

Patterns and Sequences of Multiple Query Reformulations in Web Searching: A Preliminary Study

Soo Young Rieh

Excite@Home, Redwood City, California

Hong (Iris) Xie

School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin

Abstract

While some studies have investigated query reformulation in traditional online systems, there has been little understanding of how users reformulate their queries multiple times within search sessions on the Web. This paper reports on patterns and sequences of query reformulation based on query logs from a Web search engine. The data set contained only search sessions in which multiple query modifications were made. The analysis of data resulted in three facets of reformulation: content, format, and resource. Each facet was further categorized by 10 sub-facets. The results show that while most query reformulation involves content changes, about 15% of reformulation is related to format modifications. Six patterns of query reformulation emerged as a result of sequence analysis: specified reformulation, parallel reformulation, generalized reformulation, dynamic reformulation, format reformulation, and alternative reformulation. Each pattern is discussed with definitions and examples. The results indicate that both planned and situated aspects affect query reformulation in Web searching. The implications for new Web search engine tools and features are also discussed.

INTRODUCTION

The search query is a set of words or phrases a user enters when looking for information on a specific topic or subject. Formulating a search query is a challenging task for most of users because they are required to express their anomalous states of knowledge (Belkin, 1982). Therefore, it may not be easy for users to choose the query terms that represent precisely the topic they are looking for. As a result, users sometimes need to modify their queries based on the search results retrieved. Efthimiadis (1996) differentiates these two stages as the initial query formulation stage for constructing search strategy, and the query reformulation stage for adjusting the initial query manually or with the assistance of the system. In the query reformulation stage, users modify their initial queries and submit new ones that more accurately reflect their information needs (Vélez, Weiss, Sheldon, & Gifford, 1997). Although there have been a number of studies about query reformulation, most were conducted within the context of traditional online information retrieval systems - e.g., Bates' (1990) tactics, Fidel's (1990) moves, Chen and Dhar's (1990) query reformulation process.

According to Jansen, Spink, and Saracevic (2000), while Web search engines follow the basic principles of information retrieval (IR), Web users show very different patterns of searching from those found in the traditional information retrieval systems such as online databases. For example, most Web users do not have many queries per search, and each query tends to be short. Boolean operators are seldom used in Web searching. Spink, Wolfram, Jansen, and Saracevic (2001) compared Web queries collected in 1997 and in 1999, and found that the trend appears to be even shorter queries (fewer items) and shorter sessions (fewer queries per user) over time.

Taking into account both the previous studies on query reformulation in traditional online systems, and the new searching environment on the Web, there is a need to investigate the problem of query reformulation in Web searching. Some studies (e.g., Bruza & Dennis, 1997; Lau & Horvitz, 1999) have discussed query reformulation or refinement in the Internet. However, those studies analyzed the types of "search activities" (e.g., new query, reformulation, request for additional request, substitute, add, delete) rather than focusing on patterns and sequences of query reformulation within each search session, which is what this study intends to investigate. This study addresses the following research questions:

1. How to characterize the facets of query reformulation in a Web searching environment?
2. What kinds of query reformulation patterns and sequences emerge from search sessions?

3. What are the implications for the design of Web search engines derived from the findings of this study?

RELATED LITERATURE

Several studies published in early 1990s provide a good framework for this research because they identify the types of modifications users make at the search reformulation stage. Bates (1990) first presents four levels of search activities - move, tactic, stratagem, and strategy - and further characterizes tactics as with five categories: monitoring tactics, file structure tactics, search formulation tactics, term tactics, and idea tactics. In her paper, "tactic" is defined as "one or a handful of moves made to further a search" (p.578). The tactic related to this research is "term tactics" in which Bates identifies seven different categories: SUPER, SUB, RELATE, REARRANGE, CONTRARY, RESPELL and RESPACE. Fidel (1990) found that searchers make two kinds of "moves," operational moves and conceptual moves. She defines moves as "a list of modifications of search strategies that are aimed at improving the result of a search" (p. 515). The moves are further divided the moves into three groups depending on their purpose: moves to reduce the size of a set, to enlarge the size of a set, or moves to improve a set by increasing both precision and recall. Chen and Dhar (1990) present the concept of "semantic operators" which they define as "moves and actions that change the content of the query" (p. 121). From the data of 30 student subjects, they observed five semantic operators used by searchers to move from one datum to the next. Those operators are ST (synonymous term operator), BT (broader term operator), NT (narrower term operator), AT (adjacent term operator), and DT (disjointed term operator).

The Web has created a new searching environment to users. Spink and her colleagues (e.g., Jansen, Spink, Saracevic, 2000; Spink, Jansen, & Ozmultu, 2001; Spink, Wolfram, Jansen, & Saracevic, 2001) have conducted numerous studies on Web searching behavior based on query log data from the Excite search engine. Among them, Spink, Jansen, and Ozmultu's research specifically examines Web users' query reformulation. Based on 985 Web user search sessions, they found that 33% of users went beyond their first query, and the most common modification was to change a term. According to them, about 35% of queries that were modified still had the same number of terms as the preceding ones, while 19% of queries that were modified had one more term than the preceding ones, and about 16% had one less term.

Recently, there have been at least two studies that investigate diverse search activities in general and query reformulation in particular in the Web environment. Lau and Horvitz (1999) partitioned queries into classes representing different search actions. The refinement classes include: New, Generalization, Specialization, Reformulation Interruption, Request for Additional Results, and Blank queries. Their probability distribution reveals that most actions are either new queries or requests for additional information. Relatively few users refine their searches by specialization, generalization, or reformulation. Bruza and Dennis (1997) categorized 1040 Web queries into 11 different transformation types, and found that the largest number of times users simply repeated a query that they have already submitted. They also found that other main categories of reformulation were term substitutions, additions, and deletions in that order. According to them, users do not often split compound terms, make changes to spelling, punctuation, or case, and use derivative forms of words and abbreviations. Based on these findings, Bruza and Dennis have developed "hyperindex" which aids the user in query term additions and deletions by presenting more specific terms that often contain contextual information.

METHODOLOGY

The data set for this study was generated by Excite@Home's Excite Precision Search (<http://www.excite.com>). Query logs were collected on October 9, 2000 for 24 hours from those users who accepted cookies. As the logs were captured in the bases of individual users' logs, the authors were able to examine how each user reformulated his/her Web query for 24 hours. The log data were composed of time stamp, query, results numbers shown to the user, and URL clicked by the user. Figure 1 presents a portion of the query logs from one search session.

Figure 1. Sample of Query Logs (Session 16)

```
[10/09 06:39:25] Query: holiday decorations [1-10]
[10/09 06:39:35] Query: [web]holiday decorations [11-20]
[10/09 06:39:54] Query: [web]holiday decorations [21-30]
[10/09 06:39:59] Click: [webresult][q=holiday decorations][21]
http://www.stretcher.com/stories/99/991129b.cfm
[10/09 06:40:45] Query: [web]halloween decorations [1-10]
[10/09 06:41:17] Query: [web]home made halloween decorations [1-10]
[10/09 06:41:31] Click: [webresult][q=home made halloween decorations][6]
http://www.rats2u.com/halloween/halloween\_crafts.htm
[10/09 06:52:18] Click: [webresult][q=home made halloween decorations][8]
http://www.rpmwebworx.com/halloweenhouse/index.html
[10/09 06:53:01] Query: [web]home made halloween decorations [11-20]
[10/09 06:53:30] Click: [webresult][q=home made halloween decorations][20]
http://www.halloween-magazine.com/
```

Out of the 1,451,033 users' logs collected, 87,658 users were randomly selected. Since this study concerns the patterns and sequences of query reformulation, a data set containing multiple query modifications within each search session, was needed. Therefore, search sessions that included six or more unique terms were selected. There were 5,427 user IDs in which 6 or more queries were entered for 24 hours. Among them, 272 users' query logs were randomly selected. These 272 users yielded 351 unique search sessions because some of them used the Excite search engine for more than one search session per day. While analyzing the log data manually, it was noted that some of the 351 search sessions did not have more than 5 modifications. To be consistent with the original selection criterion, those sessions were dropped. In addition, a number of search sessions containing pornographic queries were also dropped. As a result, out of the 351 user sessions randomly selected, 183 sessions remained in the final data set.

Table 1. Facets of Web Query Reformulation

Facets	Sub-facets	Examples
Content	Specification	Job+lose → lose+lose+effects [Session 56]
	Generalization	Best Buy San Antonio → Best Buy store [Session 52]
	Replacement with Synonym	Dayton equipment → dayton tools [Session 28]
	Parallel Movement	America west airlines → delta airlines [Session 14]
Format	Term Variations	FIT → Fashion Institute of Technology [Session 20] Camelot community → camelot communities [Session 21]
	Operator Usage	“trajectory methods” → trajectory AND methods [Session 23]
	Error Correction	Dayton equipment → dayton equipment [Session 28]
Resource	General Resource	Newspaper articles on matt gieger → gieger [Session 11]
	Special Resource	“data center” AND “mess” → “data center” AND “mess” AND “picture” [Session 18]
	Site URL	Camelot → camelot.com [session 21]

The coding scheme was developed in order to investigate patterns of query reformulation. The basic unit of analysis was the modification of a query. Analysis of data resulted in three distinct categories of reformulation: content, format, and resource. The distinction between content and format is rooted in Fidel's (1990) conceptual moves and operational moves. While Fidel developed those variables with respect to the results of a search (e.g., whether the results have better recall or precision), the coding scheme of this study attempts to understand users' intentions toward the meaning of a query. In other words, this study focuses on whether users intended to narrow or broaden the meaning of their queries, as opposed to examining whether users wish to narrow or broaden the results as Fidel analyzed. Content modifications refer to those instances in which users were making changes to the meaning of a query. Format modifications include cases in which users made format changes only without changing the meaning of the query by using operators, rearranging the order of multiple terms, re-spelling the query terms, re-spacing the

terms, etc. Resource modifications were developed to code those instances in which users intended to make changes in the types of information resources (e.g., news, images, music files) that they were looking for.

As seen in Table 1, each facet was further categorized as follows:

1. Content
 - *Specification*: Users specify the meaning of the query by adding more terms or replacing terms with those that have more specific meaning.
 - *Generalization*: Users generalize the meaning of the query by deleting terms or replacing terms with those that have more general meaning.
 - *Replacement with Synonym*: Users replace current terms with words that share similar meaning.
 - *Parallel Movement*: Users do not narrow or broaden previous queries. The previous queries and the follow-up queries have partial overlap in meaning, or these two queries are dealing with somewhat different aspects of one concept.
2. Format
 - *Term Variations*: Users change the format of terms while still searching for the same topic. The variations include spelling out the abbreviation, adding a preposition (e.g., of, to, for), changing from the singular to the plural of a word, or vice versa.
 - *Operator Usage*: Users change the format of terms by using Boolean operators (AND, OR, NOT), plus sign (+) for mandatory presence, minus sign (-) for absence of a term, or quotation marks (“ ”) for phrase searching.
 - *Error Correction*: Users change query terms in order to correct the typing error or spelling error they have made.
3. Resource
 - *General Resource*: Users change the query from special resource in the Web (e.g., newspaper articles, pictures, MP3 file) to general Web resource.
 - *Special Resource*: Users change the query in order to restrict the information to a particular resource such as newspaper articles, pictures, or MP3 files.
 - *Site URL*: Users change the query in order to locate the specific Web site by adding domain suffix (e.g., .com, .edu, .org).

In addition to categorizing the types of reformulation actions taken, the authors analyzed the sequences of query reformulation process to identify the patterns that emerged within one search session. Therefore, all modifications in the order of occurrence were numbered. Table 2 presents the examples of data coding.

Table 2. Examples of Data Coding

Session	Content				Format			Resource			Undefined
	Specification	Generalization	Synonym	Parallel Movement	Term	Operator	Error	General	Special	Site	
2	1,2,4	6		3	5		7				
3	3	1,5,8		2,6,10,11		4,7,9					
6	6	2		1, 4, 5		3					
7	7			2, 5			1, 3, 4, 6, 8,				

RESULTS

Based on a preliminary analysis of the data, the results answer the two research questions proposed in the first section: 1) the characterization of facets of Web query reformulation; and, 2) the identification of patterns and sequences of Web query reformulation. The implications for designing Web search engines are provided in the Discussion section.

Facets and Sub-facets of Web Query Reformulation

Three facets of Web query reformulation emerged from the data: content, format and resource. Table 3 presents the frequency and percentage of facets of Web query reformulation. Most of the query reformulations involve content change; content modifications account for more than 80% of the query reformulations. Users try to change the meaning of queries by either specifying or generalizing their previous queries. They also replace the previous query terms with synonyms, or reformulate previous queries to create new ones that share some common characteristics. About fifteen percent of query reformulations are format related. Rather than changing content, users sometimes modify the format of their queries to best represent their information needs, including term format changes (e.g. singular to plural), operators changes, and error corrections. Only one percent of changes are associated with resource reformulation, for example, from a general resource to a specific resource or to a specific Web site. Finally, 2.6% of reformulations cannot be defined.

Table 3. Frequency and Percentage of Facets of Web Query Reformulation

Facets	Frequency	Percentage
Content	874	81.5%
Format	157	14.6%
Resource	12	1.1%
Undefined	28	2.6%
Total	1072	100%

The three types of facets can be further categorized into 10 sub-facets, as discussed in detail in the Methodology section. Table 4 presents the frequency and percentage of sub-facets of Web query reformulation. The results reveal that users do not always try to specify or generalize their queries in the reformulation stage. Interestingly, they move in parallel from one query to another to search information they are looking for. The percentage of parallel movements is almost the same as the combined percentages of specification and generalization movements for content reformulation. With respect to format sub-facets, term variations, operator usage, and change or error correction play equally important roles. It is worth noting that error correction accounts for approximately thirty percent of the total format movements. Because different resources can easily be selected from the Excite Search, overall resource change plays the smallest role in query reformulation.

Table 4. Frequency and Percentage of Sub-Facets of Web Query Reformulation

Facets and sub-facets		Frequency	Percentage
Content	Specification	275	31.5%
	Generalization	149	17%
	Replacement with synonym	43	4.9%
	Parallel movement	407	46.6%
	Total	874	100%
Format	Term variations	59	37.6%
	Operator usage	52	33.1%
	Error correction	46	29.3%
	Total	157	100%
Resource	General resource	5	41.7%
	Specific resource	3	25%
	Site URL	4	33.3%
	Total	12	100%
Undefined		28	100%

Patterns of Web Query Reformulation

This study analyzes not only the facets and sub-facts of query reformulation but also the sequences of query reformulation. The investigation of both the facets and sequence of query reformulation enable us to have a complete picture and better understanding of query reformulation within search sessions. After examining the sequences of query reformulation, six patterns emerge: 1) specified reformulation, 2) parallel reformulation, 3) generalized reformulation, 4) dynamic reformulation, 5) format reformulation, and 6) alternative reformulation.

The first three patterns represent the sessions that focus mainly on one type of sub-facet movement within content. The dynamic reformulation pattern represents those sessions that contain both “specification “ and “generalization” movements, while the format reformulation pattern represents sessions that contain all the query reformulations among sub-facets of format. The alternative reformulation pattern represents the sessions where users give up their original query reformulation patterns and change from one sub-facet to another as an alternative approach to find relevant information.

These six patterns are discussed below; definitions and examples are also provided. For easy explanation, Table 5 presents the abbreviations of all the sub-facets in the following examples. In all examples, [] is used to denote explanations provided by the authors.

Table 5. Abbreviations of Sub-facets

Content				Format			Resource			Unde- fined
Specifi- cation	Generali- -zation	Syno- nym	Parallel Movement	Term	Operator	Error	General	Special	Site	
S	G	Y	P	T	O	E	N	C	I	U

Specified reformulation refers to those sessions in which users try to specify their queries as they search. Each reformulation is the specification of the previous query. While some cases of this pattern contain only specifications, other cases involve the types of query reformulations that are format related, synonym replacement, and parallel movement. Here is one example of specified reformulation ($S1 \rightarrow S2 \rightarrow S3 \rightarrow S4 \rightarrow S5$):

"no prescription" \rightarrow "no prescription" viagra \rightarrow "no prescription" viagra xenical \rightarrow "no
 $S1$ $S2$ $S3$
prescription" viagra xenical phentermine \rightarrow "no prescription" "viagra xenical phentermine herbal
 $S4$
"european pharmacies" \rightarrow "no prescription" "viagra xenical phentermine herbal "european
 $S5$
pharmacies" "order online" (Session 131)

Here is another example of specified reformulation ($S1 \rightarrow P2 \rightarrow T3 \rightarrow S4 \rightarrow S5$):

quality store \rightarrow quality farm store \rightarrow field sprayer \rightarrow field sprayers \rightarrow
 $S1$ $P2$ $T3$ $S4$
trailer field sprayers \rightarrow used trailer field sprayers (Session 107)
 $S5$

Parallel reformulation refers to those sessions where users try to modify their queries from one aspect of an entity to another, or from one thing to another that share common characteristics. Some sessions consist of only parallel reformulations. Other cases of this pattern might also involve the query reformulations that are format related or synonym replacement, while users try to concentrate on the parallel reformulation in the overall process.

Here is an example of “parallel reformulation” ($E1 \rightarrow P2 \rightarrow P3 \rightarrow P4 \rightarrow P5 \rightarrow P6$):

images of solarsystem \rightarrow images of solar system \rightarrow image of waterfall \rightarrow images of seas \rightarrow
 $E1$ $P2$ $P3$ $P4$
images of milkway \rightarrow images of earth quake \rightarrow images of forests (Session 245)
 $P5$ $P6$

Here is another example of “parallel reformulation” ($P1 \rightarrow P2 \rightarrow P3 \rightarrow P4 \rightarrow P5$):

america west airlines \rightarrow delta airlines \rightarrow america west \rightarrow air canada \rightarrow united airlines \rightarrow
 $P1$ $P2$ $P3$ $P4$ $P5$
alaska airlines (Session 14)

Generalized reformulation refers to those sessions in which users start with a narrow concept and try to generalize it step by step. In some cases, users might also have to apply parallel reformulations and other format reformulations in their quest to generalize a query.

Here is one example of generalized reformulation ($G1 \rightarrow P2 \rightarrow P3 \rightarrow G4 \rightarrow G5$):

liberal feminism \rightarrow 4 types of feminism \rightarrow definitions of feminism \rightarrow fruedian theory and
 $G1$ $P2$ $P3$
feminism \rightarrow fruedian theory \rightarrow fruied (Session 168)
 $G4$ $G5$

Here is another example of generalized reformulation ($P1 \rightarrow G2 \rightarrow G3 \rightarrow O4 \rightarrow G5$):

trajectory AND method \rightarrow "average trajectory method software" \rightarrow "average trajectory method"
 $P1$ $G2$
 \rightarrow "trajectory methods" \rightarrow trajectory AND methods \rightarrow trajectory (Session 23)
 $G3$ $O4$ $G5$

Dynamic reformulation refers to those sessions where users do not try to specify or generalize their queries completely. Instead, they might move from specification to generalization, and then try to specify again; or they might move from generalization to specification, and then try to generalize again. This pattern represents a more changeable query reformulation process. Here is one example of dynamic reformulation ($S1 \rightarrow S2 \rightarrow G3 \rightarrow G4 \rightarrow S5$).

Condo \rightarrow condo NY or CT \rightarrow condo NY or CT sales \rightarrow realstateNY or CT \rightarrow realstate NY or CT
 $S1$ $S2$ $G3$ $G4$
 \rightarrow Westcheter county NY Real Estate (Session 173)
 $S5$

Here is another example of dynamic reformulation ($S1 \rightarrow O2 \rightarrow G3 \rightarrow O4 \rightarrow S5$):

tresspassing laws \rightarrow Tennessee tresspassing laws \rightarrow "Tennessee tresspassing laws"
 $S1$ $O2$
 \rightarrow "tresspassing law" \rightarrow +Tresspass +Law \rightarrow +Tennessee +tresspass +law (Session 265)
 $G3$ $O4$ $S5$

Alternative reformulation refers to those sessions in which users have to compromise their information need in the search process after they fail to find certain information. As an alternative solution, users try to either change their queries to those that share some characteristics with their original ones, or just try to generalize their previous queries.

Here is one example of alternative reformulation ($S1 \rightarrow S2 \rightarrow P3 \rightarrow S4 \rightarrow S5 \rightarrow P6$):

SC Karate \rightarrow SC Karate Zombie squad \rightarrow SC Karate Zombie squad of Anderson \rightarrow
 $S1$ $S2$ $P3$
SCSKA [South Carolina Sports Karate Association] \rightarrow SCSKA Anderson SC \rightarrow
 $S4$ $S5$
SCSKA Anderson SC Karate \rightarrow Ray Alexander family Karate (Session 262)
 $P6$

Here is another example of alternative formulation ($S1 \rightarrow S2 \rightarrow S3 \rightarrow P4 \rightarrow G5 \rightarrow G6$):

Jobs \rightarrow +job+lose \rightarrow +lose+jobs+effects \rightarrow +late+effects+lose+jobs \rightarrow
 $S1$ $S2$ $S3$ $P4$
+tardiness+effects+lose+jobs \rightarrow +tardiness+effects+jobs \rightarrow +late to +work (Session 56)
 $G5$ $G6$

Format reformulation refers to those sessions where users try to figure out how to form their queries in a correct way to represent what they are looking for. In this type of pattern, most of the modifications focus on the format, not on the content change. Format reformulations might include term variations, operator usage, operator change, and error corrections.

Here is one example of format formulation ($O1 \rightarrow O2 \rightarrow O3 \rightarrow O4 \rightarrow T5$):

Prasesong AND for AND the AND Widow AND summary \rightarrow Prasesong OR for OR the OR

O1

Widow OR summary → Prasesong AND for AND the AND Widow AND summary → Prasesong
O2 *O3*

OR for OR the OR Widow OR summary → Prase OR song OR for OR the OR Widow OR
O4

summary → Praise OR song OR for OR the OR Widow OR summary (Session 90)
T5

Here is another example of format formulation ($E1 \rightarrow E2 \rightarrow E3 \rightarrow E4 \rightarrow E5$):

"orbitrack" → "orbatrack" → "orba track" → "orbi track" → "orbitrek" → "orbit track" (Session 225)
E1 *E2* *E3* *E4* *E5*

DISCUSSION AND CONCLUSION

Query reformulation is an important issue in information retrieval research. Previous studies have examined users' query reformulation process in searching online databases and other information retrieval systems. This study explores users' multiple query reformulations in searching a Web search engine, and investigates query reformulations with respect to multiple facets (i.e., content, format, and resource). By contrast, previous studies (e.g., Chen & Dhar, 1990) have been primarily concerned with content.

In addition to facets and sub-facets of query reformulation, this study further examines the sequence of multiple query reformulations within search sessions, and identifies the patterns of query reformulation. Previous studies in either online searching (e.g., Bates, 1990; Chen & Dhar, 1990; Fidel, 1990) or Web searching (Bruza & Dennis, 1997; Lau & Horvitz, 1999) rarely have investigated users' multiple query reformulation process within search sessions. Here, instead of investigating individual query reformulation, the authors analyze the complete reformulation process within each search session, taking account of the sequences of query reformulation.

This study shows that content reformulation plays a major role in query reformulation. It accounts for more than 80% of query reformulations, and is involved in five of the six patterns of query reformulation. The results of this study help us to answer one of the questions frequently raised in query reformulation and query expansion research: which levels of reformulation process should be automatic, interactive, or manual. It seems that content reformulation largely depends on users' own tasks and their domain knowledge which lead them to modify their original queries. While query modifications are ultimately made by users, Web search engines should provide a "term tree" which includes a structure of specified terms, generalized terms, and synonyms. In this way, the system might be able to assist users in their thinking process to modify queries that more accurately represent their information needs. Although some search engines have already offered some similar features (e.g., AltaVista's related searches, MSN's popular search topics), these features still have some limitations because most of the terms listed tend to be specific, thus supporting only one aspect of diverse query reformulation patterns. In addition, the list is incomplete, providing only 5-10 terms. Based on the preliminary analysis of this research, Excite has recently launched a new tool called "Zoom In" which offers specified query terms and a spell-checker. However "Zoom In" also has some limitations since it does not support other types of reformulation patterns such as generalized reformulation and parallel reformulation.

Although format reformulations do not occur as frequently as content reformulations do, format reformulations still account for about fifteen percent of the total reformulations. Moreover, it was found that users made numerous mistakes while attempting to reformulate their queries in format, as shown in Table 4. For example, some users wanted to correct errors but could not spell a term correctly after trying several times. More examples were found in the facet of "operators usage" in which users did not put space before and after the Boolean operators, or placed quotation marks around one word. In addition, there were a number of cases where users did not use "+" or "-" sign correctly by putting it between two words without any space. Therefore, a new tool such as "automatic corrector" would be helpful for users if it can detect users' mistakes in format reformulations, and suggest the correct format to users.

Probably one of the more interesting findings of this study is that users reformulate queries by moving one query to another in parallel. The authors believe this finding does support one of the popular features of search engines –

browsing in a Web directory or category. Currently, within one directory or category, users are able to try out several different concepts that are related, but not necessarily specified or generalized. However, one of the limitations in many search engines is that “searching” and “browsing” are treated as two separate behaviors. As a result, it may not be easy for users to change their search strategies from searching to browsing when they have to try an “alternative” approach. Therefore, Web search engines should allow users to move between directories in the same level in addition to upper and lower levels.

The identification of the six patterns indicates that users are not always able to formulate their queries as planned. Situational factors affect their query reformulation process. Web searching, as one type of information retrieval, is the product of both planned and situated aspects (Newell & Simon, 1972; Suchman, 1987; Xie, 2000). The results show that many sessions of specified, parallel, and generalized reformulations are not always straightforward, and that they involve format related reformulations in the middle of the sessions. In dynamic reformulation, users have to change from generalization to specification or vice versa to find what they need. In format reformulation, users have to modify terms, add or change operators, or correct errors to reformulate their queries. Finally, in alternative reformulations, users have to compromise their information needs to take an alternative approach to finding relevant information. The complexity of patterns of query reformulation calls for the need to explore the planned and situated aspects of Web searching, and further recommend more conceptual design ideas for the improvement of Web search engines.

This study identifies the facets and patterns of multiple query reformulations in Web searching by examining the sequences of query reformulation within search sessions. The authors associate the results of this study with existing Web search engine features and offer suggestions for improvement. Further research should focus on the following aspects: 1) extend the data set for more in-depth analysis; 2) explore factors affecting query reformulation process; 3) identify the prototypes of the sequence of query reformulations; 4) compare the similarities and differences of query reformulation process between Web searching and other types of information retrieval; 5) connect patterns of query reformulation to Web search engine design.

ACKNOWLEDGEMENTS

The authors are grateful to Jing Yee Lim at Excite@Home for collecting the query logs. We also wish to thank Halley Silver and Ron Lange at Excite@Home for generously allowing us to use the valuable data set for this paper.

REFERENCES

- Bates, M. J. (1990). Where should the person stop and the information search interface start? *Information Processing and Management*, 26, 575-591.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Parts I. *Journal of Documentation*, 38, 61-71.
- Bruza, P. D. & Dennis, S. (1997) Query reformulation on the Internet: Empirical data and the Hyperindex search engine. 5th RIAO Conference. http://www.dstc.edu.au/RDU/publications/ps_reports/raio97.ps
- Chen, H. & Dhar, V. (1990). Online query refinement on information retrieval systems: A process model of searcher/system interactions. *Proceedings of the 13th Annual International ACM SIGIR Conference*, 115-132.
- Efthimiadis, E. N. (1996). Query expansion. *Annual Review of Information Systems and Technology*, 31, 121-187.
- Fidel, R. (1990). Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science*, 42(7), 515-527.
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management*, 36, 207-227.

Lau, T. & Horvitz, E. (1999). Patterns of search: Analyzing and modeling web query refinement. *Proceedings of the 7th International Conference on User Modeling*, 119-128.

Newell, A. & Simon, A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Spink, A., Jansen, B. J., & Ozmutlu, C. (2001). Use of query reformulation and relevance feedback by Excite users. *Internet Research*, 10(4), 317-328.

Spink, A., Wolfram, D., Jansen, B. J. & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science*, 52(3), 226-234.

Suchman, L.A. (1987). *Plans and situated actions: The problems of human-machine communication*. Cambridge: Cambridge University Press.

Vélez, B., Weiss, R., Sheldon, M. A., & Gifford, D. K. (1997). Fast and effective query refinement. *Proceedings of the 20th Annual International ACM SIGIR Conference*, 6-15.

Xie, H. (2000). Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science*, 51(9), 841-857.