unanswerable questions among the submitted questions to an online community.

7.4.4 Study on user-participated library reference (SR)

The success of social services, such as Yahoo Answer or Ask Metafilter, demonstrates the users' willingness to participate in the services. Future libraries would be expected to encourage users to participate in their library services. Library reference service is not an exception to the library

trend of user-participated library services. Expanding human information sources from only libraries to the large population of library users would be a highly useful resource in the future libraries. Further study on user-participated reference services in digital library settings needs to be done in order to develop a user-participated library that utilize these library users' expertise in the future.

BIBLIOGRAPHY

Adamic, Lada A, Zhang, Jun, Bakshy, Eytan, & Ackerman, Mark S. (2008). *Knowledge sharing and yahoo answers: everyone knows something.* Paper presented at the Proceedings of the 17th international conference on World Wide Web.

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). *Finding high-quality content in social media*.

Amitay, E., Carmel, D., Golbandi, N., HaríEl, N., Ofek-Koifman, S., & Yogev, S. *Finding people and documents, using web 2.0 data*.

Bailey, P., Craswell, N., de Vries, A.P., & Soboroff, I. (2007). *Overview of the TREC 2007 enterprise track.*

Balog, K., Azzopardi, L., & De Rijke, M. (2006). *Formal models for expert finding in enterprise corpora*.

Balog, K., Azzopardi, L., & de Rijke, M. (2009). A language modeling framework for expert finding. *Information Processing & Management, 45*(1), 1-19.

Balog, K., Thomas, P., Craswell, N., Soboroff, I., Bailey, P., De Vries, A.P., & ., AMSTERDAM UNIV. (2008). *Overview of the TREC 2008 enterprise track*: Citeseer.

Blei, David M, Ng, Andrew Y, & Jordan, Michael I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research, 3*, 993-1022.

Bohr, Niels. (1995). Expert Systems Defined. In J. V. R. Jr. (Ed.), *Knowledge-Based Systems for General Reference Work: Applications, Problems, and Progress*. New York: Academic Press.

Borgatti, S.P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management science*, 432-445.

Bouguessa, Mohamed, Dumoulin, Benoit, & Wang, Shengrui. (2008). *Identifying authoritative actors in question-answering forums: the case of Yahoo! answers*. Paper presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, Nevada, USA.

Boyd, K.M. (1983). Triage and Justice. *Journal of Medical Ethics, 9*(2), 117.

Bunge, Charles A. (1999). Reference Services. *The Reference Librarian, 31*(66), 185-199. doi: 10.1300/J120v31n66_17

Burns, R.B., & Burns, R.A. (2008). *Business research methods and statistics using SPSS*: SAGE Publications Ltd.

Chen, H., Shen, H., Xiong, J., Tan, S., & Cheng, X. (2006). *Social network structure behind the mailing lists: Ict-iiis at trec 2006 expert finding track*.

Coats, P.K. (1988). Why expert systems fail. *Financial Management, 17*(3), 77-86.

Craswell, N., de Vries, A.P., & Soboroff, I. (2005). *Overview of the trec-2005 enterprise track*.

Ericsson, K. Anders. (2007). Expertise. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology*. Thousand Oaks, CA: SAGE Publications, Inc.

Experience. (n.d.). *Merriam-Webster*. Retrieved April 25, 2013, from http://www.merriam-webster.com/dictionary/experience

Fang, H., & Zhai, C.X. (2007). Probabilistic models for expert finding. *Advances in Information Retrieval*, 418-430.

Farzan, R., & Brusilovsky, P. (2010). Encouraging user participation in a course recommender system: An impact on user behavior. *Computers in Human Behavior*.

Flammer, August. (1981). Towards a theory of question asking. *Psychological Research, 43*(4), 407-420.

Gazan, R. (2007). Seekers, sloths and social reference: Homework questions submitted to a question-answering community. *New Review of Hypermedia and Multimedia, 13*(2), 239-248.

Goldman, Alvin I. (1999). *Knowledge in a social world*. Oxford; New York: Clarendon Press ; Oxford University Press.

Goodall, A. (1985). *The guide to expert systems*: Information Today Inc.

Green, S.S. (1876). Personal relations between librarians and readers. *Library Journal, 1*(2), 3.

Harper, F. Maxwell, Moy, Daniel, & Konstan, Joseph A. (2009). *Facts or friends?: distinguishing informational and conversational questions in social Q&A sites.* Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems.

Harper, F. Maxwell, Raban, Daphne, Rafaeli, Sheizaf, & Konstan, Joseph A. (2008). *Predictors of answer quality in online Q&A sites*.

Hofmann, K., Balog, K., Bogers, T., & de Rijke, M. (2010). Contextual factors for finding similar experts. *Journal of the American Society for Information Science and Technology, 61*(5), 994-1014.

Horne, Esther E. (1983). Question Generation and Formulation: An Indication of Information Need*. *Journal of the American Society for Information Science, 34*(1), 5-15.

Inglis, Jennifer. (2003). The Cooperative Librarian. In K. Bridges (Ed.), *Expectations of Librarians in the 21st Century*. Westport, Connecticut: Greenwood Press.

Jeon, Jiwoon, Croft, W Bruce, & Lee, Joon Ho. (2005). *Finding similar questions in large question and answer archives.* Paper presented at the Proceedings of the 14th ACM international conference on Information and knowledge management.

Jiang, J., Han, S., & Lu, W. *Expertise retrieval using search engine results*.

Jurczyk, Pawel, & Agichtein, Eugene. (2007). *Discovering authorities in question answer communities by using link analysis*. Paper presented at the Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, Lisbon, Portugal.

Kamel Boulos, Maged N., & Wheeler, Steve. (2007). The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education1. *Health Information & Libraries Journal, 24*(1), 2-23. doi: 10.1111/j.1471-1842.2007.00701.x

Kao, Wei-Chen, Liu, Duen-Ren, & Wang, Shiu-Wen. *Expert finding in question-answering websites: a novel hybrid approach*. Paper presented at the Proceedings of the 2010 ACM Symposium on Applied Computing, Sierre, Switzerland.

Karimzadehgan, M., White, R., & Richardson, M. (2009). Enhancing expert finding using organizational hierarchies. *Advances in Information Retrieval*, 177-188.

Kim, Soojung, Oh, Jung Sun, & Oh, Sanghee. (2007). Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective. *Proceedings of the American Society for Information Science and Technology, 44*(1), 1-15.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political analysis, 9*(2), 137-163.

Knowledge. (n.d.). *Merriam-Webster.* Retrieved April 25, 2013, from http://www.merriam-webster.com/dictionary/knowledge

Lankes, R.D. (1998). *Building and Maintaining Internet Information Services: K-12 Digital Reference Services*: Information Resources Publications, Syracuse University, 4-194 Center for Science and Technology, Syracuse, NY 13244-4100; Tel: 800-464-9107 (Toll-Free).

Lankes, R.D. (2004). The digital reference research agenda. *Journal of the American Society for Information Science and Technology, 55*(4), 301-311.

Lee, Chong Tong, Rodrigues, Eduarda Mendes, Kazai, Gabriella, Milic-Frayling, Nataša, & Ignjatovic, Aleksandar. (2009). *Model for voter scoring and best answer selection in community Q&A services.* Paper presented at the Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on.

Li, Baichuan, & King, Irwin. (2010). *Routing questions to appropriate answerers in community question answering services.* Paper presented at the Proceedings of the 19th ACM international conference on Information and knowledge management.

Liu, Mingrong, Liu, Yicen, & Yang, Qing. (2010). Predicting best answerers for new questions in community question answering *Web-Age Information Management* (pp. 127-138): Springer.

Liu, X., Croft, W.B., & Koll, M. (2005). *Finding experts in community-based question-answering services*.

Liu, Yandong, Bian, Jiang, & Agichtein, Eugene. (2008). *Predicting information seeker satisfaction in community question answering.* Paper presented at the Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval.

Low, K. (1996). *The roles of reference librarians: Today and tomorrow*: Routledge.

Markgren, Susanne. (2006). Digital Reference Services Bibliography. *Public Services Quarterly, 2*(2/3), 194-195.

McDonald, D.W., & Ackerman, M.S. (1998). *Just talk to me: a field study of expertise location*.

Mitchell, G.W. (2008). A brief history of triage. *Disaster Medicine and Public Health Preparedness, 2*(Supplement 1), S4.

Nam, Kevin K., Ackerman, Mark S., & Adamic, Lada A. (2009). *Questions in, knowledge in?: a study of naver's question answering community*.

Naughton, J. (2001). *A brief history of the future: from radio days to Internet years in a lifetime*: Overlook Press, The.

Naylor, C. (1987). *Build your own expert system*: Halsted Press New York, NY, USA.

Oh, Sanghee. (2010). *Answerers' motivations and strategies for providing information and social support in social Q&A: an investigation of health question answering.* (Ph.D. dissertation), University of North Carolina.

Petkova, D., & Croft, W.B. (2006). Hierarchical language models for expert finding in enterprise corpora.

Pomerantz, J., Nicholson, S., Belanger, Y., & David Lankes, R. (2004). The current state of digital reference: validation of a general digital reference model through a survey of digital reference services. *Information Processing & Management, 40*(2), 347-363.

Pomerantz, J., & Stutzman, F. (2006). Collaborative reference work in the blogosphere. *Reference services review, 34*(2), 200-212.

Pomerantz, Jeffrey. (2003). *Question Taxonomies for Digital Reference.* (Ph.D. dissertation), Syracuse University, New York. ProQuest database.

Pomerantz, Jeffrey. (2004). Factors Influencing Digital Reference Triage: A Think-Aloud Study. *Library Quarterly, 74*(3), 235-264.

Pomerantz, Jeffrey, Nicholson, Scott, & Lankes, R. David. (2003). Digital Reference Triage: Factors Influencing Question Routing and Assignment^1. *Library Quarterly, 73*(2), 103.

Qu, Mingcheng, Qiu, Guang, He, Xiaofei, Zhang, Cheng, Wu, Hao, Bu, Jiajun, & Chen, Chun. (2009). *Probabilistic question recommendation for question answering communities.* Paper presented at the Proceedings of the 18th international conference on World wide web.

Richardson, J.V. (1995). *Knowledge-based systems for general reference work: Applications, problems, and progress*: Academic Press, Inc.

Rothstein, S. (1953). The development of the concept of reference service in American libraries, 1850-1900. *The Library Quarterly*, 1-15.

RUSA. (2008). Definitions of Reference. from http://www.ala.org/ala/mgrps/divs/rusa/resources/guidelines/definitionsreference.cfm

Salih, S. (2007). On Judith Butler and Performativity. *Sexualities and communication in everyday life: A reader*, 55-68.

Serdyukov, P., & Hiemstra, D. (2008a). *Being omnipresent to be almighty: The importance of the global web evidence for organizational expert finding*.

Serdyukov, P., & Hiemstra, D. (2008b). Modeling documents as mixtures of persons for expert finding. *Advances in Information Retrieval*, 309-320.

Serdyukov, P., Rode, H., & Hiemstra, D. (2008). *Modeling multi-step relevance propagation for expert finding*.

Shachaf, P. (2010). Social reference: Toward a unifying theory. *Library & Information Science Research, 32*(1), 66-76.

Shachaf, P., & Rosenbaum, H. (2009). Online social reference: A research agenda through a STIN framework. *iConference, 9*, 8-11.

Shah, C., Oh, J.S., & Oh, S. (2008). Exploring characteristics and effects of user participation in online social Q&A sites. *First Monday, 13*(9).

Shah, C., Oh, S., & Oh, J.S. (2009). Research agenda for social Q&A. *Library & Information Science Research, 31*(4), 205-209.

Shah, Chirag. (2011). Measuring effectiveness and user satisfaction in Yahoo! Answers. *First Monday, 16*(2-7).

Shah, Chirag, & Pomerantz, Jefferey. (2010). *Evaluating and predicting answer quality in community QA.* Paper presented at the Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.

Shami, N.S., Ehrlich, K., & Millen, D.R. (2008). *Pick me!: link selection in expertise search results*.

Shera, Jesse. (1964). Automation and the reference librarian. *RQ, 3*(6), 3-7.

Skill. (n.d.). *Merriam-Webster.* Retrieved April 25, 2013, from http://www.merriam-webster.com/dictionary/skill

Soboroff, I., de Vries, A.P., & Craswell, N. (2006). *Overview of the trec 2006 enterprise track.*

Straw, J.E. (2001). From Magicians to Teachers: The Development of Electronic Reference in Libraries: 1930-2000. *Reference Librarian, 74*, 1-12.

Subject. (n.d.). *Dictionary.com.* Retrieved April 25, 2013, from http://dictionary.reference.com/browse/subject

Terveen, L., & McDonald, D.W. (2005). Social matching: A framework and research agenda. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(3), 401-434.

Triage. (n.d.). *Oxford English Dictionary.* Retrieved April 25, 2013, from http://www.oed.com

Umbert, Pere R. Comas i. (2012). *Factoid question answering for spoken documents.* (Doctoral dissertation), Universitat Politècnica de Catalunya, Barcelona, Spain.

Voorhees, E.M., & Tice, D.M. (2000). *Building a question answering test collection*.

Waterman, D. (1986). A guide to expert systems.

Weil, C.B. (1968). Automatic retrieval of biographical reference books. *Journal of Library Automation, 1*(4), 239-249.

Woudstra, L., & van den Hooff, B. (2008). Inside the source selection process: Selection criteria for human information sources. *Information Processing & Management, 44*(3), 1267-1278.

Yahoo! Answers Team. (2010). Yahoo! Answers Blog. Retrieved from http://yanswersblog.com/index.php/archives/2010/05/03/1-billion-answers-served/

Zhang, J., Tang, J., & Li, J. (2010). Expert finding in a social network. *Advances in Databases: Concepts, Systems and Applications*, 1066-1069.