Examining Categorical Transition and Query Reformulation Patterns in Image Search Process

이미지 검색 과정에 나타난 질의 전환 및 재구성 패턴에 관한 연구

EunKyung Chung*
JungWon Yoon**

ABSTRACT

The purpose of this study is to investigate image search query reformulation patterns in relation to image attribute categories. A total of 592 sessions and 2,445 queries from the Excite Web search engine log data were analyzed by utilizing Batley's visual information types and two facets and seven sub-facets of query reformulation patterns. The results of this study are organized with two folds: query reformulation and categorical transition. As the most dominant categories of queries are specific and general/nameable, this tendency stays over various search stages. From the perspective of reformulation patterns, while the Parallel movement is the most dominant, there are slight differences depending on initial or preceding query categories. In examining categorical transitions, it was found that 60-80% of search queries were reformulated within the same categories of image attributes. These findings may be applied to practice and implementation of image retrieval systems in terms of assisting users' query term selection and effective thesauri development.

초 록

이 연구는 이미지 특성 범주와 관련하여 질의 재구성 패턴을 탐색하고자 하였다. 이러한 연구 목적을 수행하기 위해서 Excite 웹검색 엔진 로그 데이터가 사용되었으며, 총 592 세션과 2,445 질의어가 분석되었다. 데이터 분석은 Batley의 정보 형태 구분과 선행 연구에서 밝혀진 팻싯과 서브팻싯을 활용하여 수행되었다. 분석결과는 두가지 형태로 구분하여 제시되었다. 첫째, 질의 재구성에 관한 분석결과이다. 질의 분석 결과, 가장 많은 부분을 차지하는 범주는 특정어(specific)와 지칭어(nameable)이며, 이러한 경향은 다양한 정보 탐색 단계에서도 지속적으로 나타났다. 둘째, 질의 재구성 패턴과 관련하여, 평행이동이 가장 많이 나타났으며, 이러한 경향은 최초 혹은 직전 질의 범주에 따라 근소한 차이를 보였다. 범주 전환 분석에서는 높은 비율(60%-80%)로 검색 질의의 범주가 지속적으로 동일한 범주에 머무르는 경향을 밝혀내었다. 이러한 결과는 이미지 검색 시스템 설계와 구현에 있어서, 이용자의 질의 선정 과정에 도움을 제공하고 효과적인 시소러스 구축 등에 활용될 수 있을 것으로 기대된다.

Keywords: image retrieval, query analysis, query reformulation, query transition, web search
이미지검색, 질의분석, 질의재구성, 질의전환, 웹검색

---

* Assistant Professor, Department of Library and Information Science, Ewha Womans University, Seoul, Korea(corresponding author, echung@ewha.ac.kr)
** Assistant Professor, School of Library and Information Science, University of South Florida, Tampa, FL, USA(jyoon@cas.usf.edu)

▪ Received : 13 May 2010 ▪ Revised : 4 June 2010 ▪ Accepted : 13 June 2010
▪ Journal of the Korean Society for Information Management, 27(2): 37-60, 2010.
〔DOI:10.3743/KOSIM.2010.27.2.037〕

1. Introduction

Visual features of image documents influence not only viewing and interpreting image documents but also indexing and searching image documents. A well-known characteristic of visual materials is that an image conveys multi-layered and inter-woven messages, so an image can be interpreted heterogeneously by image viewers, searchers, and indexers. Another feature is that verbal representation, although it has been broadly adopted as a conventional image retrieval mechanism, has limitations in describing visual features. In other words, since index terms cannot fully represent an image document and users cannot express their needs completely with a query, an image retrieval system which entirely relies on matching between index terms and queries often fails to return what users intended to find. Given the context of matching components between two incomplete representations (i.e. index terms and users' queries), this study bases on the assumption that understanding users' queries in terms of iterative sequences over search stages is beneficial to design image retrieval systems. Information retrieval is an iterative process and rarely a single occurrence. In an image retrieval environment, since visual features of images seem to require more iterative and interactive processes, the iterative characteristics in searches become more distinguished. Previous studies on image search queries have focused on identifying message types as index terms or entry points (Enser and McGregor 1992; Keister 1994; Ornager 1997; Collins 1998; Chen 2001; Choi and Rasmussen 2003) rather than exploring users' query reformulation in terms of sequential processes.

Accordingly, this study aims to investigate how queries are reformulated throughout the search stages. Based on examination of query reformulation over search sequences, this study suggests ways to design image retrieval systems and image indexing mechanisms so they guide query reformulation over the entire search process. Compared to textual documents, image documents have unique visual features and characteristics - different categories of pictorial messages are interrelated in a single image. Therefore, this study will examine not only query term reformulation patterns, which have been a typical approach of previous query reformulation studies, but also categorical transitions during the search process. Specifically, the following research questions will be addressed:

- RQ1: What are the overall characteristics of categorical transitions of image search queries?
- RQ2: What are the overall patterns of image search query reformulations?
- RQ3: How can sets of two consecutive queries be characterized in terms of categorical transition and query reformulation pattern?
- RQ4: How can sequential image search processes be characterized in terms of categorical transition and query reformulation pattern?

2. Related Studies

Information retrieval is an iterative process. From

one single interaction with an information system, in practice, it is oftentimes hard for users to search successfully. Once users become unsatisfied with search results, they attempt to search again keeping the same information needs. Users are likely to reformulate their queries based on the search results, their previous search experiences, subject knowledge, and other associated environments. In a previous Web search engine log analysis study (Spink and Jansen 2002), approximately half of Web users modify their initial queries. Undesirably, this query reformulation process is found to be more frustrating for users than initial query formulation (Rieh and Xie 2006). Given the iterative characteristics of information retrieval, it is clear that understanding patterns of query reformulation are beneficial in designing information retrieval system interfaces. As a result, there have been several studies analyzing query reformulation patterns on the Web (Rieh and Xie 2006; Lau and Horvitz 1999; Bruza and Dennis 1997) as well as on online databases (Wilemuth 2004; Vakkari, Pennanen, and Serola 2003; Sutcliffe, Ennis, Watkinson 2000).

Compared to studies on query reformulation for text document collections, there are only a few recent studies which investigate query reformulation patterns for image collections. Goodrum, Bejune and Siochi (2003) analyzed 71 image searches conducted by 18 graduate students in several Web platforms including web search engines, commercial image collections, and digital library image collections. They identified 18 transition stages including search tools and collection selection (e.g., change search engine, log in specific collections, etc.), queries (e.g., queries which are submitted), relevance judgments (e.g., end search, save image, etc.), and context moves (e.g., move from image surrogate to individual image, etc.). Of the initial queries, 83% were revised in some way, and there were a number of common search strategies - for example, approximately 70% of state transitions include browsing activities. Goodrum et al. indicated the results may suggest automatically generated image search navigation aids. Whereas Goodrum, Bejune and Siochi analyzed overall search behavior patterns, Jörgensen and Jörgensen (2005) focused on search term modification: add, delete and change terms (change to related, broader or narrower terms). They analyzed two search logs from a commercial image provider. Jörgensen and Jörgensen's analysis revealed 62% of searches were modified. Approximately 29-30% of modifications were adding terms, 11-12% were deleting terms, and 58-60% were changing terms. Most of changes were moves to related terms (91-97%), whereas moves to broader and narrower terms were relatively small percentages. In addition, they demonstrated users generally attempt to modify their queries without having clear ideas on the process. These findings basically lead to the concept that query modification supports during users' reformation may be beneficial to retrieving pertinent results (2005).

Although the previous two studies on image search query reformulation provide insightful ideas on image search behavior and term modification process, the authors of this paper believe an application of image documents features to query reformulation and tran-

sition patterns will be beneficial to identifying the characteristics of search queries. Because image documents have layered multiple meanings, categorization analysis has been used to examine image documents, search queries, and users' needs. More specifically, Batley (1988) identified four visual information needs: Specific, General/Nameable, General/abstract, and General/subjective needs. Shatford (1986) determined Specifics, Generics, and Abstracts categories, with each of the categories divided by Who, What, Where, and When facets. Hastings (1995) analyzed queries of art historians using four levels of complexity: 1) who, where, and when, 2) text and artist, 3) color, subject, and style, and 4) subject and why. Jörgensen (2003) identified three image attributes: perceptual, interpretive, and reactive attributes. Eakins and Graham (1999) categorized query types into level 1 primitive features (i.e. color, texture, or shape), level 2 logical features (i.e. objects and persons), and level 3 abstract attributes (i.e. named events, types of activities, emotional or religious meanings). In addition to these studies, several studies examined types of pictorial messages and their importance in perceiving and searching images (Choi and Rasmussen 2003; Griesdorf and O'Connor 2002). This study presumes it is significant to examine query reformulation and transition patterns by considering the types of pictorial meanings identified by previous studies.

3. Research Design

3.1 Data Set

This study used the Web search log of Excite 2001, which has been used frequently in several Web query studies (Spink and Jansen 2002; Eastman and Jansen 2003; Jansen and Spink 2005) as well as the authors' previous study (Chung and Yoon 2009) where more detailed features of the data set are explained. In summary, this study selected 32,664 image search queries, and eliminated the following queries from the data set: consecutive queries which were resent by users without any change, 2) pornographic terms in queries[1)], 3) queries containing simply 'image, picture, photo, etc.', and 4) others (non-English queries, etc.). As a result, a total of 8,434 queries and 5,680 sessions remained. Among the 5,680 sessions, 74% (4,204 sessions) included only initial queries and 26% (1,476 sessions) include revised queries.

As shown in Table 1, overall revision status indicates the number of query revisions drops dramatically. For this study, search sessions which have three or more revisions per session were identified; as a result, a total of 592 sessions and 2,445 queries remained for the final data set. The reason for choosing the queries with three or more revisions is to focus on query reformulation and transition in a more sequential context, rather simple revisions such as spelling corrections, spelling words out, and so on.

1) According to Goodrum and Spink (2001)'s study which used the same data set, 25 terms among the 100 most frequent search terms dealt with sexual context. Queries having pornographic terms were eliminated for analyzing general image search terms. Eliminating pornographic terms are also accordance with a previous study using the same data set (Rieh and Xie 2006).

〈Table 1〉 The revision numbers in queries and sessions

No. of Revision	No. of Session	No. of Query
30	1	30
26	1	26
18	1	18
17	1	17
15	1	15
14	3	42
13	1	13
11	4	44
10	5	50
9	7	63
8	7	56
7	18	126
6	24	144
5	53	265
4	141	564
3	324	972
2	884	1,768
Total	1,476	4,213

3.2 Data Analysis

In order to examine the features of categorical transitions and reformulation patterns of image search queries, this study analyzed the queries from two perspectives. First, overall information needs expressed in a single query were analyzed using Batley's four visual information types. For instance, as shown in Table 2, a query of "multicultural photographs" is categorized in General/Abstract, while a query of "mom and son pic" is in General/Nameable. On the other hand, a query of "pictures of Toronto" is classified as Specific, but a query of "pretty pictures" falls into a General/Subjective category. The analyses using Batley's four visual information types identify the characteristics of categorical transitions of search queries.

A second method for analyzing image queries is at the level of individual terms in each query. The reformulation patterns of individual terms were examined using the facet and sub-facet schemes shown in Table 3. The facets and sub-facets have been developed based on previous query reformulation studies (Rieh and Xie 2006; Lau and Horvitz 1999: Bruza and Dennis 1997). Two facets, Content and Format, include five sub-facets and two sub-facets, respectively.

The Content facet deals with changes and modifications in meanings and semantics from the preceding queries to the current one, and includes sub-facets of Narrower, Broader, Replacement with synonyms, Parallel movement, and Repeated query. The Narrower sub-facet is an attempt to specify meanings and seman-

〈Table 2〉 Batley (1988)'s four visual information types (p.376)

Type	Example
Specific	The Scott Monument, Edinburgh, The River Dee
General/Nameable	A ruined castle, A tram
General/Abstract	An arch shape, A busy street scene
General/Subjective	A pretty scene, A scene which illustrates how times have changed

tics in order to advance the search processes. For instance, a query changing from 'musician photo' to the more specific 'Led Zeppelin photo' is classified as a Narrower sub-facet. In contrast, a Broader sub-facet provides the opposite function, expanding the semantics of a query when moving into the next stage of searches. A query change from 'pre-teen pictures' to 'young pictures' exemplifies a Broader sub-facet. The Replacement with synonyms sub-facet is about utilizing synonyms when users advance their search stages. For example, a query transition from 'female birthing photos' to 'female childbirth photos' involves using the term "childbirth" as a synonym replacement for "birthing." The Parallel movement sub-facet attempts to move to another concept. This movement might overlap with the previous search in meanings or concepts. However, it does not include broader or narrower concepts. An example of Parallel movement is changing a search from 'pictures of candles,' 'pictures of toothpicks,' and to 'pictures of matches(fire).' The Repeated query sub-facet is when users reuse the same exact query within a single session, but its use is not consecutive.

On the other hand, the Format facet contains two sub-facets: Term variation and Operator usage. The Term variation specifies further different variations

〈Table 3〉 Facets/Sub-facets used in query reformulation for images

Facet	Sub-facet
Content	Narrower
	Broader
	Replacement with synonyms
	Parallel movement
	Repeated query
Format	Term variation • Spelling change • Spelling out/Abbreviation • Singular/Plural • Add/delete punctuation • Term splitting/joining • Case change • Term order change • Add/delete preposition
	Operator usage

such as spelling change, spelling out or abbreviation, singular or plural, adding or deleting punctuation, term splitting or joining, case change, term order change, and adding or deleting preposition. The other sub-facet, Operator usage, occurs through using Boolean operators such as AND, OR, and NOT.

4. Findings

4.1 Overall Characteristics of Query Reformulation

According to Batley's visual information types, categories of search queries were classified as shown in Table 4. The majority of query types were identified as two categories, the Specific and the General/Nameable. Comparatively, only 4.13% and 0.49% are associated with General/Abstract and General/Subjective categories, respectively.

The categorical distributions of search queries were analyzed in terms of different search stages including the initial search queries, the second search queries, the third search queries, and the fourth search queries, and more than the fifth search queries. As shown in Table 5, overall distributions among individual search stages demonstrate similar percentage patterns to those in Table 4. The majority of queries are in the Specific and the General/Nameable categories, while only small portions of queries are classified in the General/Abstract and General/Subjective categories.

More specifically, as shown in Table 6, the overall characteristics of query reformulation are identified as having two facets, the Content and the Format. In fact, among 1,926 occurrences of sub-facets, 79.55%

〈Table 4〉 Categories of search queries

Batley's Type	Number	%
Specific	1,189	48.63
General/Nameable	1,143	46.75
General/Abstract	101	4.13
General/Subjective	12	0.49
Total	2,445	100.00

〈Table 5〉 Categories of search queries in sequence

Batley's Type	1st query	%	2nd query	%	3rd query	%	4th query	%	〉= 5th query	%
Specific	282	47.64	289	48.82	296	50.00	123	45.90	63	49.61
General/Nameable	284	47.97	274	46.28	271	45.78	131	48.88	54	42.52
General/Abstract	23	3.89	26	4.39	23	3.89	14	5.22	7	5.51
General/Subjective	3	0.51	3	0.51	2	0.34	0	0.00	3	2.36
Total	592	100.00	592	100.00	592	100.00	268	100.00	127	100.00

〈Table 6〉 Frequency of query reformulation in facets and sub-facets

Facet	Sub-facet	Frequency	%
Content (79.55%)	Narrower	378	24.67
	Broader	224	14.62
	Replacement with synonyms	72	4.69
	Parallel movement	754	49.22
	Repeated query	104	6.79
	Sub total	1,532	100.00
Format (20.45%)	Term variation	287	72.84
	Operator usage	107	27.15
	Sub total	394	100.00
Total		1,926*	

* the total number of sub-facets from only revised queries including multiple sub-facets in a query.

are associated with the Content facet, while only 20.45% are related to the Format facet. In the Content facet, five sub-facets are further analyzed, whereas two sub-facets are identified in the Format facet. In the Content facet, the Parallel movement sub-facet is the most frequently used in queries consisting of 49.22%, and the Narrower and the Broader sub-facets accounts for 24.67% and 14.62%, respectively. The dominant usage of Parallel movement sub-facets reveals that users primarily attempt to change their search queries with associated concepts in some senses - athough those concepts are not connected equivalently or hierarchically. In the cases of the Broader and the Narrower sub-facets, users utilize hierarchically related concepts as well as drop or add terms for broadening or narrowing their searches. Only a small portion of facets (4.7%) account for users adding or replacing synonyms. Moreover, a few users attempt to search images using the exact same query which is used in previous search stages (6.79%).

On the other hand, the Format facet includes two sub-facets, Term variation and Operator usage. The Term variation and the Operator usage account for 72.84% and 27.15%, respectively. More specifically, the sub-facet of Term variation deals with slight query term changes, for instance, from 'beer funneling pictures' to 'beer funnel pictures', while the Operator usage is related to including one or more of several Boolean operators (AND, OR, and NOT) or mathematical operators (plus and minus signs) in revised queries. As presented in Table 7, eight types of term variation have been found. Primarily, the Spelling change, the Spelling out/Abbreviation, the Term order change, and the Add/delete preposition consist of the majority of term variation types accounting for 22.53%, 19.75%, 18.21%, and 14.51%, respectively. By contrast, the rest of term variation types are identified as the Term splitting/joining, the Singular/Plural, the Case change, and the Add/delete punctuation.

The analyses of query reformulation patterns are conducted on different search stages as shown in Table 8. The distributions over different search stages are not only comparable with each other, but similar to overall distributions in Table 6. The Parallel move-

〈Table 7〉 Frequency of term variation types in the Format facet

Type of term variation	Frequency	%
Spelling change	73	22.53
Spelling out/Abbreviation	64	19.75
Term order change	59	18.21
Add/delete preposition	47	14.51
Term splitting/joining	28	8.64
Singular/Plural	22	6.79
Case change	16	4.94
Add/delete punctuation	15	4.63
Total	324*	100.00

* the total number includes multiple sub-facets in a single query.

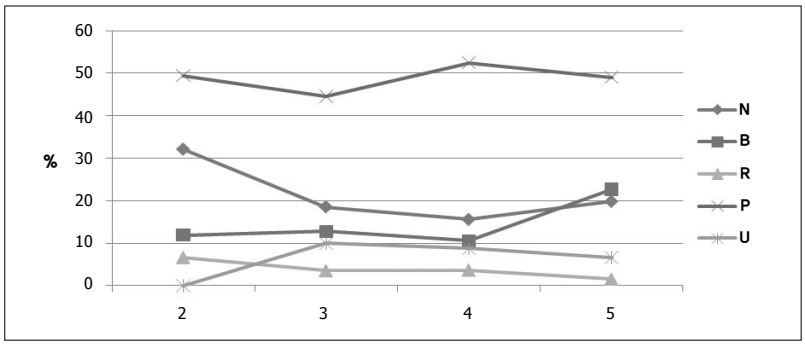
〈Table 8〉 Sequential frequency of facets and sub-facets as searches progress

Facet	Sub-facet	2nd query	%	3rd query	%	4th query	%	5th query	%
Content	Narrower	151	32.06	112	18.54	43	15.58	21	19.81
	Broader	56	11.89	77	12.75	29	10.51	24	22.64
	Replacement with synonyms	31	6.58	21	3.48	10	3.62	2	1.89
	Parallel movement	233	49.47	217	44.56	117	52.47	52	49.06
	Repeated query	0	0.00	60	9.93	24	8.70	7	6.60
	Sub total	471	100.00	487	100.00	223	100.00	106	100.00
Format	Term variation	101	74.26	87	74.36	37	69.81	18	64.29
	Operator usage	35	25.74	30	25.64	16	30.19	10	35.71
	Sub total	136	100.00	117	100.00	53	100.00	28	100.00
Total		607		604		276	100.00	134	

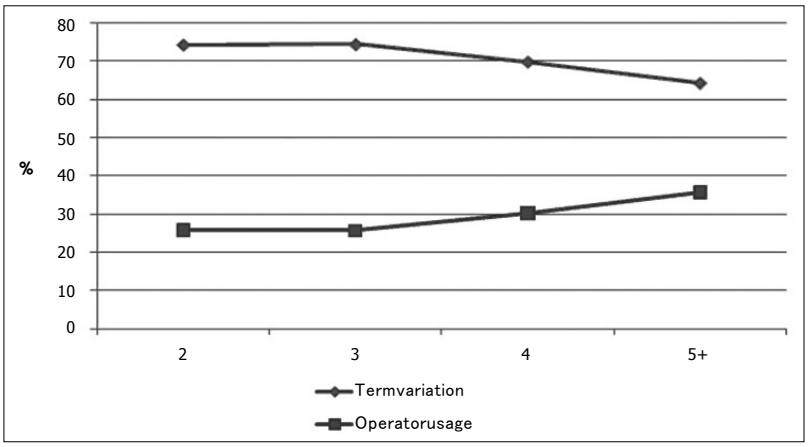
ment occurred most frequently, the Narrower and the Broader are next, and the Replacement with synonyms and the queries in the Repeated query sub-facet have the fewest occurrences.

To recognize the trends of query reformulation over several search stages, the sub-facets in the Content facet are presented in Figure 1. In this representation, two distinctive patterns are recognized in terms of increasing and decreasing trends. The use of the Narrower sub-facet decreases in the latter stages of searches, while the Broader increases as searches progress. Considering the shape of the Narrower and the Broader, as the search progresses, users are more likely to reformulate their queries into broader concepts than into narrow ones. By contrast, the use of Parallel movement in reformation remains steady throughout the various search stages, occurring at approximately 50% of the time in each stage.

The sub-facets of the Format facet, as shown in Table 8, tend to stay consistent as image searching progresses. The proportions between the Term variation and the Operator usage are shown in Figure 2.



〈Figure 1〉 The Content facet transition patterns as searches progress
(N: Narrower, B: Broader, R: Replacement with synonyms,
P: Parallel movement, U: Repeated query)



〈Figure 2〉 The Format facet transition patterns as searches progress

The Term variation decreases as searches progress, while the Operator usage increases. Users are likely to use more operators in later stages of searches, rather than to change terms in terms of format in their queries.

4.2 Categorical Transitions of Consecutive Queries

Categorical transition characteristics of image search queries are analyzed in terms of query categories and facets (Appendix A). More specifically, pairs of consecutive queries are analyzed, in order to examine whether the category of a preceding query affects the category and reformulation pattern of a following query. First, according to the Batley's visual information types, the characteristics of categorical transitions have been recognized as shown in Table 9. Most frequently occurring transitions are the movement within the same category. When users form

〈Table 9〉 Categorical transition frequency and percentage (%)

Preceding category / Following category	A		S		G		P	
	Frequency	%	Frequency	%	Frequency	%	Frequency	%
A	52	61.18	7	0.77	19	2.12	2	22.22
S	8	9.41	778	85.21	143	15.92	0	0.00
G	22	25.88	127	13.91	733	81.63	5	55.56
P	3	3.53	1	0.11	3	0.33	2	22.22
total	85	100.00	913	100.00	898	100.00	9	100.00

(A: General/Abstract, S: Specific, G: General/Nameable, P: General/Subjective)

their queries in a category, the reformulated following query is most likely to be in the same category. In case of the Specific and the General/Nameable categories, transitions within the same categories account for more than 80%. On the other hand, the Abstract category, although it occurred only 85 times, shows relatively diverse categorical transitions to a following query compared to the Specific and the General/Nameable categories; the queries following the Abstract category which are classified in the General/Nameable (25.88%) and the Specific categories (9.41%) are considerable. In addition, although the General/Subjective category may present highly diverse transitions from the original category, the number of occurrences (9 times) is too small; it may be difficult to interpret meaningful transition patterns. However, it underscores that there are very few queries originating from the General/Subjective category.

To recognize trends of categorical transitions between two consecutive queries in terms of facets and sub-facets in query reformulation, Table 10 presents dominant reformulation patterns. In other words, for each preceding category, sub-facets occurring more than 5% are included in Table 10. Consistent with the analysis results in the previous section (4.1), Parallel movement occurs in most categorical transitions. Analysis also showed Term variation frequently used when query reformulation remains within the same category. Each categorical transition demonstrates notable characteristics. First, when users transition from the General/Abstract(A) category to the same category, the Broader (e.g., from “Scenic Nature Photography” to “Scenic Photography”) and the Replacement with synonyms categories (e.g., from “photographs of multi-cultured children” to “photos of multi-cultured kids”) are frequently used. When queries move to the General/Namable category, the Narrower (e.g., from “Weather illustration/pictures” to “hurricane pictures”) pattern is often used. Parallel movement occurs in all types of transitions, from the General/Abstract (A) to the other three categories. These results suggest that users dissatisfied with a preceding abstract query tend to broaden the search query by dropping terms, replacing them with synonyms, or moving to the General/Nameable (G) category by adopting nameable concepts which relate to the preceding abstract concept.

Second, from the Specific category (S), users either transited to the same category or the General/Nameable category (G). Regarding the first type of

transition, to the same category, Parallel (e.g., from "photos of madrid" to "photos of "rio de janerio"") and Narrower movement (e.g., from "ronin warriors picture gallery" to "miya and ronin warrior pics") were frequently found as well as Broader movement (e.g., from "jennifer beals flash dance pictures" to "flash dance pictures"), Operator usage, and Term variation. The second type of transition, to the General/Nameable (G) category, contained two types of reformation: Broader (e.g., from "roger rabbit pictures" to "cartoon pictures") and Parallel movements. Depending on categorical transition, two different types of broadening and narrowing patterns are found: when transitioning to the same category (S), users add or delete terms rather than using conceptually narrower or broader terms, whereas when transitioning to the General/Nameable category, users included generic terms for specific terms.

Third, from the category of the General/Nameable (G), users are likely to reformulate their queries into the same and the Specific (S) categories. In the General/Nameable category, there are six sub-facets recognized: Broader, Narrower, Parallel movement, Repeat query, Operator usage, and Term variation. The Specific category contains two facets: Parallel and Narrower movement. In the case of the transition to the same category (G), queries were broadened or narrowed either by adding or dropping terms or by applying hierarchically related concepts; whereas in the case of transition to the Specific (S) category, specific examples of the nameable concept were used (e.g., from "cartoon pictures" to "tarepanda pictures").

〈Table 10〉 Categorical transitions with facets occurring more than 5%

Preceding category Following category	A	S	G
A	Broader **Parallel movement*** Replacement with synonyms ***Term variation****		
G	**Parallel movement*** Narrower	Broader Parallel movement	**Broader*** **Parallel movement*** **Narrower*** Repeat query *Operator usage* ***Term variation****
S	Parallel movement	Broader **Parallel movement*** **Narrower*** *Operator usage* ***Term variation****	Parallel movement Narrower

(A: General/Abstract, S: Specific, G: General/Nameable; * stands for an occurrence of more than 10%; The General/Subjective category is not included because of its small data size, but raw data can be found in Appendix A.)

4.3 Categorical Transition of Sequential Queries

Whereas the previous section examined the query reformulation characteristics appearing in two consecutive queries, this section focuses on sequential query transition processes, from the initial to final queries. Based on the categorical analyses of queries for image, the categorical transitions are recognized as three patterns: the generalization, the specification, and the abstraction transition patterns. Each of these three patterns represents the query transitions approaching a single query category as image searches progress. For instance, when a query moved toward the General/Nameable (G) category regardless of its starting category, this type of transition is defined as the generalization pattern. Similarly, when a query moved toward the Specific (S) category, this pattern is referred as the specification pattern. The abstraction transition pattern is defined when a query moved toward the General/Abstract (A) category.

The generalization transition patterns can be described as the processes of query transition from any starting query category which concludes with the G category. As shown in Table 11, this pattern is grouped into three sub-types: generalization starting with A, G, or S. Each of these will be explained in detail. The generalization starting with the A category takes place only five times out of 274 total generalizations. Out of these five occurrences, the transition pattern starting with the A category is identified as moving toward the G category through three possible paths: solely through the A category, solely through the G category, or alternately using the A and G categories. By comparison, generalization starting with the S category has more varied combinations. However, the most frequently occurring of these combinations are similar. A major transition from the S category occurs either solely through the G category, solely through the S category, or through a combination of the G and S categories. While these types of transit patterns accounts for approximately 43 times or 93.5% of the Specific starting generalizations, the remaining various transitions (those involving the A category or combining all three categories) each occur only once. Lastly, the generalization starting with the G category exhibits a recursive transit pattern, which occurs when the starting generalization (in this case the G category) matches the final generalization in a group of sequential queries. This generalization starting with the G category consists mainly of transitions solely within the G category, which account for 189 transition patterns out of 223 (84.7%). By comparison, transitions involving a combination of the S and G categories are 10.76%, and transitions including the A category or three different categories occur even more infrequently.

Moving on to the specification transition pattern described in Table 12, the specification transition pattern can be defined as the processes of query transition which begins with any query category and concludes with the S category. Within this pattern, the starting categories of A, S, or G can be compared. Transitions starting with the A category were the most infrequent, occurring only six times out of the

〈Table 11〉 Generalization transition patterns

Categories	Transition	Frequency	% (within pattern)
Generalization starting with A (1.82%)	A→[G] →G	2	40.00
	A→[A]→G	2	40.00
	A→[A,G]→G	1	20.00
	Sub total	5	100.00
Generalization starting with S (16.79%)	S→[G]→G	19	41.30
	S→[S]→G	18	39.13
	S→[A]→G	1	2.17
	S→[S,G]→G	6	13.04
	S→[A,G]→G	1	2.17
	S→[S,P,G]→G	1	2.17
	Sub total	46	100.00
Generalization starting with G (81.39%)	G→[G]→G	189	84.75
	G→[S,G]→G	24	10.76
	G→[A,G]→G	8	3.59
	G→[P,G,S]→G	1	0.45
	G→[S,G,A]→G	1	0.45
	Sub total	223	100.00
Total		274	

Categories within [ ] imply that those categories occur multiple times. (A: General/Abstract, S: Specific, G: General/Nameable, P: General/Subjective)

296 specification transition patterns. Half of these six occurrences only included the S category, and transitions with only the A category, the G category, or a combination of the A, G, and S categories each occurred only once. In contrast the most frequent occurring transitions were those starting with the S category. This demonstrates the recursive pattern for the specification transition patterns, which was similar to the recursive pattern for the generalization transition patterns (see Table 11).Within transitions starting with the S category, the most common trend was using the S category exclusively (85%). Although not as common, the next trend was those patterns which alternated the S and the G categories (7.3%), and this was followed by patterns involving only the G category in the middle (5.58%). Patterns alternating between the A and S categories occurred two times, and those involving only the A category in the middle occurred once. The last pattern described in table 12, transitions starting with the G category, occurred 57 times overall. Out of these 57 times, transitions which involved either solely the S category or the G category comprise 46 (80.7%) of the transitions, alternating the S and the G categories accounts for 10 (17.54%) of the transitions, and combining all categories, S, G, P, and A, occurs once.

Table 13 describes the abstraction transition patterns. Overall abstraction transition patterns occur only 17 times in total, far fewer than the generalization (274 times) and the specification (296 times)

〈Table 12〉 Specification transition patterns

Categories	Transition	Frequency	% (within pattern)
Specification starting with A (2.03%)	A→[S]→S	3	50.00
	A→[A]→S	1	16.67
	A→[G]→S	1	16.67
	A→[A,G,S]→S	1	16.67
	Sub total	6	100.00
Specification starting with S (78.72%)	S→[S]→S	200	85.84
	S→[G]→S	13	5.58
	S→[A]→S	1	0.43
	S→[S,G]→S	17	7.30
	S→[A,S]→S	2	0.86
	Sub total	233	100.00
Specification starting with G (19.26%)	G→[S]→S	28	49.12
	G→[G]→S	18	31.58
	G→[S,G]→S	10	17.54
	G→[S,G,P,A]→S	1	1.75
	Sub total	57	100.00
Total		296	

(A: General/Abstract, S: Specific, G: General/Nameable, P: General/Subjective)

〈Table 13〉 Abstraction transition patterns

Categories	Transition	Frequency	% (within pattern)
Abstraction starting with A (70.6%)	A→[A]→A	9	75.00
	A→[G,A]→A	2	16.67
	A→[G]→A	1	8.33
	Sub total	12	100.00
Abstraction starting with S (17.65%)	S→A→A	2	66.67
	S→[G,A]→A	1	33.33
	Sub total	3	100.00
Abstraction starting with G (11.76%)	G→[A]→A	1	50.00
	G→[G]→A	1	50.00
	Sub total	2	100.00
Total		17	

(A: General/Abstract, S: Specific, G: General/Nameable, P: General/Subjective)

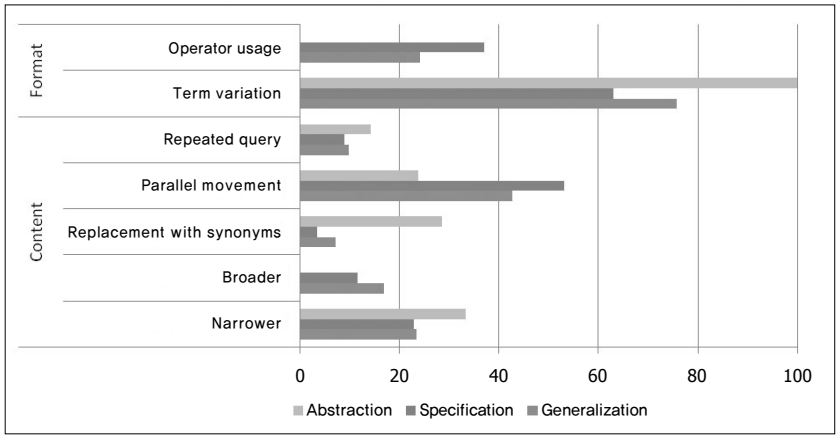
transition patterns described in tables 11 and 12, respectively. Despite its infrequency, the abstraction transition patterns reveal trends similar to the generalization and specification patterns. First, the abstraction starting with the A category, which is the recursive pattern, is found as the most common occurring pattern. It occurs 12 times out of the 17 (70.6%) total abstraction patterns. Of these 12 occurrences,

the major pattern is the transition pattern which includes solely the A category (75%). The remainder are patterns containing the G category or patterns alternating the G and the A categories. Abstraction transition patterns starting with the S or G categories occurred three times and two times, respectively. Since the number of occurrence is so small, it may be difficult to find meaningful semantics.

As the three transition patterns are analyzed from tables 11-13, a primary pattern among three transition patterns is recognized as the recursive patterns (70-81% for three transition patterns). In terms of recognizing the recursive transition patterns, results suggest users are more likely to stay in the same category as the query's starting category (75-86% of each recursive patterns exclusively stay in the same category). Figure 3 presents these exclusive recursive patterns, classifying them into facets and sub-facets in order to identify the transition patterns. The generalization and the Specification transition patterns demonstrate a similar trend. Within the Content facet, queries in both transition patterns are most frequently categorized in the sub-facet of Parallel movement. In addition, the Narrower sub-facet occurs more than the Broader sub-facet in both transition patterns. Although among the recursive patterns the Abstraction transition pattern occurred considerably less than the Generalization and Specification transition patterns, it demonstrated a significantly different trend. It has more queries in the Narrower and Replacement with synonyms sub-facets, rather than the Parallel movement sub-facet. Within the Format facet, however, the Term variation sub-facet was the most frequent for all three transition patterns; although for the Specification transition, the Operator usage sub-facet was also frequently used.

5. Discussion and Implication

As the information needs of users do not always



〈Figure 3〉 Percentage comparison among three recursive patterns

translate into the desired search results, there is a gap between users' initial and repeatedly modified queries and IR systems' appropriate reactions to them. This gap may occur because IR systems seem to consider modified queries as simply another individual search query, although users reformulate their queries hoping for more relevant search results. This gap may be more problematic in image searches, since an image document contains more complicated, multi-layered, and interwoven semantic meanings. By focusing on categorical transition and reformulation patterns of queries, this study examined a dataset of image search queries as users modified their queries in different search stages.

First, overall, queries can be divided into two main categories: the Specific category and the General/Nameable category. There are also the General/Abstract and General/Subjective categories, but the occurrence of these categories is small (approximately under 5% of queries). Users are likely to formulate queries consisting of specific and nameable terms rather than abstract and subjective terms. More importantly, this broad categorical distribution of image search queries stays in a similar pattern as users' searches move toward later stages. Although previous studies did not investigate query transitions, the aforementioned categorical distributions are compatible with previous studies which analyzed image search queries (Chen 2001; Choi and Rasmussen 2003). In conjunction with these previous studies, this study's findings support the importance of the Specific and the General/Nameable categories as access points for image searches. However, the dominance of these two categories might be primarily based on users' previous experiences and perceptions on the capabilities of image retrieval systems. Supporting this claim, image sorting and describing studies demonstrate that abstract and subjective meanings significantly affect how users perceive and understand images (Jörgensen 1998; Rorissa and Hastings 2004; Laine-Mernandez and Westman 2006; O'Connor, O'Connor, and Abbas 1999) As Jörgensen (2003) addressed, users are not likely to use abstract and subjective queries because experienced users recognize that typical image retrieval systems do not support these types of pictorial messages. Therefore, the authors believe that although abstract or subject meanings occur infrequently in queries, they need to be explored further as possible access points for users because of their significance as image descriptors.

Second, categorical transition analysis over various search stages were conducted at two levels: between two consecutive queries and in context of a sequence of queries from a starting query to an ending query. The categorical analysis between two consecutive queries identified that categorical transitions within the same category occurred the most frequently (60 - 80% depending on categories). The tendency of maintaining the same category occurs more frequently in the Specific and the General/Namable categories compared to the General/Abstract category. Similarly, analysis of a sequence of queries from a starting query to an ending query demonstrated that when users start their query in one category they are more likely to stay in that same category. More specifically, transition patterns of these queries are classified into

the following transition patterns: generalization (i.e. moving toward the General/Namable category), the specification (i.e. moving toward the Specific category), or abstraction (i.e. moving toward the General/Abstract category). Among these three patterns, a common distinguishing trend is the exclusive recursive transition pattern, which involves maintaining only one category (of the G, S, or A categories) throughout the transition processes. Accordingly, these findings suggest users tend to transit their queries within the same category they started with initially or previously, and propose that an image retrieval system should be able to help users navigate their searches with terms from the same category. While taking into account that searches tend to maintain the same category, transitions which cross categories are also substantial, occurring between 20% and 40% of the time depending on the starting or previous category. Since users do not often encounter image retrieval systems which allow for searching by abstract and subjective terms, addressing this aspect of searching is important. As Enser, Sandom, Hare and Lewis (2007) note, the semantic content of image documents consist of a blend of general, specific, and abstract facets. Exploring this mixture of general, specific, and abstract concepts contained in an image document needs to be considered a sound approach to improving image retrieval systems.

Third, query reformulation is identified as two facets, the Content and the Format, in the context of query transition between two consecutive queries. In general, more query reformulations fall into the Content, rather than the Format, facet. This overall pattern remains consistent over different search stages. However, when analyzing search queries in detail via facets and sub-facets, there are substantial differences consistently occurring over different search stages. From the Content facet, at the first reformulation stage (e.g., formulating the 2nd query), users tend to narrow down their search queries. Later, as users progress, they are likely to broaden out their queries from a previous query. This result suggests the first reaction of users receiving a huge number of search results is an attempt to reduce their search results by narrowing down their searches. However, still dissatisfied with results, users keep modifying their queries several times seeking to retrieve more images which utilize the query terms in broader concepts. Regardless of search stages, Parallel movement, which changes search topic to related concepts, was the most frequently used sub-facet. This finding is consistent with Jörgensen and Jörgensen (2005)'s study which demonstrated that most query changes were to related terms, with changes to broader or narrower terms occurring infrequently. Within the Format facet, Boolean operators were used infrequently for reformation; instead, Term variation was more likely to be incorporated over different search stages. As can be noticed from Table 7, most types of Term variation sub-facet can be implemented in automated ways. For instance, most current Web search engines correct users' spellings in order to reduce their query revisions and consequent dissatisfaction.

Fourth, the impact of category transition on query reformulation pattern was analyzed. In categorical

transitions between two consecutive queries, the Parallel movement was the most frequently found for all types of categorical transitions, but other sub-facet distributions seem to be slightly different depending on categories. For instance, from A to A categorical transition, the Broader and the Replacement with synonym sub-facets are frequently found; from A to G transition, the Narrower sub-facet is frequently found; and from S to S and from G to G, the Broader and Narrower sub-facets are frequently found. When analyzing the exclusive transition patterns among the types of reformulation patterns, with Generalization and Specification, the Parallel movement is the dominant sub-facet followed by the Narrower; whereas with Abstraction, the Narrower, Replacement with synonym and Parallel movement are often used. Although slightly different reformulation patterns were found depending on the previous query category, the authors conclude that Parallel movement is dominantly used in all sorts of categorical transitions (approximately 50%), and Broader and Narrower movements follow the Parallel movement in frequency, generally speaking; however, it should be noticed the Broader and the Narrower sub-facets include not only hierarchically related concepts but also adding or dropping terms in queries. In other words, hierarchically related concepts are used much less frequently than parallel concepts. Synonymous relations are not frequently used compared to other relations, except in the case of the Abstract category, where synonyms are frequently used.

Since it is difficult for users to retrieve the most relevant image documents with their initial queries, in many cases, query reformulation and the resulting transitions to search stages are an inevitable component during the image search process. As Marchionini (2005) observed, users' analytical search strategies involve iterative query reformulations and transitions based on their examination of previous search results. Considering this, one method for identifying the users' analytical search strategies is to reveal query reformulation and transition patterns for image searches. Having revealed these reformations and patterns, the findings from this study could affect image retrieval systems design and implementation in various ways.

In terms of assisting with query reformulation and transitions, two overall approaches have been studied and implemented in practice: global and local approaches (Efthimiadis 1996; Manning, Paghaven, and Shutze 2008). The global approach, or collection-independent approach, attempts to provide query reformulation aids using lexical words databases, controlled vocabulary or thesauri which are independent of any specific collections. On the other hand, the local approach, or collection-dependent approach, is dependent on any given collections. Accordingly, the local approach is likely to use index terms, co-occurrence analysis, or clustering results based on given collections. Using this framework of implementation on image retrieval systems, the findings of this study could be used with both approaches by considering the followings.

First, although one critical finding was that the most frequent categorical transition occurs within the same category, there are also a substantial number

of queries crossing different categories when transiting. As a result, it might be beneficial if image information retrieval/indexing systems support both aspects: within the same category and crossing to other categories. This way, users might be exposed to the possibility to explore query terms from both categories. Therefore, it is necessary to provide some sort of mechanism or connections between different categories in addition to within the same categories. (e.g., for a user looking for a hurricane picture, 'Katrina' can be a good search term in addition to 'hurricanes' or 'storms')

Second, another potential mechanism for implementation in image retrieval systems is the use of Parallel movement, which corresponds to a RT (related term) relation. In other words, the authors emphasize that an image retrieval system should include an access mechanism which enables users to review related (or parallel) terms with preceding query terms. For this, controlled vocabulary systems should include terms in each category and identify relationships among terms. However, some controlled vocabulary systems simply focus on terms in the General/Nameable category and not on terms in the Abstract and the Specific categories. In addition, compared to NT, BT or USE/UF relations, a RT relation seems to depend on a thesaurus developer's intuitive judgment. Therefore, for providing terms in the Specific category and RT relations which are meaningful in a given domain or collection, it would be reasonable to utilize term co-occurrence analysis or a clustering technique - this can show related terms not only within the same category but also among different categories. Yoon (2009)'s study showed the possibility of utilizing co-occurrence analysis for identifying related terms within and across categories through user-supplied tags on images. In the case of the Abstract category, synonyms are frequently used. Abstract terms are generally subtle, delicate, and subjective; therefore, a controlled vocabulary system could be made more coherent by providing a cluster of abstract terms based on given collections.

6. Conclusion

There have been several studies attempting to understand the iterative process of information retrieval by understanding the query reformulation process (Rieh and Xie 2006; Lau and Horvitz 1999; Bruza and Dennis 1997) One of the significant keys to sound understanding of the iterative process of image searches is to identify the patterns and characteristics of query reformulation. Moreover, by analyzing the patterns and the characteristics in a series of query reformulations, this study examined image search queries by categorical transition in conjunction with query reformulation. Several conclusions were discovered. First, users consistently tend to use query terms in the Specific and the General/Nameable categories over several search stages. Second, regarding categorical transitions, users are more likely to stay with their initial categories during query reformulation. This trend of maintaining initial categories is more common with Specific and the General/Nameable

categories, rather than General/Abstract and General/ Subjective categories. Third, more detailed analysis on query reformulation identified significant findings: early search stages often incorporate narrowing terms while latter search stages tend to use broader terms; moreover, Parallel movement, moving toward a related term, is the most common form of query reformation. This is in contrast with movement that is hierarchical or synonymous. Fourth, the relationships between categorical transitions and reformulation patterns were examined. While there are slight differences in reformulation patterns depending on initial query categories and transition patterns, the Parallel movement consistently occurs most frequently, followed by the Narrower and Broader movements. An exception is with the case of transition from the Abstract category; in this instance the Replacement with synonyms movement is often found.

These findings of this study can be realized in various aspects of image retrieval systems, such as assisting users' analytical search experiences and building effective thesauri. Based on the findings of this study, search term suggestion over several search stages might be an ideal interface for users. When search terms are considered, they should not be limited in the same categories, since there are substantial proportions of search terms with cross-categories. Furthermore, this study should be followed by evaluation studies which adopt some findings of this study to an image retrieval system or access mechanism. Questions for implementation, can social tagging systems providing related term clusters be a useful source for assisting parallel movement? how should manual and automated thesauri be integrated? how do these approaches improve image search effectiveness? and so on., can be answered through these future evaluation studies.

Acknowledgement

The authors gratefully acknowledge Dr. Jim Jansen's generous sharing of the Excite data. We also thank master student assistants at Ewha, EunHye Lee and Jungah Namgoong, for their assistance in data analysis.

References

Batley, S. 1988. "Visual information retrieval: Browsing strategies in pictorial database." In *Proceedings of 12th International Online Information Meeting,* 373-381.

Bruza, P. D. and S. Dennis. 1997. "Query reformulation on the Internet: empirical data and the hyper-index search engine." In *Proceedings of the 5th RIAO Conference.*

Chen, H. 2001. "An analysis of image retrieval tasks in the field of art history." *Information Processing & Management,* 37(5): 701-720.

Choi, Y. and E. M. Rasmussen. 2003. "Searching

for images: The analysis of users' queries for image retrieval in American history." *Journal of the American Society for Information Science and Technology,* 54(6): 498-511.

Chung, E. and J. Yoon. 2009. "Categorical and specificity differences between user-supplied tags and search query terms for images. An analysis of *Flickr* tags and Web image search queries." *Information Research*, 14(3), paper 408. [cited 2009.8.15] <http://InformationR.net/ir/14-3/paper408.html>.

Collins, K. 1998. "Providing subject access to images: A study of user queries." *The American Archivist,* 61: 36-55.

Eakins, J. and M. Graham. 1999. *Content-Based Image Retrieval: a Report to the JISC Technology Applications Program.* Institute for Image Data Research, University of Northuinbria at Newcastle. [cited 2009.8.15] <http://www.jisc.ac.uk/uploaded_documents/jtap-039.doc>.

Eastman, C. and B. J. Jansen. 2003. "Coverage, relevance, and ranking: the impact of query operators on Web search engine results." *ACM Transactions on Information Systems,* 21(4): 383-411.

Efthimiadis, E. N. 1996. "Query expansion." *Annual Review of Information Systems and Technology,* 31: 121-187.

Enser, P. G. B., and C. G. McGregor. 1992. *Analysis of Visual Information Retrieval Queries: British Library R&D Report 6104.* London: British Library.

Enser, P. G. B., C. J. Sandom, J. S. Hare, and P. H. Lewis. 2007. "Facing the reality of semantic image retrieval." *Journal of Documentation,* 63(4): 465-481.

Goodrum, A., M. M. Bejune, and A. C. Siochi. 2003. "A state transition analysis of image search patterns on the Web." *Lecture Notes in Computer Science,* 2728: 281-290.

Goodrum, A and A. Spink. 2001. "Image searching on the Excite Web search engine." *Information Processing & Management,* 37: 295-311.

Griesdorf, H and B. O'Connor. 2002. "Modeling what users see when they look at images: a cognitive view point." *Journal of Documentation,* 58(1): 1-24

Hastings, S. K. 1995. "Query categories in a study of intellectual access to digitized art images." *Proceedings of the American Society for Information Science,* 32: 3-8.

Jansen, B. J. and A. Spink. 2005. "How are we searching the World Wide Web?: an analysis of nine search engine transaction logs." *Information Processing & Management,* 42(1): 248-263.

Jörgensen, C. 1998. "Attributes of images in describing tasks." *Information Processing & Management,* 34(2/3): 161-174.

Jörgensen, C. 2003. *Image Retrieval: Theory and Research.* Scarecrow, Lanham, Md.

Jörgensen, C. and O. Jörgensen. 2005. "Image querying by image professionals." *Journal of the American Society for Information Science and*

Technology, 56(12): 1346-1359.

Keister, L. A. 1994. "User types and queries: Impact on image access systems." In: R. Fidel, T.B. Hahn, E. M. Rasmussen, and P. J. Smith (eds.), *Challenges in Indexing Electronic Text and Images.*

Laine-Hernandez, M and S. Westman. 2006. "Image semantics in the description and categorization of journalistic photographs." *Proceedings of the American Society for Information Science and Technology,* 43: 1-25.

Lau, T. and E. Horvitz. 1999. "Patterns of search: Analyzing and modeling Web query refinement." In *Proceedings of the 7th International Conference on User Modelling,* 119-128.

Manning, C. D., P. Raghaven, and H. Schutze. 2008. *Introduction to Information Retrieval.* Cambridge: Cambridge University Press.

Marchionini, G. 2005. *Information Seeking in Electronic Environments.* Cambridge university press, Cambridge.

O'Connor, B., M. O'Connor, and J. Abbas. 1999. "User reactions as access mechanism: An exploration based on captions for images." *Journal of the American Society for Information Science,* 50(8): 681-697.

Ornager, S. 1997. "Image retrieval: Theoretical analysis and empirical user studies on accessing information in images." *Proceedings of the 60th Annual Meeting of the American Society for Information Science,* 34: 202-211.

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

Rorissa, A. and S. K. Hastings. 2004. "Free sorting of images: Attributes used for categorization." *Proceedings of the American Society for Information Science and Technology,* 41: 360-366.

Shatford, L. 1986. "Analyzing the subject of a picture: A theoretical approach." *Cataloguing & Classification Quarterly,* 6(3): 39-62.

Spink, A., B. J. Jansen, D. Wolfram, and T. Saracevic. 2002. "From E-sex to E-commerce: Web search changes." *IEEE Computer,* 35(3): 133-135.

Sutcliffe, A. G., M. Ennis, and S. J. Watkinson. 2000. "Empirical studies of end-user information searching." *Journal of the American Society for Information Science and Technology,* 51(13): 1211-1231.

Vakkari, P., M. Pennanen, and S. Serola. 2003. "Changes in search terms and tactics while writing a research proposal: A longitudinal case study." *Information Processing & Management,* 39: 445-463.

Yoon, J. 2009. "Towards a user-oriented thesaurus for non-domain-specific image collections." *Information Processing & Management,* 45(4): 452-468.

Appendix A. Transition frequency from column to row

Type			A		S		G		P	
			Freq.	%	Freq.	%	Freq.	%	Freq.	%
A	C	Broader	7	8.24	2	0.22	5	0.56		0.00
A	C	Parallel movement	10	11.76	5	0.55	11	1.22	2	22.22
A	C	Replacement with synonyms	7	8.24		0.00		0.00		0.00
A	C	Narrower	9	10.59		0.00	3	0.33		0.00
A	C	Repeated query	4	4.71		0.00		0.00		0.00
A	F	Operator usage	1	1.18		0.00		0.00		0.00
A	F	Term variation	14	16.47		0.00		0.00		0.00
		Sub-total	52	61.18	7	0.77	19	2.12	2	22.22
G	C	Broader	1	1.18	51	5.59	92	10.24	2	22.22
G	C	Parallel movement	9	10.59	66	7.23	233	25.95	2	22.22
G	C	Replacement with synonyms	2	2.35		0.00	41	4.57		0.00
G	C	Narrower	7	8.24	2	0.22	139	15.48	1	11.11
G	C	Repeated query	3	3.53	4	0.44	46	5.12		0.00
G	F	Operator usage		0.00	2	0.22	45	5.01		0.00
G	F	Term variation		0.00	2	0.22	137	15.26		0.00
		Sub-total	22	25.88	127	13.91	733	81.63	5	55.56
P	C	Broader	1	1.18		0.00		0.00		0.00
P	C	Parallel movement	2	2.35	1	0.11	3	0.33		0.00
P	C	Replacement with synonyms		0.00		0.00		0.00	1	11.11
P	C	Narrower		0.00		0.00		0.00		0.00
P	C	Repeated query		0.00		0.00		0.00		0.00
P	F	Operator usage		0.00		0.00		0.00		0.00
P	F	Term variation		0.00		0.00		0.00	1	11.11
		Sub-total	3	3.53	1	0.11	3	0.33	2	22.22
S	C	Broader		0.00	63	6.90		0.00		0.00
S	C	Parallel movement	5	5.88	324	35.49	60	6.68		0.00
S	C	Replacement with synonyms		0.00	22	2.41		0.00		0.00
S	C	Narrower	3	3.53	138	15.12	76	8.46		0.00
S	C	Repeated query		0.00	43	4.71	4	0.45		0.00
S	F	Operator usage		0.00	55	6.02	1	0.11		0.00
S	F	Term variation		0.00	133	14.57	2	0.22		0.00
		Sub-total	8	9.41	778	85.21	143	15.92	0	0.00
		Total	85	100.00	913	100.00	898	100.00	9	100.00

A: General/Abstract, S: Specific, G: General/Nameable, C: Content, F: Format