

# Query Modification Patterns and Concept Analysis of Web Image Queries

Youngok Choi

School of Library and Information Science,  
Catholic University of America  
choiy@cua.edu

## ABSTRACT

This study investigated query modification patterns and concepts used in query construction during users' searching images on the Web. It examined whether query modifications were related to different content collections, and analyzed what attributes were used to formulate a query in an interactive web searching process. Findings of the study show that query modification patterns were significantly associated with content collections. Terms related to format or specific objects to represent an image were found to be frequently used in reformulations and specializations. The findings suggest that system features for flexible query formulation and navigation would support users' image search process.

## Keywords

Image queries modification, concept analysis of queries

## INTRODUCTION

Ample studies demonstrate that query formulation is an essential activity and users change their queries frequently during a Web information search (Jansen, Booth, & Spink, 2009). This is apparent in text-based image search on the Web. Query modification is an important aspect in image searching on the Web as well (Goodrum and Spink, 2001; Jorgensen and Jorgensen, 2005). Research indicates that while searching images on the Web, users formulate an initial query as well as many consecutive queries on different content collections on the Web – e.g., general textual Web content search sites, images search sites, or a local content (Choi, 2010; Ozmutlu, Spink, & Ozmutlu, 2003; Spink, Jansen, & Narayan, 2007; Tjondronegoro, Spink & Jansen, 2009).

While previous studies have demonstrated that frequent query modification and browsing are critical in image search, little is known whether query modifications are

related to content collections such as general textual search engines, image search engines, and local content sites. Even less is known what conceptual attributes users incorporate in their query as a result of browsing search results and shaping their image needs accordingly. The purpose of this study, therefore, is to examine whether there is any relationship between query modification patterns and content collections of the Web where a query was issued, and to investigate what attributes of image search need users incorporate in formulating a query during query modification process.

## RELATED STUDIES

Previous studies reported image searches queries and search strategies on the Web by analyzing search logs from general or commercial search engines (Goodrum & Spink, 2001; Jansen, 2008; Jansen, Spink, & Pedersen, 2005; Jorgensen & Jorgensen, 2005; Ozmutlu, Spink, & Ozmutlu, 2003; Pu, 2005; Tjondronegoro, Spink, & Jansen, 2009). Their findings suggest that Web users type in short queries not only when searching for textual information, but also when searching for visual information. Query modification is important in image searching (Goodrum & Spink 2001; Jorgensen & Jorgensen 2005). Pu (2005) and Jansen (2008) found that users' image queries tended to focus on people and people-related queries. A few studies also analyzed multimedia searching characteristics on the Web among different content collections on search engines (Jansen et al., 2005; Spink & Jansen, 2006; Ozmutlu, Spink, & Ozmutlu, 2003; Tjondronegoro, Spink & Jansen, 2009). The findings show that image search is the most popular. Both the mean terms per query and the session lengths for image searching are larger than the other categories of multimedia searching. Westman, Lustila, and Oittinen (2008) found that 84.5% of all first queries were modified; purely content-based searches accounted only for 5% of queries; text queries included an average of 1.30 terms. Choi and Hsieh-Yee (2010) confirmed similar characteristics of image query formulation and modification strategies. Their study shows the most frequent strategy in modifying a query was replacing one term with another.

While general characteristics of how users formulate image queries and patterns of reformulation on the Web were reported in studies, unknown is any relationship between query modifications and content collections when users

explore different content collections on the Web and accordingly modify search queries. An identification of query modification patterns as well as conceptual attributes in query modification would enhance an understanding of user' image searching behavior and contribute to a system development for image searching.

**METHOD**

**Query Collection**

The study collected 978 image search queries from 29 college students' three separate search sessions for finding images on the Web (Choi, 2010). The total number of search sessions was 87. The participants conducted searches based on their own aptitudes and at their own pace with their own search goals during each search session separately. Therefore, each search session treated as an independent case. The search session was recorded by Camtasia v5.0 as screen capture. The screen capture was analyzed to collect the queries.

An average number of query iterations among 87 sessions was 11.24 with a maximum 39 and SD=8.73. An average term per query was 3.25 (SD=1.69). Three hundred sixty-five queries (37.32%) out of the total queries were issued in general search engines whereas about fifty percent (483 queries) were in image search engines and the rest were in a local site.

**Analysis of Query Modifications and Attributes**

Types of query modification were adapted from Lau and Horvitz (1999) – *New, Continuation, Generalization, Specialization, Reformulation, and Request*. The analysis of query modification was conducted by two coders independently categorizing each query refinement into one of 6 types. For inter-coder reliability, Cohen's kappa test was performed, and a test result shows that the agreement between two coders was outstanding at K=0.83.

The attributes of terms in image queries were analyzed with categories adopted from the study of Hollink, et al. (2004). Table 1 shows the categories and associated attributes of those.

Category	Attribute
Non-visual	Title/phrase; Creator/author; Date; Collection; Others
Perceptual	Color; Shape; Format/genre; Texture
Conceptual	Object, Person, Organization/brand name; Concept/theme; Event/activities/action; Place; Time period

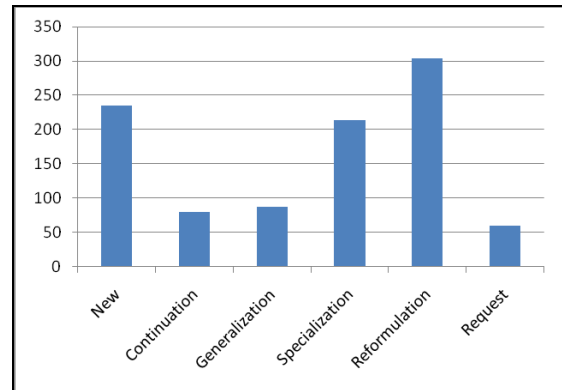
**Table 1. Classification of attributes in image descriptions**

**RESULTS**

*Query Modification Patterns*

The most frequently used modification type is Reformulation (31.08%) followed by New (24.03%), Specialization (21.78%), Generalization (8.90%),

Continuation (8.08%), and Request (6.13%) (Figure 1). It is interesting to note that about 14% of queries (combined from Continuation and Request) were repeated queries from previous queries.



**Figure 1: Frequency of query modification**

Modification	Content collection		
	General Web	Image	Local
New	108 (29.6%)	100 (20.7%)	27 (20.8%)
Continuation	21 (5.8%)	33 (6.89%)	25 (12.9%)
Generalization	28 (7.7%)	40 (8.3%)	19 (14.6%)
Specialization	74 (20.3%)	123 (25.5%)	16 (12.3%)
Reformulation	112 (30.7%)	172 (35.6%)	20 (15.4%)
Request	22 (6.0%)	15 (3.1%)	23 (17.7%)
Total	365 (37.3%)	483 (49.4%)	130 (13.3%)

**Table 2. Count of query modification within content collection**

*Association of query modification patterns with content collection*

Query modification patterns were analyzed with content collections where queries were issued with a chi-square test to determine the relationship. Results of the analysis indicated that the participants were significantly more likely to construct new queries on general search engine sites while they tended to modify queries with reformulations and specializations on image search engine sites (Value= 93.27, p=0.00) (See Table 2).

*Conceptual attributes in image queries*

An analysis of attributes in queries reveals that new queries include more contextual attributes like bibliographical information associated with images that users looked for. On the other hand, specializations and reformulations of queries include more perceptual attributes related to a format like picture, image, print etc. and represented

objects, person, and specific organization or brand names (See Table 3).

Attribute	N	C	G	S	RF	RQ	Total
Non-visual	58	21	18	53	69	24	243
Perceptual	48	18	31	137	138	16	388
Object	46	23	21	75	118	6	289
Concept/theme	20	9	0	13	23	2	67
Person	47	13	9	47	69	8	193
Organization	43	16	39	86	120	16	320
Event/activities	28	10	8	39	49	7	141
Place	27	13	13	70	69	12	204
Time	8	9	11	33	52	6	119

Key: N- New; C- continuation; G- generalization; S - specialization; RF- reformulation; RQ- request

**Table 3. Frequency of attributes in image queries**

### CONCLUSION

This study confirms a frequent query modification behavior of users during image search on the Web. Findings of the study suggest that users tend to form a general or broad query as an original search description, which suggests that users experience difficulties in forming a query in words. While browsing images search results from an initial search query, users were able to further articulate their search query with format-related terms and descriptions of more specific objects depicted in an image. It seems clear that system interface features are necessary for query formulation and navigation support. For example, a faceted browsing or a query formulation feature based on facets would be useful to allow users to describe image needs in various levels.

### REFERENCES

- Choi, Y. (2010). Investigating variation in querying behavior for image searches on the Web. In: Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem. 47: 1-10.
- Choi, Y., & Hsieh-Yee, I. (2010). Finding images on an OPAC: Analysis of user queries, subject headings, and description notes. *Canadian Journal of Information and Library Science*, 34(3), 271-296.
- Goodrum, A., & Spink, A. (2001). Image searching on the EXCITE Web search engine. *Information Processing and Management*, 37, 295-312.
- Hollink, L., Schreiber, A.Th., Wielinga, B.J., & Worring, M. (2004). Classification of user image descriptions. *International Journal of Human-Computer Studies*. 61 (5), 601-626.
- Jansen, B. J. (2008). Searching for digital images on the web. *Journal of Documentation*. 64(1), 81-101.
- Jansen, B. J., Booth, D. L., & Spink, A. (2009). Patterns of query modification during Web searching. *Journal of the American Society for Information Science and Technology*. 60(7), 1358-1371.
- Jansen, B. J., Spink, A. & Pedersen, J. (2005). The effect of specialized multimedia searching on web searching. *Journal of Web Engineering*. 3(3/4), 182-199.
- Jorgensen, C., & Jorgensen, P. (2005). Image querying by image professionals. *Journal of the American Society for Information Science and Technology*, 56(12), 1346-1359.
- Lau, T., & Horvitz, E. (1999). Patterns of search: Analyzing and modeling web query refinement. In: Proceedings of the Seventh International Conference on User Modeling, Banff, Canada, June 1999. New York: Springer Wien, pp.1190-128.
- Ozmutlu, S., Spink, A., & Ozmutlu, H. C. (2003). Multimedia web searching trends: 1997-2001. *Information Processing and Management*, 39(4), 611-621.
- Pu, H. (2005). A comparative analysis of Web image and textual queries. *Online Information Review*, 29(5), 457-467.
- Spink, A., & Jansen, B. J. (2006). Searching multimedia federated content web collections, *Online Journal Review*, 30(5), 485-495.
- Jansen, B., Spink, A., & Narayan, B. (2007). Query modifications patterns during Web searching. In *Fourth International Conference on Information Technology, 2007. ITNG '07*, 2-4 April 2007, Las Vegas, Nevada. pp. 439 - 444
- Tjondronegoro, D. W., Spink, A., & Jansen, B.J. (2009). A study and comparison of multimedia Web searching: 1997-2006. *Journal of the American Society for Information Science and Technology*, 60(9), 1756-1768.
- Westman, S., Lustila, A., & Oittinen, P. (2008). Search strategies in multimodal image retrieval. In *Proceedings of the Second international Symposium on information interaction in Context* (London, United Kingdom, October 14 - 17, 2008). P. Borlund, J. W. Schneider, M. Lalmas, A. Tombros, J. Feather, D. Kelly, and A. P. de Vries, Eds. IliX '08, vol. 348. ACM, New York, NY, 13-20.