Investigating Factors Affecting Automated Question Triage for Social Reference: A Study of Adopting Decision Factors from Digital Reference*

Jong Do Park (박종도)**

Contents

ABSTRACT

The efficiency and quality of the social reference sites are being challenged because a large quantity of the questions have not been answered or satisfied for quite a long time. Main goal of this study is to investigate important factors that affect the performance of question triage to relevant answerers in the context of social reference. To achieve the goal, expert finding techniques were used to construct an automated question triage approach to resolve this problem. Furthermore, important factors affecting triage decisions in digital reference were first examined, and extended them to the social reference setting by investigating important factors affecting the performance of automated question triage in the social reference setting. The study was conducted using question-answer pairs collected from Ask Metafilter. For the evaluation, logistic regression analyses were conducted to examine which factors would significantly affect the performance of predicting relevant answerers to questions. The results of the current study have important implications for research and practice in automated question triage for social reference. Furthermore, the results will offer insights into designing user-participatory digital reference systems.

Keywords: Question Triage, Question Routing, Social Reference, Digital Reference

* This article was originally published as a Ph.D dissertation at the University of Pittsburgh (Park 2013).

** Instructor, Department of Library and Information Science at Chung Ang University (dlibrary@gmail.com)

논문접수일자: 2015년 1월 26일 최초심사일자: 2015년 2월 6일 게재확정일자: 2015년 2월 17일
한국문헌정보학회지, 49(1): 483-511, 2015. 〔http://dx.doi.org/10.4275/KSLIS.2015.49.1.483〕

www.kci.go.kr

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.

Question triage is the assignment of a question to a reference or subject expert answerer (Pomerantz, Nicholson and Lankes 2003). In the field of libraries, reference librarians have practiced question triage, 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 question triage 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 question triage 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.

2. Literature Review

The term "social reference (Gazan 2007; Shachaf 2010)" is used in the study to refer to online question answering 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

www.kci.go.kr

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 and Konstan 2009; Kim, Oh and Oh 2007)", "*community Q&A* (Lee et al. 2009; Li and King 2010; Nam, Ackerman and Adamic 2009; Shah and Pomerantz 2010)" or "*community-based QA* (Jeon, Croft and Lee 2005; Shah and Pomerantz 2010)".

Since research on Q&A sites has mainly focused on information retrieval (Agichtein et al. 2008; Bian et al. 2008; Jeon, Croft and Lee 2005) and information seeking behavior (Gazan 2007; Shah, Oh and Oh 2008), a few researchers have paid attention on Q&A sites in the context of reference service. Shachaf (2010) argued that social reference is similar to library reference, but at the same time, it may be significantly different from the traditional (and digital) dyadic reference encounter since it involves a collaborative group effort and uses Web 2.0 infrastructure. Furthermore, she proposed the Input-Process-Output (IPO) model for social reference to provide a clear account of the various components that shape social reference. One of the implications of her study is that she attempted to provide theoretical framework for social reference in the context of reference service.

While theoretical research on social reference is premature, that of digital reference is mature. Since reference service has been one of core services in library services, many researchers and librarians developed theories and practices for reference service and those are extended in the context of digital reference. Pomerantz, Nicholson & Lankes (2003) proposed a general digital reference model that consists of five steps-question acquisition, triage, answer formulation, tracking and resource creation- to illustrate the process of question answering in the context of digital reference. The implication of this study is in that they included question triage as a step of the process. However, it has limitations when used to illustrate the process of question triage in the context of social reference; 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 social reference. This model can be a starting point to develop a framework for SR for this research.

www.kci.go.kr

〈Table 1〉 Fifteen factors affecting question triage for DR (Pomerantz, Nichoson and Lankes 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.

Researchers in the field have also interested in the factors that affect the decision of question triage. Pomerantz (2004) found that question subject is the single most important attribute that affects triage decision among fifteen factors determined by Pomerantz, Nichoson & Lankes (2003).

In the field of information retrieval, many researchers have focused on identifying best answerers. In approaching this, researchers use different terms to refer to the best answerers (Bouguessa, Dumoulin and Wang 2008; Qu et al. 2009; Liu, Liu and Yang 2010), such as experts (Zhang, Tang and Li 2010), authoritative users (Bouguessa, Dumoulin and Wang 2008), etc. Some researchers employed link analysis techniques, such as Page Rank and HITS algorithms to identify authoritative answerers in a social reference site (Jurczyk and Agichtein 2007; Agichtein et al. 2008). Bouguessa, Dumoulin & Wang (2008) proposed a model to identify authoritative actors based on the number of best answers provided by them. Zhang, Tang & Li (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. 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.

www.kci.go.kr

3. Methodology

The goal of this research is to describe and gain a further understanding of automated question triage in the context of social reference. In order to achieve the goal of this study, the following research question is raised:

"What are the factors that affect the performance of automated question triage for social reference?"

3.1 Selecting Factors for Testing

In order to answer the research question, the researcher tried to test triage factors from digital reference in the context of social reference. The literature provides a comprehensive investigation of the identifying attributes of key elements as factors that affect the decisions of question triage. Pomerantz, Nicholson & Lankes (2003) determined the fifteen most important factors that affect a triager's decisions in the digital reference setting, based on a survey of experts about the decision-making process in digital reference. In the current study, those fifteen factors, determined by them were adopted and redefined in the context of social reference. Those factors are, in fact, the chosen attributes of question, answerer and service. For this study, the chosen attributes of question, answerer, and service of digital reference are mapped and redefined in the context of social reference.

〈Table 2〉 Mapping the triage factors of Digital Reference to Social Reference

Element	Attribute		Factor	Applicability
	Digital Reference	Social Reference		
Question (Q)	Subject area	Subject area		Category posted
		Topic area		Tag assigned by the questioner
Answerer (A)	Area of subject expertise	Subject area of interest		Category in which answers were provided previously
		Topic area of interest	x_1	Tag associated with questions to which the user answered
	Scope of the subject expertise[a)]	Scope of subject interest	x_2	Generality (subject)
		Scope of topic interest	x_3	Specialty (topic)
	Experience & skill	Answering activity	x_4	General experience in answering questions
		Questioning activity	x_5	General experience in posting questions
	Past performance	Performance	x_6	Providing good answers
	Turnaround time	Response time	x_7	Count the response time
	Number of question per unit for times	Quota	x_8	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.

www.kci.go.kr

The question's subject was considered as the single most important factor affecting decision-making in question triage in digital reference (Pomerantz 2004). In order to make the decision to assign a question to relevant answerers, it is needed to identify the primary matter of the discussion (subject) of the question. For this study, the researcher 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.

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 that represents the answerer's general interest. In the current study, the answerer's subject area and topic area are regarded as attributes of the answerer. The conceptual difference between subject and topic in terms of the boundary of the answerer's interest is explained in <Table 3>.

〈Table 3〉 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

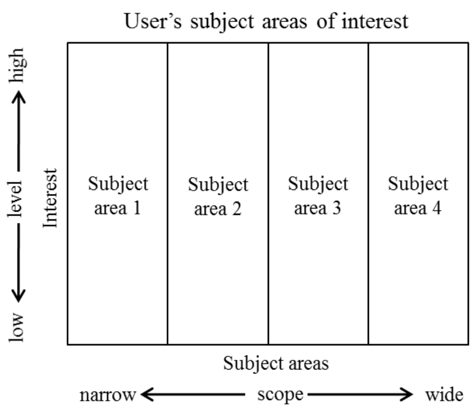
In order to assess the answerer's subject area of interest using the main collection of this study, the researcher examined the answerer's previous category preference in providing answers, and interpreted the answerer's interest on categories as evidence of the answerer's subject interest.

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

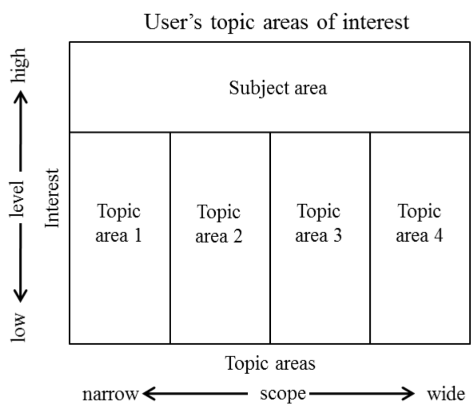
www.kci.go.kr

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

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



〈Figure 1〉 Map of user's subject areas of interest



〈Figure 2〉 Map of user's topic areas of interest

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. Furthermore, some answerers provide more answers to the community than others are doing. Considering this kind of difference would be useful in making decisions for question triage.

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.

Instead of using the term 'turnaround time', the researcher uses 'response time' in the context of SR. In the current SR setting, we can infer the answerer's response time to questions that

www.kci.go.kr

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.

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 question triage; 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 A framework of expertise: subject interest, performance, and contribution

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

As an expert finding task, finding relevant experts or answerers to the question is the key to success in question triage. 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. In this research, a framework consisting of subject interest, performance and contribution is used to address user expertise in the context of social reference.

Employing the above framework of expertise, the factors affecting to the decision of question triager can be mapped into three dimensions as the aspects of expertise (see, Table 4).

〈Table 4〉 The framework of user expertise for social reference

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, it is needed first to know who

www.kci.go.kr

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 question triage by other researchers (Pomerantz et al. 2004). In fact matching subject matter between the question and the answer is the key to success in question triage, 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, it is also needed 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.

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.

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.

3.3 Hypotheses

In order to investigate important factors affecting the performance of automated question triage for social reference, nine attributes of the answer were chosen and null hypothesis and alternative hypotheses were defined in order to test the regression coefficients of those factors (see, Table 5). It is expected that a statistically important factor would increase the performance of the automated question triage algorithm.

For each factor, a null hypothesis ($H_{x\text{-}0}$) and an alternative hypothesis ($H_{x\text{-}A}$) were defined. $H_{x\text{-}0}$ represents that the factor will not affect to the performance of automated question triage ($\beta_x = 0$, $P < .05$); and $H_{x\text{-}A}$ represents that the factor will affect to the performance of automated question triage ($\beta_x \neq 0$, $P < .05$). For example, $H_{1\text{-}0}$ means that the user’s level of topic interest (x_1) will not affect to the performance of automated question triage ($\beta_1 = 0$, $P < .05$). Likewise, $H_{1\text{-}A}$ means that the user’s level of topic interest (x_1) will affect to the performance of automated question

www.kci.go.kr

triage ($\beta_1 \neq 0$, $P < .05$).

〈Table 5〉 The Hypotheses of the study

Factor (IV)	Null Hypothesis ($H_{x\text{-}0}$)	Alternative Hypothesis($H_{x\text{-}A}$)
Level of topic of interest (x_1)	$H_{1\text{-}0}$	$H_{1\text{-}A}$
Scope of subject interest (x_2)	$H_{2\text{-}0}$	$H_{2\text{-}A}$
Scope of topic interest (x_3)	$H_{3\text{-}0}$	$H_{3\text{-}A}$
Answering activity (x_4)	$H_{4\text{-}0}$	$H_{4\text{-}A}$
Questioning activity (x_5)	$H_{5\text{-}0}$	$H_{5\text{-}A}$
Relevant answering (x_6)	$H_{6\text{-}0}$	$H_{6\text{-}A}$
Response time (x_7)	$H_{7\text{-}0}$	$H_{7\text{-}A}$
Quota (x_8)	$H_{8\text{-}0}$	$H_{8\text{-}A}$
Indri relevance score (x_9)[a)]	$H_{9\text{-}0}$	$H_{9\text{-}A}$

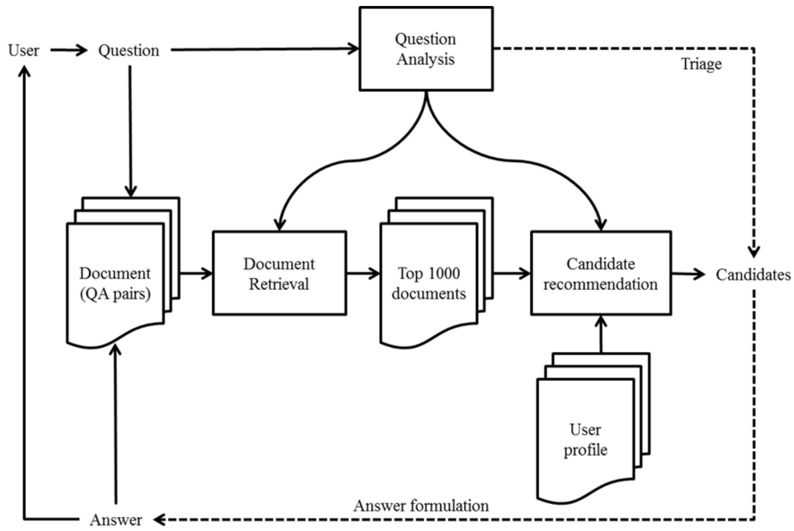
Note. [a)] Indri relevance score was added as the ninth factor since this score is provided as default by the search engine to generate candidate pool.

3.4 A Framework for Automated Question Triage for Social Reference

For this study, a process framework for automated question triage for social reference was developed as shown in <Figure 3>.

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, and question filtering. The document retrieval module retrieves documents that are expected to be relevant to the submitted query from the corpus. In this experiment, the researcher selected the top 1000 documents from the search result for doing candidate recommendation. The candidate recommendation module retrieves possible candidates based on the documents selected. In this experiment, it is considered that 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, it is also considered that 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.

www.kci.go.kr



〈Figure 3〉 A process framework for automated question triage for social reference

4. Experimental Design

In order to answer the research question, the experiment was designed and conducted in order to investigate factors that would affect the performance of an automated question triage model for social reference, and then evaluations were conducted in order to assess the model proposed.

4.1 Data Collection

For the experiment, a SR site needed to be selected. For this study, Ask MetaFilter 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. Some advantages of using Ask Metafilter for collecting dataset are as the following:

- 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

www.kci.go.kr

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.

Questions and answers are the information artifact of knowledge from users in social reference 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.

〈Table 6〉 Summary of the dataset

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

For this experiment, only five categories among 20 categories were randomly selected to make this research achievable since each category required a lot of calculation to generate different set of user profiles by the combination of the conditions (see, Table 7).

〈Table 7〉 Summary of categories selected

Category	Computers-internet	Education	Health	Law-government	Science-nature
# of Questions	17,969	2,713	6,730	2,919	1,831
# of Answers	128,280	32,424	93,256	32,773	21,697
# of Questioners	6,570	1,856	3,371	2,064	1,368
# of Answerers	10,168	6,695	9,699	5,568	4,904

4.2 Data Processing

4.2.1 Creating user document

In this study, the researcher 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, the researcher 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

www.kci.go.kr

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, it is possible to 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, the document-centered approach was chosen in order to build user documents.

4.2.2 Creating an index of user documents for search engine retrieval

In this experiment, the researcher 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, the researcher combined a question and all answers to the question as a document.

4.2.3 Topic identification

In the social reference 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.

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 the researcher was interested in tags for using them as keywords that contained some subject meaning, tag "resolved" was ignored in this study.

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

www.kci.go.kr

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 order to identify user's topic of interest, the researcher investigated user's topic distribution using the Stanford Topic Modeling Toolbox. In order to use LDA topic modeling technique, it is required to set the number of topics as a parameter for training; the exact number of topics in the documents is hidden in the concept of LDA topic modeling. In order to set the number of topics for each category, perplexity score (Steyvers et al. 2004) that represent topics, calculated by the Stanford Topic Modeling Toolbox was used.

4.2.4 Creating user expertise profile

A prerequisite for developing question triage systems that recommends candidate answerers to the submitted questions is user expertise profile. User expertise profile 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 question triage system should able to locate relevant candidate answerers to the given question based on that profile applying some algorithms.

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 may result in the failure of getting answers from the expert or extension of turnaround time for answer without estimation.

In tag-based approach, user's topic interest 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.

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

www.kci.go.kr

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

User's performance to providing good answers was calculated by the number of answers that 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 that were answered by the user.

4.2.5 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.

(1) Topic detection of the question

Once a question is given, it needs first to identify the subject and topics of the question. Since a question is submitted to a subject category in real world, the researcher 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. In topic-clustering approach, sub-topics of the question were investigated using LDA topic model. In order to determine topics of the question, a threshold to limit the number of topics less than 5 was set.

```
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 4> Pseudo code for determining topics of the question (topic-clustering approach)

www.kci.go.kr

(2) Selecting the pool of candidates

The goal of this step is 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* (Balog, Azzopardi and De Rijke 2006). In order to obtain feasible document to the given question, the researcher used Indri as a search tool. <Figure 5> explains the procedure of choosing the candidate pool. In this step, relevance score is used in order to retrieve candidate answerers. Furthermore, this relevance score is used as the last factor (x_9) in the estimation.

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

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

, 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:

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

```
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 5〉 Pseudo code for candidate selection - document-based approach

www.kci.go.kr

4.3 Generating Dataset for Logistic Regression Analysis

In order to investigate important factors affecting to decisions of question triage 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.

4.3.1 Conditions to generating dataset for training

(1) Topic modeling approach: the tag-based vs. the topic-clustering-based

In this study, the researcher 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.

(2) Balancing

Examining the generated samples for training, the researcher 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, the researcher tried to append randomly selected relevant cases to the sample dataset.

(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 noise in the dataset. In order to generate good samples for the training, the researcher selected top N number of candidates for each question.

(4) Precision-filtering

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

www.kci.go.kr

(5) Normalize the score

In order to understand the characteristic of our dataset in user profile, the researcher 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. For the purpose of understanding the characteristics of the dataset, the researcher 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.

4.4 Evaluation

4.4.1 Measures for performance evaluation

For the evaluation, the researcher used precision and Mean Average Precision (MAP) in order to measure the performance of the algorithm. In this study, MAP is the mean of the average precision scores for each query for a set of queries.

4.4.2 Relevance judgment

In order to judge relevance of a candidate recommended, the researcher 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 ‘favorite’ answerer to a question; they are interpreted as highly relevant answerers to the question.

4.4.3 Evaluating the impact of factors on the performance

In order to evaluate the impact of factors on the performance of the automated question triage system, logistic regression analysis was used. Once the researcher obtained the best logistic model in terms of performance, then the researcher 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 the researcher 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 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

www.kci.go.kr

or not a candidate is relevant answerer to the question assigned.

4.4.4 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

4.4.5 Logistic regression models

Trying to investigate the impact of the factors selected earlier on the performance of automated question triage for social reference, the researcher defined the following logistic regression model for the evaluation.

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

, where $x_1 \cdots x_9$ are the values of each factors; $\beta_1 \cdots \beta_9$ are the regression coefficients of factors, β_0 is the intercept from the linear regression equation. In the above model, the researcher added relevance score (x_9) since the researcher tried to estimate relevant answerers among candidates who were recommended by the search engine result using relevance score discussed earlier. 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:

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

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

4.5 Training

For training, eight different models were prepared by the combination of the condition of generating dataset discussed earlier. Using the different datasets of the models, binary logistic regression analysis was performed to investigate important factors affecting the performance of estimating

www.kci.go.kr

relevant answerers to the given questions.

Once the researcher obtained the coefficient of each factor for the logit function, the researcher could test the model to evaluate its performance using a real dataset. In order to do the evaluation, the researcher 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.

In addition, the researcher also tried to select the best candidate 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, the researcher had to devise different sets of training data using the top N number of user profiles to the given questions. For this experiment, the top 20, 30, 40, and 50 candidates were selected in order to prepare the training dataset.

5. Results

5.1 Comparison between paired models

In order to select best models between paired models, a paired-samples t-test was conducted. In terms of balancing the number of cases by adding dummy cases (B1) hurt the performance. It is observed that 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; $t(2079) = 3.955$, $p = 0.000$. In terms of precision filtering, training model with dataset that include at least one relevant answerer (F1) does not improve the performance. It is 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; $t(2079) = -1.132$, $p = 0.258$.

In terms of selecting the best performing topic modeling, tag-based approach (TAG) performed better than topic-clustering-based approach (TC). For this reason, TAG-B0F1 model was chosen as the best model, which was tested later with testing dataset.

www.kci.go.kr

〈Table 8〉 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*	.141*	.138*	.134*
TAG-B1F0	MAP	.135	.136	.133	.128
TAG-B0F1	MAP	.366*	.318*	.287*	.267*
TAG-B1F1	MAP	.351	.302	.278	.256
TC-B0F0	MAP	.134*	.131*	.128*	.123*
TC-B1F0	MAP	.131	.125	.122	.117
TC-B0F1	MAP	.336*	.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.

Based on these results, the researcher can conclude that TAG-B0F0N20 and TAG-B0F1N20 models performed best of all the models of different training datasets. Previously, it is 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, the researcher 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).

5.2 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, the researcher 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; $t(5872) = -6.473$, $p = 0.000$.

www.kci.go.kr

〈Table 9〉 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

Note. * indicates significant statistical difference between the two models.

5.3 Testing the selected model with the testing dataset

Previously, the researcher 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, the researcher 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 10〉 Performance comparison between the proposed model and baseline (testing dataset)

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

Note * indicates significant difference.

[a)] 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.

[b)] 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.

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

www.kci.go.kr

the suggested model estimates relevant candidates versus the baseline system.

5.4 Selecting the best fit logistic model and verifying hypotheses

In order to investigate important factors for the automated question triage, the researcher needed to test the full model using the tag-based approach. In order to do this, the researcher 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. The result supports that topic of interest (x_1), the scope of topic area (x_3), contribution to answering (x_4), the performance of providing good answers (x_6), answer quota per day (x_8), and Indri relevance score (x_9) have an impact on estimating relevant answerers to the given question for automated question triage. On the other hand, the scope of subject (x_2), contribution to submitting questions (x_5), and response time to answer (x_7) seemed not to have an impact on the estimation with condition; their coefficients $\neq 0$, but $P > .05$.

〈Table 11〉 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 activity (x_4)	1.742	.100	300.299	1	.000	5.711
Questioning activity (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

In order to make sure whether those factors have an impact on the estimation, the researcher tested a reduced model excluding those three factors — x_2, x_5, and x_7. 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

www.kci.go.kr

model 1 ($\beta_2 = 0$) 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 ($\beta_5 = 0$) 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 ($\beta_7 = 0$) with the condition; LR chi2 (1) = 2.69, P=.101. Thus, it can be concluded that the scope of subject areas (x_2), contribution to submitting questions (x_5), and response time (x_7) do not affect the performance of automated question triage.

〈Table 12〉 Result of Hypotheses test of the study

Factor (IV)	Null Hypothesis (H_{x-0})	Alternative Hypothesis(H_{x-A})
Level of topic of interest (x_1)	H_{1-0} (rejected; β1=.194, P=.000)	**H_{1-A} (favored)**
Scope of subject interest (x_2)	**H_{2-0} (favored)**	H_{2-A} (rejected; β1=.044, P=.120)
Scope of topic interest (x_3)	H_{3-0} (rejected; β1=-1.810, P=.000)	**H_{3-A} (favored)**
Answering activity (x_4)	H_{4-0} (rejected; β1=1.742, P=.000)	**H_{4-A} (favored)**
Questioning activity (x_5)	**H_{5-0} (favored)**	H_{5-A} (rejected; β1=-.031, P=.096)
Relevant answering (x_6)	H_{6-0} (rejected; β1=.287, P=.000)	**H_{6-A} (favored)**
Response time (x_7)	**H_{7-0} (favored)**	H_{7-A} (rejected; β1=.032, P=.101)
Quota (x_8)	H_{8-0} (rejected; β1=.222, P=.000)	**H_{8-A} (favored)**
Indri relevance score (x_9)[a)]	H_{9-0} (rejected; β1=.415, P=.000)	**H_{9-A} (favored)**

Note. Bold indicates favored hypotheses.

Based on the result above, the researcher tested the hypotheses and accepted the favored hypotheses as shown in <Table 12>. Thus, it is revealed that: the user's topic interest, scope of topic interest, answering activity, relevant answering (performance), quota, and Indri relevance score affected the performance of automated question triage; the user's scope of subject interest, questioning activity, response time did not affect the performance of the question triage.

6. Discussion

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

www.kci.go.kr

(category).

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 may vary as the number of topic areas is changed. With a real dataset, it is difficult 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 may attract more answerers than vaguer topic area postings, where the specific question is hidden. In order to verify this explanation, another study focusing on users' answering behavior is required.

In DR settings, the area of subject expertise of the answerer was considered as the second most important factor, following the subject area of the question. In the current study, the subject area of interest of the answerer was not implemented in the suggested user expertise model, but this factor was used prior 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 means that the area of subject expertise of the answerer is still one of the most important factors in both DR and SR settings.

Among the nine factors expected to affect the decision of automated question triage, the user's topic of interest (x_1 and x_9), scope of topic area (x_3), answering activity (x_4), performance in providing good answers (x_6), and quota (x_8) were observed as important factors that should affect automated question triage decisions within social reference. Among them, the user's answering activity was identified as the most important factor to affect the automated question triage decisions within SR, followed by topic of interest (the Indri relevant score and tag-based topic of interest).

In this study, user activities in providing answers and submitting questions were interesting factors to compute user contribution for the estimation. The researcher observed that there is a strong positive relationship between answering activity and the estimation ($\beta = 1.742$, $P = .000$), but no relationship between questioning activity and the estimation ($\beta = -.031$, $P > .05$). Thus, the user's contribution to answers provided was one of the important factors for estimating relevant

www.kci.go.kr

answerers to the question, and answer activity is revealed as the most important factors among them. It also 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.

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 (x_2) and the scope of topic area (x_3) 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 ($\beta = .044$, $P > 0.05$), but the user's scope of topic area was identified as one important factor ($\beta = -1.810$, $P = .000$), when trying to estimate relevant answerers to the question given. 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.

The researcher 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 ($\beta = .287$, $P < .05$, 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.

The factor of response time was also regarded as an important factor affecting the pathway of question triage 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 question triage 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 ($\beta = .032$, $P > .05$).

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 ($\beta = -.222$, $P = .000$, 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

www.kci.go.kr

than other answerers who were not. In other words, a relevant answerer provided answers to a smaller number of questions.

The area of subject expertise of the answerer, which is considered as the third most important factor affecting question triage decisions within digital reference, was not included in the proposed user expertise model for social reference. 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.

In this study, the researcher borrowed factors affecting the decision-making of question triage from DR and tested them for question triage 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 question triage 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.

References

[1] Agichtein, E. et al. 2008. "Finding High-Quality Content in Social Media." *Proceedings of the 2008 International Conference on Web Search and Data Mining*.

[2] Balog, K., Azzopardi, L. and De Rijke, M. 2006. "Formal models for expert finding in enterprise corpora." *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*.

[3] Bian, J. et al. 2008. "Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media." *Proceedings of the 17th international conference on World Wide Web*.

[4] Blei, D. M., Ng, A. Y. and Jordan, M. I. 2003. "Latent dirichlet allocation." *The Journal of Machine Learning Research*, 3: 993-1022.

[5] Bouguessa, M., Dumoulin, B. and Wang, S. 2008. "Identifying authoritative actors in question-answering forums: the case of Yahoo! answers." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*.

[6] Gazan, R. 2007. "Seekers, sloths and social reference: Homework questions submitted to

www.kci.go.kr

a question-answering community." *New Review of Hypermedia and Multimedia*, 13(2): 239-248.

[7] Goldman, A. I. 1999. *Knowledge in a social world*. USA: Oxford University Press.

[8] Harper, F. M., Moy, D. and Konstan, J. A. 2009. "Facts or friends?: distinguishing informational and conversational questions in social Q&A sites." *Proceedings of the 27th international conference on Human factors in computing systems*.

[9] Jeon, J., Croft, W. B. and Lee, J. H. 2005. "Finding similar questions in large question and answer archives." *Proceedings of the 14th ACM international conference on Information and knowledge management*.

[10] Jurczyk, P. and Agichtein, E. 2007. "Discovering authorities in question answer communities by using link analysis." *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. Lisbon, Portugal: ACM.

[11] Kim, S., Oh, J. S. and Oh, S. 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.

[12] Lee, C. T. et al. 2009. "Model for voter scoring and best answer selection in community Q&A services." *Proceedings of the Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on*.

[13] Li, B. and King, I. 2010. "Routing questions to appropriate answerers in community question answering services." *Proceedings of the 19th ACM international conference on Information and knowledge management*.

[14] Liu, M., Liu, Y. and Yang, Q. 2010. "Predicting best answerers for new questions in community question answering." *Web-Age Information Management*, 127-138.

[15] Nam, K. K., Ackerman, M. S. and Adamic, L. A. 2009. "Questions in, knowledge in?: a study of naver's question answering community." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

[16] Park, J. D. 2013. *Automated Question Triage for Social Reference: A Study on Adopting Decision Factors from Digital Reference*. PhD dissertation, University of Pittsburgh.

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

[18] Pomerantz, J., Nicholson, S. and Lankes, R. D. 2003. "Digital Reference Triage: Factors Influencing Question Routing and Assignment." *Library Quarterly*, 73(2): 103.

[19] Qu, M. et al. 2009. "Probabilistic question recommendation for question answering communities." *Proceedings of the 18th international conference on World wide web*.

www.kci.go.kr

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

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

[22] Shah, C. and Pomerantz, J. 2010. “Evaluating and predicting answer quality in community QA.” *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*.

[23] Steyvers, M. et al. 2004. “Probabilistic author-topic models for information discovery.” *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.

[24] Zhang, J., Tang, J. and Li, J. 2010. “Expert finding in a social network.” *Advances in Databases: Concepts, Systems and Applications*, 1066-1069.

www.kci.go.kr

www.kci.go.kr