Effects of Topic Familiarity and Search Skills on Query Reformulation Behavior

Rong Hu
School of Information Management,
Wuhan University
Luojiashan Road #1, Wuhan, Hubei, China
hu_lotus@foxmail.com

Kun Lu
School of Library and Information Studies,
University of Oklahoma
401 West Brooks, Norman, OK, 73019
kunlu@ou.edu

Soohyung Joo
School of Information Studies,
University of Wisconsin-Milwaukee
2025 E Newport, Milwaukee, WI
sjoo@uwm.edu

ABSTRACT
This study examined the effects of two factors, topic familiarity and search skills, on users' query reformulation behavior in health information searching. Four hypotheses were tested. Forty five graduate students participated in our study and searched for health-related topics on our experimental retrieval system. Their search actions were recorded by a server-side log system and their demographic information were collected, including their familiarity with the topics and their major. Our results suggest that topic familiarity and search skills do not have statistically significant impact on users' selection of query reformulation types. However, participants with a higher topic familiarity tend to make less spelling errors and prefer to use specific terms or search from different aspects. Participants with better searching skills are more likely to generalize and specify their queries and make less errors. In addition, significantly fewer reformulations were observed from the participants with higher topic familiarity. This indicates participants with higher topic familiarity can complete their task with less reformulation effort. There is no significant difference on the time spent on different types of query reformulations. The findings from this study yield practical implications for designing health information retrieval systems that support query reformulation for users with different knowledge and skills.

Keywords
Query reformulation, topic familiarity, search skills, factors.

INTRODUCTION

Query reformulation is one of fundamental search strategies that users apply to obtain more precise, relevant search results in achieving a search goal. According to recent studies, more than half of search sessions include at least one query reformulation attempt during the search process in web searching environments (Joo & Lee 2011). Users modify their initial or previous queries based on the judgment of search results. In an effort to achieve better search results, users apply different approaches to reformulating search terms. For example, when search results are too broad, they try to narrow down the results by adding more terms or using more specific terms. On the contrary, when search results are insufficient, they try to extend search results by bringing up more general terms or other terms involving different aspects. Researchers identified different strategies of query reformulation efforts in web searching, such as specialization, broadening, and parallel movement (Rieh & Xie, 2006; Jansen, Spink & Narayan 2007; Huang & Efthimiadis 2009).

Understanding different types of query reformulation strategies is important for IR system designers to develop more effective search algorithms or system features that support users' query manipulation. The study of query reformulations has attracted lots of attention from researchers over the last decade. However, while previous researchers explored query reformulation types and assessed effects of query reformulation on search outputs less research focused on user side factors affecting query reformulation behavior. This study intends to investigate two user factors, including topic familiarity and search skills, which would influence users' selection of query reformulation types. User search logs of an experimental IR system on health information were analyzed to examine the relationships between topic familiarity and search skills and query reformulation behaviors.

LITERATURE REVIEW

Researchers have paid much attention to different types of query formulation strategies in online search environments. Based on open coding and content analysis, Rieh and Xie (2006) analyzed 313 information retrieval sessions of query

logs from the Excite Web search engine. They defined three facets and associated types of query reformulation strategies, such as specification, generalization, parallel movement, term variation, and error correction. In their analysis of Yahoo's logs, Teevan et al. (2007) explored users' information re-retrieval behavior and identified taxonomies of query reformulation such as word merge, remove words, and stemming and pluralization. Jansen, Spink and Narayan (2007) clustered query modification patterns into specialization, content change, and generalization during Web search sessions. While proposing a unified and discriminative model for query refinement, Guo et al. (2008) defined query reformulation types into spelling correction, word splitting, word merging, phrase segmentation and others. Boldi et al. (2011) developed an algorithm to trace query reformulation types automatically. Their algorithm was designed to classify user query reformulations observed in web search engines into several different types, such as generalization, specialization and error correction. From the AOL query log analysis, Huang and Efthimiadis (2009) detected 13 types of query reformulations, including word reorder, whitespace and punctuation, remove words and others.

Researchers also tried to assess the effectiveness of query reformulation. Based on the 13 types of query reformulations they defined, Huang and Efthimiadis (2009) evaluated the effectiveness of different types of query reformulations by analyzing users' click behavior. They found that certain reformulation strategies like adding/removing words, word substitution, acronym expansion and spelling correction would be more effective in obtaining improved search results. Liu et al. (2010) found that the effectiveness of query reformulation types vary by search task type. For example in simple search task tasks, new and specialization strategies were relatively more effective, while in parallel search tasks, word substitution was the most effective query reformulation type. By analyzing information retrieval diary data, Joo and Lee (2011) examined how query reformulation types would affect search output improvement. They found that users obtained relatively better search results after parallel term movement compared to generalization and specialization methods.

More importantly, several researchers investigated factors related to users' query reformulation behavior. In particular, users' domain knowledge, system knowledge, and cognitive abilities were investigated in relation to query reformulation behavior. Hembrooke et al. (2005) scrutinized the effects of domain knowledge on users' query operation, including query creation and following modification. Their findings revealed that expert users would engage in more elaboration and employ more complex queries whereas novice users come up with less effective query strategies, such as plural making/taking, redundancy, poke and hope, and backtracking. Liu, Gwizdka and Belkin (2010) investigated how users' cognitive abilities are associated with query reformulation behaviors, and they found non-significant effect of users' working memory on query reformulation behavior. Joo and Lee (2011) explored how domain knowledge and system familiarity would influence the effectiveness of query reformulations. Their multiple regression analysis revealed that domain knowledge and system familiarity do not have a significant influence on search outputs after query reformulation.

Task type is another important factor in relation to query reformulation behavior. Liu et al. (2010) looked into the relationship between information search task types and query reformulation strategies. Their experimental study confirmed that users' query reformulation patterns are significantly different by task types. For example, specialization was more frequently used in simple and hierarchical tasks while word substitution was more significantly used in parallel tasks. Xie and Joo (2012) found that query reformulation tactics were more frequently employed in scholarly search tasks than occupational or leisure search tasks.

Query behavior has been investigated in health information related tasks. In their early study, Spink et al. (2004) analyzed medical or health queries extracted from commercial web search engines, and identified the top subjects of user queries such as general health, weight issues, and reproductive health and puberty. Toms and Latter (2007) explored consumer users' search process focusing on query behavior based on transaction log analysis. They found that users exhibited significant problems in query formulation and search result evaluation. In an effort to support users' query creation in the health information domain, Zeng et al. (2006) developed a query recommendation system. By analyzing the relationship between original queries and medical vocabularies, they tried to support users to come up with more adequate, specific queries. Recently, Zhang et al. (2012) investigated users' searching and browsing behavior in using Medline Plus. In their study, they analyzed unique query terms in different task situations, and applied semantic analysis to identify different types of query reformulations.

Previous studies have identified a range of users' query reformulation strategies based on query log analysis or diary data, and evaluated the effectiveness of query reformulations. In addition, different factors were investigated in consideration of query reformulation behavior. However, there are fewer studies in query reformulation behavior in the context of health information retrieval. This study focuses on two user side factors, topic familiarity and search skills, that could affect users' selection of query reformulations in the domain of health information. This study is one of the few studies that investigate query reformulation behavior and associated user-side factors in the context of health information searching.

RESEARCH QUESTIONS AND HYPOTHESES

The purpose of this study is to investigate factors that could affect users' selection of query reformulation types in health information searching. We established the following four research questions and corresponding hypotheses:

Q1: Does users' topic familiarity influence their selection of query reformulation types?

H1: Users' topic familiarity has a significant impact on their selection of query reformulation types.

Q2: Does users' search skills influence their selection of query reformulation types?

H2: Users' search skills have a significant impact on their selection of query reformulation types.

Q3: Does users' topic familiarity influence the frequencies of query reformulations in a session?

H3: User topic familiarity has a significant impact on the frequencies of query reformulations per session.

Q4: Is there a significant difference in the time spent on each type of query reformulation?

H4: There is a significant difference in the time spent on each type of query reformulation.

These questions were answered based on a user experimental study of searching health information.

METHODOLOGY

Sampling and Data Collection

To investigate users' query behaviors, we designed an experimental IR system of health information with a Google-like search interface (Figure 1). The query likelihood model implemented by Indri was employed as our retrieval algorithm (Metzler and Croft, 2004). We used the OHSUMED test collection for the experiment. It is an IR test collection of medical articles derived from medical information database, MEDLINE 10, which contains titles and abstracts from 270 medical magazines. Along with it, there are 106 test topics and the corresponding relevance judgments. Six search topics were randomly selected out of them and served as the test topics for our user studies. The detailed descriptions of our test topics are listed in Appendix 1.

Forty five graduate students were recruited from a state university in the United States. They were invited to our usability lab and instructed to carry out assigned information search tasks on health information using our IR system. Two search tasks were assigned in the pre-experiment instruction. One was to provide a definition of a key biomedical term selected from each search topic. The other was to find the relationships between biomedical concepts and to provide an answer for each search topic. Participants were not allowed to access the search system while answering questions. They were given monetary incentive (USD 10 or 15) upon completion of assigned search tasks. Each subject was assigned a User ID. A pre-questionnaire was posted to collect their demographic information (such as gender and how frequently they search the Internet for health information), as well as their familiarity with the six topics and their major. A nine-level Likert scale was used to measure users' perceived topic familiarity from one to nine. Table 1 provides the descriptive statistics of their topic familiarity before searches. The average topic familiarity before searches is approximately 2, ranging from 1 to 8 with a standard deviation of 1.51.

	Mean	Std	Min	Max
Topic 1	2.27	1.51	1	6
Topic 2	2.29	1.70	1	7
Topic 3	1.71	1.41	1	5
Topic 4	2.64	1.58	1	6
Topic 5	2.02	1.78	1	8
Topic 6	1.36	0.86	1	5
All topics	2.05	1.51	1	8

Table 1. Descriptive statistics of topic familiarity.

Each participant was asked to complete three tasks, which are random subsets of the six topics in Appendix 1, using our health information search system. The server recorded all their interactions with the system during their search process. Our data include their queries, the timestamps of their search submissions, the document IDs they viewed, the search task, and their User IDs.

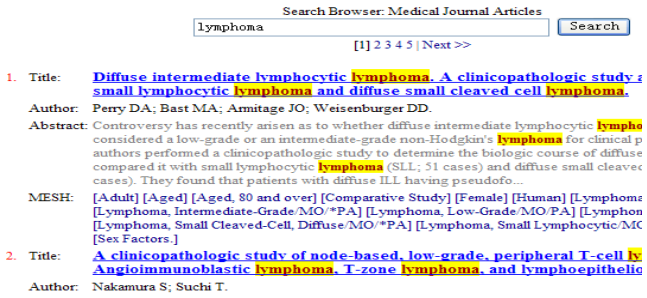

Figure 1. Search interface.

Table 2 presents demographic information of the participants. According to Table 2, nearly 65% of the participants are female, more than 65% of them search health information on the Internet occasionally, and only about 11% never searched.

Data Analysis

In total, we collected 135 search sessions (i.e. each participant completes one search task is considered as one search session), out of which 112 sessions with at least one reformulation action. The purpose of this study is to examine the effects of two factors, topic familiarity and

Gender	male	16	35.56%
	female	29	64.44%
Search health information on the Internet	daily	3	6.67%
	weekly	7	15.56%
	occasionally	30	66.67%
	never	5	11.11%

Table 2. Characteristics of participants (N=45).

search skills, on users' selection of query reformulation types. That is, the two independent variables are 1) topic familiarity and 2) search skills while the dependent variable is their selection of query reformulation types.

We categorized the 135 sessions into two groups according to users' familiarity with the topic: expert sessions (participants' topic familiarity greater than 2) and novice sessions (participants' topic familiarity less than or equal to 2). Although the participants in our study are generally not familiar with the search topics, we believe that participants are relatively more familiar with the topics in expert sessions than in novice sessions. We categorized the participants at the session level as their topic familiarity could differ by topics. It would be more accurate to investigate how they perform when they are familiar or unfamiliar with the topic rather than treat them as domain expert or novice as an individual.

On the other hand, we divided 45 participants into two groups according to their major: Library & Information Science (LIS) major and other major. We assume that students from LIS field are more skillful in searching information. At the time of data collection, LIS major students completed or were taking LIS core courses that covered information accessing and retrieval skills. As a result, we have 34 expert sessions and 101 novice sessions, and 28 LIS major subjects and 17 other major subjects.

The unit of our analysis is at each action of query reformulation. We categorized their query reformulation into six types according to the scheme proposed by Rieh and Xie (2006) with slight modifications. Table 3 presents our detailed coding scheme. Each facet was further categorized as follows:

Content changed

Specification: If users do not satisfy search results, they add terms or use more specific terms to specify the query.

Generalization: If the search results are too specific and users want to obtain more general information, they remove some terms or substitute for more general terms.

Parallel movement: If they fail to find relevant information, they shift query terms including the topic from another aspect.

Content unchanged

Facets	Sub-facets	Example
Content changed	Specification (S)	lymphoma→ lymphoma definition
	Generalization (G)	Diabetic gastroparesis →gastroparesis
	Parallel movement(P)	lymphoma definition→ lymphoma small bowel
Content unchanged	Synonym(Y)	Menopausal→ menopause
Format	Format(F)	definition:menopausal→ definition menopausal
Error	Error(E)	COPT→COPD

Table 3. Classification of query reformulation

Synonym: the content of the topic is not changed, but users use more common terms to substitute the same term to obtain better search results.

Format

Format: format change observed in this study includes: term variations and search operators, such as abbreviation, preposition, singular, plural, and Boolean operators.

Error

Error: query reformulations are needed to correct the previous terms in wrong format or with a typo.

Two of the authors independently coded the 135 sessions according to the proposed scheme. The inter-coder reliability turned out 0.843 measured by Cohen's kappa, which indicates a high level of reliability. Disputed coding was resolved by consulting an external expert in query reformulation studies.

In total, 334 query reformulation actions were observed. Table 4 presents descriptive statistics of the coding results. Altogether the reformulation types "Generalization", "Parallel movement" and "Specification" account for about 78%, "Format", "Error" and "Synonym" accounts for 16.8%, 3.9% and 0.3% respectively and we only observed "Synonym" once. This study reaffirms previous studies that specification is most frequently used as a method of query reformulations (Rieh & Xie, 2006; Liu et al., 2010; Joo & Lee, 2011).

RESULTS

This section reports the findings from this study. First, we examined how topic familiarity and search skills affect users' selection of query reformulation types. Chi-square tests were conducted to examine the first and second hypotheses of this study. As we observed the reformulation type "Synonym" once (Table 4), it was excluded from the statistical tests.

Reformulation Type	frequency	percentage
E	13	3.9%
F	56	16.8%
G	68	20.4%
P	83	24.9%
S	113	33.8%
Y	1	0.3%

Table 4. Frequency of reformulation types.

This study also examined whether participants' familiarity with the topics influences the frequencies of query reformulations in a session, and whether there are any differences in the time spent on each query reformulation type.

Topic Familiarity and Reformulation Types

In this study, topic familiarity was measured by users' perceived familiarity with the given topic. All the 135 sessions were categorized into two groups (Expert sessions vs. novice sessions) according to users' perceived level of topic familiarity.

Table 5 presents the cross-tabulation of the two groups and their query reformulation types (the value of the cell represents the percentage in the row). We found the expert sessions more frequently applied "Generalization" (22.4% vs. 15.0%) and "Format" (22.4% vs. 19.9%) than the novice sessions did. Users with higher topic familiarity usually start their search with specific terms as they have knowledge in terminology used in the topic. Then, they extend search results to explore more items by applying generalization reformulation type. By contrast, the "novice sessions" group exhibited relatively higher percentage of the reformulation type "Error" (4.1%), "Parallel movement" (25.6%) and "Specification" (35.3%) compared to the "expert sessions" group. This reveals that users with low topic knowledge are more likely to make mistakes, such as word misspelling or typo, due to their lack of knowledge in terminology on the topic. Also, they more often start their sessions with general terms since they usually could not come up with appropriate terms at the initial query construction. Accordingly, they need to specify terms to obtain more precise results or try to shift to other aspects of the topic to find relevant information.

Topic familiarity	Reformulation type				
	E	F	G	P	S
Expert sessions (%)	3	22.4	22.4	22.4	29.9
Novice sessions (%)	4.1	19.9	15.0	25.6	35.3

Table 5. Cross-tabulation of topic familiarity and query reformulation type.

However, the statistical analysis did not find significant associations according to the Pearson's Chi-square test: X^2 (4, N=333) =2.779, p>0.05. This is consistent with the finding from the previous study that domain knowledge has no significant influence directly on query reformulation (Joo & Lee, 2011).

Search Skills and Reformulation Types

The relationship between search skills and query reformulation types was also examined. Table 6 shows the cross-tabulation of search skills and query reformulation types (the value of the cell represents the percentage in the row). The "LIS major" group applied more "Format" (17.7% vs. 15.6%), "Generalization" (21.7% vs. 18.5%) and Specification" (35.9% vs. 31.1%).Their knowledge of searching skills may lead them to select "Format", "Generalization" and "Specification" more frequently in reformulating queries. Interestingly, LIS major users more often changed the format of queries. They were aware of that format change could generate more precise search results. Also they were likely to make less error owing to their prior skills in manipulating and inputting query terms to the system.

On the contrary, the "other major" group applied relatively higher percentage of reformulation type "Error" (7.4%) and "Parallel movement" (27.4%) compared to "LIS major" group. They are usually less skillful in how to compose an efficient query to obtain relevant search results and to move searching from one aspect to another.

However, there was no statistically significant association between search skills and the selection of query reformulation types. A Chi-square test accepted the null hypothesis: X^2 (4, N=333) =8.850, p=0.065>0.05.

Topic Familiarity and Reformulation Times

Table 7 presents the descriptive statistics of frequencies of query reformulations in each session. On average, "Expert sessions" group showed fewer query reformulations per session compared to "Novice sessions". Subjects in expert sessions applied less than 2 query reformulations whereas those in novice sessions made approximately 2.7 changes during the session. This result indicates that users with more topic knowledge would be able to accomplish their

Search skills	Reformulation type				
	E	F	G	P	S
LIS major(%)	1.5	17.7	21.7	23.2	35.9
Other major(%)	7.4	15.6	18.5	27.4	31.1

Table 6. Cross-tabulation of search skills and query reformulation type.

search task with less effort in query reformulation, and vice versa.

We also tested if users' topic familiarity influences the frequencies of query reformulations in a session. As the query reformulation frequencies per session are not normally distributed, a non-parametric test was employed in the analysis (Wilcoxon Signed-Rank Test). The result confirmed a significant difference between the two groups ($Z=-2.004$, $p=0.045<0.05$). That is, users with higher topic familiarity would apply fewer query reformulations in completing a search session.

Time Spent on Each Type of Query Reformulation

The last research question is to investigate how much time users spent on each type of query reformulation (i.e. the intervals between query reformulation actions).

As shown in Table 8, we observed extremely high standard deviation, so we report their median values (the unit is in seconds) instead of mean values. "Error" showed a highest standard deviation (131.71) and median (59s), which indicates that users spent relatively longer time when correcting their queries, but has greater varibility in general.

We also examined whether there is a significant difference in the time spent on different types of query reformulations based on Kruskal-Wallis H test. The result indicates that there is no significant difference in the time spent on each reformulation type ($H=6.282$, $p>0.05$).

DISCUSSION

This study investigated two user side factors that would influence users' query reformulation behavior. The results of this study revealed that topic familiarity affects the frequencies of query reformulations as well as difference of each reformulation type spent on time. Although most results turned out statistically insignificant, this study provides some insights into users' query reformulation behavior in the domain of health information.

First, users who have a better understanding of the topics are likely to make less spelling errors and prefer to apply general terms when modifying their search terms (Generalization). Only two "Error" type query reformulations were observed from this group and standardized residuals in "Format" and "Generalization" types of "novice sessions" group are minus which suggest the observed counts are lower than their expected values. These findings indicate that users with high topic familiarity create more correct search statements as they are familiar with the terms in the topic. Also, they can initiate their search with specific terms based on their prior knowledge of the topic. This might naturally lead their next queries more general terms as their first search results are likely to be specific. We observed that many expert sessions include user query manipulation to broaden search results, such as reducing terms, substituting to general terms, or conduct some format change (e.g. abbreviation, preposition, singular, plural, and plus sign). In this way, they are able to improve search results with higher recall rate. On the contrary, users that are unfamiliar with the topic experience difficulty in finding appropriate search terms at the beginning of a session, and they are likely to generate their first query with broader term. Consequently, they have a higher chance to encounter the initial search results that are broad. Then, they look for more specific terms to obtain more precise search results. This explains why users with lower topic knowledge relied more on specification method in this study. In addition, inadequate initial queries fetch irrelevant search results. Thus, users with less topic familiarity sometimes tried to shift their search aspect to explore another aspect of information to solve the assigned search task, which implies Parallel reformulation. This study reassured the typical characteristics of users with lower topic knowledge that they have difficulty in selecting appropriate search strategies in initiating a certain task (Xie & Joo, 2010). Also, Wildermuth (2004) found that low domain knowledge is associated with less efficient selection of concepts to include in the search and with more errors being made in the reformulations of the search tactics.

	N	mean	SD	min	max
Expert sessions	34	1.85	1.82	0	8
Novice sessions	101	2.67	2.63	0	16

Table 7. Descriptive statistics of reformulation times in two topic familiarity group.

	N	median	SD	min	max
E	9	59.00	131.71	7	358
G	43	27.00	54.05	5	245
F	40	30.00	60.84	4	323
P	50	51.50	46.03	3	193
S	79	45.00	65.11	2	322

Table 8. Descriptive statistics of time spent on each reformulation type (measured in seconds).

Second, users with better search skills more frequently applied either generalization or specification types of query reformulations while make less errors. Standardized residuals in query reformulation type "Error" and "Parallel movement" of "information major" group turned out to be negative (-1.7, -0.5), and standardized residuals in "Format", "Generalization", and "Specification" query reformulation type of "other major" group also showed minus values (-0.4, -0.5 and -0.6). These results affirm that users with more search skills make fewer mistakes while apply more various types of query reformulations, such as specification, generalization, and format. Higher search skills lead them to employ Boolean operators more frequently, such as AND, NOT, and among others. In this

way, they kept trying to either generalize or specify search terms in an attempt to obtain more relevant search results. On the contrary, users with lower level of search skills made more mistakes in submitting search terms to the system. Also, when their initial queries are not effective, they preferred to apply parallel reformulation strategies to explore different aspects of the topic. Saito (2002) found that experienced search users well organize their behavior during the search process by constructing queries in a more structured way.

Third, users' topic familiarity affects the frequencies of query reformulations in a session. Users who are more familiar with topics tried less query reformulations in accomplishing their search tasks. This finding reveals that users that are more familiar with the topic accomplish search tasks more efficiently in terms of the number of reformulations. As their queries usually fetch more relevant search results, they do not need as many reformulations as the other group. Thus, users with higher topic familiarity can accomplish their tasks more efficiently with less query manipulation effort.

Fourth, users required more time in conducting "Error" type query reformulations, which involve checking and correcting previous words before sending the modified terms. Also, standard deviation of time spent on "Error" type turned out high, which reveals restoring time from query errors might differ by users.

In this study, users' reformulations of queries in search health information were analyzed based on the coding scheme suggested by Rieh and Xie (2006). However, this study has some limitations. First, 135 sessions from the forty-five subjects might not be enough to generalize the findings. Second, topic familiarity was measured by user perception. Self claimed familiarity might not be reliable as it depends on Likert scale. Thus, some pre-tests might be needed to better represent the levels of user topic knowledge. Third, similarly, user search skills were not objectively measured in this study. Since LIS major subjects took or were taking the course of "Information Access and Retrieval" as required courses, we assume that they have more advanced search skills than other major subjects.

CONCLUSION

In this paper, we proposed four hypotheses to examine the relationship between two factors (topic familiarity and search skills) and users' selection of query reformulation types. We also examined the number of times each reformulation type observed in groups with different levels of topic familiarity, as well as the difference in time spent on different types of query reformulations. The statistical analyses suggest that two factors would not be associated with users' selection of query reformulation type. However, we observed that users with a higher topic familiarity tend to make more correct queries with less spelling errors and start with specific queries and then broaden search results by selecting generalization type. Users with low topic familiarity may make more mistakes and frequently specialize or change aspects to narrow the initial search results or to explore different aspects of the topic. Users with better search skills tend to generalize or specify their queries and make fewer errors than the others, while users with lower search skills are easily to make mistakes and prefer to change terms parallel to find out relevant information from other aspects of the topic. As users with less topic familiarity are less likely to generate adequate search statement from the first query, they require more reformulations of terms. Thus, to support they select more specific terms, query expansion or suggestion can be useful for uses with less topic familiarity (Xie & Joo, 2010). Users that are familiar with the topic have less reformulation actions in completing a search session, and reformulation type "Error" requires more time to correct words before acting next type. To help them to select correct terms, controlled vocabularies, query suggestions and auto filling can be applied in the IR system.

In the future, we plan on a next study involving larger sample and more factors. We will investigate different dimensions of factors comprehensively, such as such as user characteristics (e.g. age, gender, etc.), user knowledge (e.g. domain knowledge, search skills, etc.), task types (e.g. search task type, task difficulty, etc.), and system features (e.g. help features, etc.). More importantly, our next study intends to assess how query reformulation behavior affects search outputs in different search tasks.

ACKNOWLEDGMENTS

We appreciate the help from Hohyon Ryu from University of Texas – Austin in implementing the system. We also thank Iris Xie for her help in resolving the disputed coding.

REFERENCES

Boldi, B., Bonchi, F., Castillo, C. & Vigna, S. (2011). Query reformulation mining: models, patterns, and applications. *Information retrieval, 14*(3), 257-289.

Guo, J., Xu, G., Li, H., and Cheng, X. (2008). A unified and discriminative model for query refinement. In *Proceedings of SIGIR '08* (pp. 379-386), Singapore.

Hembrooke, H. A., Granka, L. A., Gay, G. K., & Liddy, E. D. (2005). The effects of expertise and feedback on search term selection and subsequent learning: Research Articles. *Journal of the American Society for Information Science and Technology*, *56*(8), 861-871.

Huang, J. & Efthimiadis, E. (2009). Analyzing and evaluating query reformulation strategies in Web search logs. In *Proceedings of CIKM' 09* (pp.77-86), Hong Kong, China.

Jansen, B.J., Spink, A., & Narayan, B. (2007). Query modifications patterns during Web searching. In *Proceedings of ITNG '07* (pp. 439-444), Las Vegas, Nevada, USA.

Joo, S., & Lee, J. (2011). Assessing effectiveness of query reformulations: Analysis of user-generated information retrieval diaries. In *Proceedings of ASIS&T 2011*, October 9-13, 2011, New Orleans, LA, USA.

Liu, C., Gwizdka, J., & Belkin, N.J. (2010). Analysis of query reformulation types on different search tasks. In *Proceedings of the iConference 2010, Urbana Champaign*, IL, February 3-6, 2010.

Liu, C., Gwizdka, J., Liu, J., Xu, T., & Belkin, N.J. (2010). Analysis and evaluation of query reformulations in different task types. In *Proceedings of ASIST 2010*, Pittsburgh, Pennsylvania, USA.

Metzler, D., & Croft, W.B. (2004). Combining the language model and inference network approaches to retrieval. *Information Processing and Management, 40*(5), 735-750.

Rieh, S.Y., & Xie, H. (2006). Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing and Management, 42*,751-768.

Satio H., & Miwa K. (2002). A cognitive study of information seeking processes in the WWW: The effects of searcher's knowledge and experience. I*n Proceedings of the 2nd International Conference on Web Information Systems Engineering* (Vol.1, pp.321-327).

Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D. P., Ozmutlu, S., & Ozmutlu, H. C. (2004). A study of medical and health queries to web search engines, A study of medical and health queries to web search engines. Health Information & Libraries Journal, *Health Information & Libraries Journal, 21*(1): 44-51.

Teevan, J., Adar, E., Jones, R., and Potts, M.A. (2007). *Information re-retrieval: repeat queries in Yahoo's logs.* In *Proceedings of SIGIR '07* (pp. 151-158), Amsterdam, The Netherlands.

Toms, E. G., & Latter, C. (2007). How Consumers Search for Health Information. *Health Informatics Journal, 13*(3): 223-235.

Wildermuth, B.M. (2004).The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science, 55*(3), 246-258.

Xie, I. & Joo. S. (2010). Transitions in Search Tactics during the Web-based Search Process. *Journal of American Society for Information and Science Technology. 61*(11): 2188-2205.

Xie, I., & Joo, S. (2012). Factors affecting the selection of search tactics: Tasks, knowledge, process, and systems. *Information Processing and Management, 48*, 254-270.

Zeng, Q. T., Crowell, J., Plovnick, R. M., Kim, E., Ngo, L., & Dibble, E. (2006). Assisting consumer health information retrieval with query recommendations. *Journal of the American Medical Informatics Association: JAMIA, 13*(1): 80-90.

Zhang, Y., Wang, P., Heaton, A., & Winkler, H. (2012). Health information searching behavior in MedlinePlus and the impact of tasks. In *Proceedings of the ACM International Health Informatics (IHI) Symposium 2012*, (pp. 641-650). Miami, FL.

APPENDIX 1.SIX SEARCH TOPICS USED IN THE STUDY

(1) You are a 60-year-old menopausal woman without hormone replacement therapy. You want to know if there are adverse effects on lipids when progesterone is given with estrogen replacement therapy.

(2) You are a patient with cerebral palsy and depression. You want to know about the relationship between cerebral palsy and depression.

(3) You have diabetic gastroparesis. You want to know what the best treatment is for diabetic gastroparesis.

(4) You are a patient with a migraine. You want to know about the treatment of migraine headaches with beta-blockers and calcium channel blockers.

(5) You are a 50-year-old person with COPD. You want to know how theophylline should be used for chronic and acute asthma.

(6) You are a 60-year-old man with severe malabsorption. You want to know about the processes of infiltrative small bowel and information about small bowel lymphoma and heavy alpha chain disease.