Tactics for Information Search in a Public and an Academic Library Catalog with Faceted Interfaces

Xi Niu
University of North Carolina at Chapel Hill
CB#3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
xiniu@email.unc.edu

Bradley M. Hemminger
University of North Carolina at Chapel Hill
CB#3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
bmh@ils.unc.edu

ABSTRACT

This study examined a large number of searches conducted when the users are interacting with two Endeca-based faceted library catalogs (University of North Carolina at Chapel Hill [UNC] Library catalog and Phoenix Public Library [PPL] catalog). The goal is to investigate people's search tactics with the faceted catalogs in an academic library and a public library environment. Two large data sets (with 504,142 logs for 40 days, and 1,010,239 logs for 60 days respectively) are analyzed. State transition analysis and maximal repeating patterns [MRP] analysis are conducted to identify the search tactics and action patterns. The results have both implications for designers in developing faceted library catalogs, and methodological contributions to the empirical research on faceted search systems.

1. INTRODUCTION

In recent years, faceted navigation has grown to be a well-accepted approach. It has been applied as a standard technique on commercial Web sites for a number of years (Breeding, 2007). Since the adoption of the faceted search in some university libraries, it has become adopted by to many academic and public libraries. Many librarians and IT professionals take it for granted that the categorized labels will help people find what they want. However, little has been known about how people are searching through these faceted systems. While significant work has examined the usability of facets (for example, Antelman et al., 2006; Lown, 2008), few investigated the sequences of moves, i.e., search tactics, made by searchers in order to understand the cognitive process when they are interacting with a faceted search system.

The current paper tackles this problem by examining the tactics of users searching two faceted library catalogs—University of North Carolina at Chapel Hill (UNC) Library catalog and Phoenix Public Library (PPL) catalog, focusing on the comparison between the academic library and the public library.

The two library catalogs, catering to different audiences, represent faceted OPACs in leading research universities and leading public libraries respectively. UNC library catalog (http://search.lib.unc.edu/), implemented faceted search on top of their original catalog system in 2008 to enhance the search ability. PPL's catalog (http://www.phoenixpubliclibrary.org/default.jsp), , has excellent facet implementation added on its traditional ILS system. The commonality of the two catalogs is both are using the commercial Endeca search engine to implement faceted navigation. However, for UNC catalog, facets are primarily serving the purpose of narrowing down a search, rather than browsing the collection. In contrast, for PPL catalog, facets are supporting browsing as well as refining a search.

2. BACKGROUND

Bates in her paper (1979) first introduces the concept of the search tactic as moves to further a search in information seeking context, as compared to the search strategy which focuses on the plan of the whole search. Since then, a few studies have examined search tactics adopted by users in various search systems. According to Wildemuth (2004), a search tactic is a set of search moves that are temporally and semantically related in order to accomplish a search goal. In her study, she investigated the effects of domain knowledge on search tactic formulation, and found that the search tactics changed over time as the students' domain knowledge changed. Qiu (1993) used a regression model to examine the fit of Markov models to search tactics in a hypertext system. It was found that a second-order Markov model best fit the data. Chen and Cooper (2002) studied users' moves of a library catalog through semi-Markov chains and identified six different usage groups with distinct patterns of system usage. Kiestra, Stokmans, and Kamphuis (1994) derived that the minimum "meaningful unit" of search behavior consisted of three consecutive actions, therefore analyzed these fragments of move sequence, and discovered 65 commonly used patterns. Jansen, Spink, and Saraevic (2000) conducted transaction log analysis for over 50,000 Web search engine queries and found that only 22% queries were modifications of a previous query. They also analyzed search moves of 191 search sessions and concluded that the most common session was "a unique query followed by a request to view the next page of results."

Although the analyses were similar in each of these studies, it is difficult to draw general conclusions due to different search move definitions, and different search systems. In addition, none of this study investigated the search tactics through a faceted search system—almost a standard feature for today's commercial search engines.

3. RESEARCH PROBLEM

This study addresses two research questions. First, it examines the tactics actually used by users to understand how they formulate and reformulate their cognitive process and search strategies when they are searching the faceted library catalogs. Second, it focuses on differences in these tactics that might be attributable to differences between an academic and a public library setting.

4. METHODS

4.1 Data Extracting and Processing

The transaction log datasets in this study are from UNC OPAC and PPL OPAC servers. Description of the two datasets is summarized in Table 1. The data are extracted and coded using Perl scripts and MySQL database. Details about analyzing the data have been elaborated elsewhere (Niu, Lown, & Hemminger, 2009; Lown, 2008). Table 2 below summarizes the available action codes and their descriptions.

Table 1. Description of transaction log datasets

Log Dataset	Time Frame	Size	Available Fields
UNC library Apache server logs	40 days 1/1/2010—2/9/2010	491M raw data 504,142 useful records	IP address /data,time/URL/ reference URL/user agent
PPL Apache server logs	2 month 3/1/2010—4/30/2010	371M raw data 1,010,239 useful records	IP address /data,time/URL/ reference URL/user agent

4.2 State Transition Matrix

A Markov model is a stochastic process with the Markov property which means that the description of the present state fully captures all the information that could influence the future evolution of the process. Future states will be reached through a probabilistic process instead of a deterministic one. The order of Markov models means how many previous states (including the current state) influence the choice of the next state probabilistically. The simplest form of a Markov model is called a zero-order model. It is simply the frequency with which each state occurred. The first-order Markov model, also called a state transition matrix, reports the probability of the transition from all the possible current states to all the possible future states. First-order Markov models are the types of models most frequently found in the ILS literature (Wildemuth, 2009). In this paper, first-order Markov model is adopted to report the frequency of each transition. A graphical representation of the most frequent state transitions is to be created by linking together these individual transitions.

4.3 Maximal Repeating Pattern (MRP)

Previous literature indicates that people's information behavior varies greatly from one person to another. It is helpful to find patterns that are frequently adopted by searchers. Siochi and Ehrich's (1991) algorithm for identifying maximal repeating patterns (MRPs) among sequences of behavior is applied to serve this purpose. They defined an MRP as "a repeating pattern that is as long as possible, or is an independently occurring substring of a longer pattern" (Siochi & Ehrich, 1991). Thus, the algorithm systematically identifies those sequences of events that occur repeatedly within the data set. In this study, each of the two data sets is analyzed with the MRP software. Any frequently occurring fragment strings are selected for further analysis. Selected MRPs are then grouped into families of patterns. Each family could be summarized using a regular expression and exemplified with two most frequent sequences.

5. RESULTS

The results from each library are reported separately and include both the state transition analysis and maximal repeating pattern (MRP) analysis.

5.1 Results From UNC Library Catalog

After processing the UNC data, 1243 search sessions, composed of 6416 moves, are identified. The most frequent state transitions among those moves are illustrated in Figure 1.

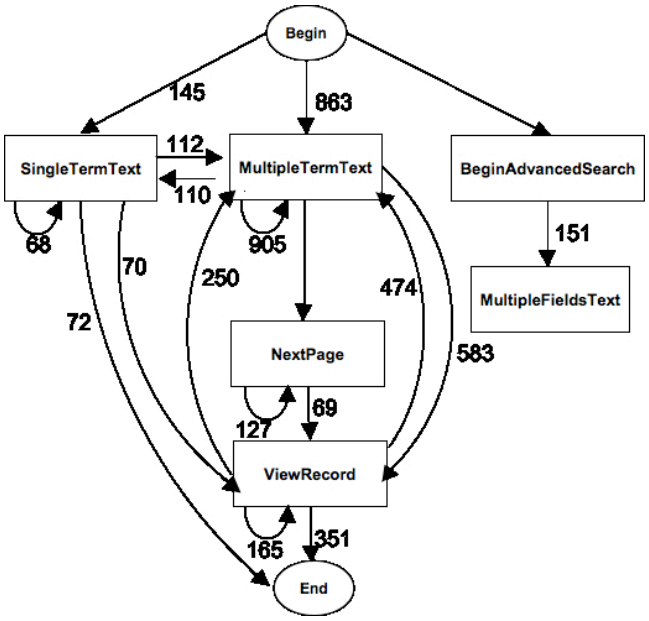

Figure 1. State transition network of search tactics for UNC (1243 search sessions, 6416moves), the numbers only include the transitions > 1% of total.

As shown in Figure 1, the most common transitions during a search are *MultipleTermText* to *MultipleTermText*, *Begin* to *MultipleTermText*, and *MultipleTermText* to *ViewRecord.* These transitions account for over 1/3 of the total transitions. It suggests that a number of searchers of the UNC library catalog adopt rather simple strategies, searching similarly with this faceted catalog as they did in a traditional catalog. "Advanced" features, like facet operations, sorting the result, are not as often used as expected.

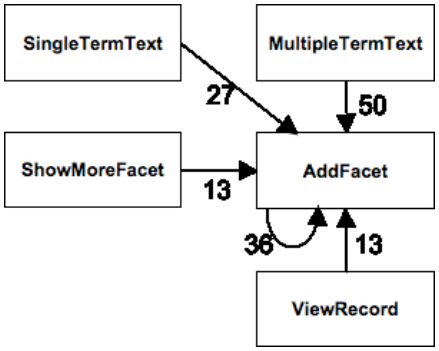

Figure 2. state transition network of facet operations for UNC (only illustrates the top 5 transitions).

Table 2 Action codes

UNC action code	Description	PPL action code
SingleTermText	Submit a single word query in the text search box	SingleTermText
MultipleTermText	Submit a multiple-word query in the text search box	MultipleTermText
SwitchTextField	Switch the search field (e.g. author, title) using the same text query as before	SwitchTextField
MultipleFieldsText	Submit a query in a multiple search boxes in different fields, typically in the advanced search page	MultipleFieldsText
EmptyTextSearch	Submit a empty query in the search box	EmptyTextSearch
BooleanTextSearch	Submit a query in the Boolean search box	
AddFacet	Click a facet value to incorporate it to the current search. For UNC library, AddFacet is only for refining the search. For PPL, could be either refining the search or browsing the collection	Refine Browse
RemoveFacet	Click "x" next to the already chosen facet value to take off it	RemoveFacet
ShowMoreFacet	Click "more" under a facet group to show more values	ShowMoreFacet
RefineYears	Under "Publication Year" facet group, manually type the starting and ending years and submit	N/A
OpenFacet	Click the "+" next to a facet group to show the values of the facet	N/A
CloseFacet	Click the "-" next to a facet group to hide the values of the facet	N/A
BeginSimpleTextSearch	Open the "Search" tab to begin a simple text search	N/A
BeginAdvancedSearch	Open the "Advanced Search" tab to begin a multiple fields text search	N/A
BeginCallNumberSearch	Open the "Browse by Call Number" tab to begin a call number search	N/A
BeginNewTitlesSearch	Open the "Browse New Titles" tab to begin a new title search	N/A
N/A	Open the "books" tab to search in the book collection	BookSearch
N/A	Open the "movies" tab to search in the movie collection	MoviesSearch
N/A	Open the "music" tab to search in the music collection	MusicSearch
N/A	Open the "downloadables" tab to search in the media collection	DownloadablesSearch
N/A	Open the "magazines & newspapers" tab to search in the book collection	MagNewsSearch
ViewRecord	Click through a record link to view details about the record	ViewRecord
NextPage	Click a page number or next button in the result page to view what in next page	NextPage
SortResult	Choose the options (relevance, publication year …) from the "Sort by" drop down menus to sort the result list	SortResult
N/A		SaveItem
FollowupAction	Click a link provided within a record to find the related records	FollowupAction
Refresh	Click refresh button of the browser	Refresh

Figure 2 above represents the state transition network of most frequently occurring facet operations for the UNC catalog. The degree of involvement for each action in this network indicates that the most used facet operation is *AddFacet.* Facets are typically added following a text search with either single term or multi term query. Most likely, these facets are serving the purpose of refining the previous text search. Sometimes, *AddFacet* is followed by another *AddFacet*. Compare to Figure 1, the numbers are much smaller, suggesting less usage in facets.

Focusing on the search sessions having at least one facet operation which are called faceted search sessions, MRP analysis was conducted The result indicates that there are three distinct families of patterns for the tactics frequently adopted by the UNC catalog searchers.

Table 3 Maximal repeating patterns for UNC catalog

Family of Pattern	Examples
F+V*	AddFacet>AddFacet>ViewRecord ShowMoreFacet>AddFaccet>ViewRecord
T+F+V*	MultipleTermText>MultipleTermText>ShowMoreFacet>AddFacet MultipleTermText>AddFacet>ViewRecord>ViewRecord
T*F+V*T+F*V*	AddFacet>ViewRecord>MultipleTermText>AddFacet MultipleTermText>AddFacet>SingleTermText

F denotes facet operations; V denotes viewing an item; T denotes text search
+ denotes occurring one or more times; * denotes occurring any time

The family patterns are summarized using regular expressions shown in the left column of Table 3. The examples in the right column are the two most frequent individual patterns of that family. Even for the most frequent individual patterns, they only account for approximately 1% of all the possible patterns, which are over 1 thousand identified through MRP analysis. This indicates that search tactics are rather idiosyncratic with much difference that can be attributed to individuals.

5.2 Results From Phoenix Public library catalog

428 search sessions with 6987 moves are extracted from the PPL data. The average number of actions per search triples that of UNC (1243 sessions with 6416 moves). It's probably because of the public library setting or the different implementation of the library catalog. The frequent state transitions among those moves are demonstrated in Figure 3.

The most common transitions happen between *MultipleTermText* and *ViewRecord*, or self-repetitions of *ViewRecord. Begin* to *MultipleTermText* is relatively less than that of the UNC data, due to more choices offered by Phoenix Public Library, rather than just text querying to start a search. In addition to typing a text query, users may browse the collection or select an appropriate tab to begin a search. One remarkable thing is that *Refine*, as one way of adding a facet, appears in the network as one of the frequent states. It implies a boost of facet usage compared to the UNC catalog. *Refine* is typically followed by *MultipleTermText*, and leads to either *ViewRecord* or a repetition of itself. Focusing on the transitions among facet operations (Figure 4), in addition to *Refine*, *ShowMoreFacet* and *Browse* are also the top used facet moves. They are most likely to follow or to be followed by an identical move.

The result from MRP analysis suggests 4 families of patterns, as illustrated in Table 4.

Table 4 Maximal repeating patterns for PPL catalog

Family of Pattern	Examples
F+V*	Refine>Refine>ViewRecord ShowMoreFacet>Refine
T+F+V*	MultipleTermText>ShowMoreFacet>Refine MultipleTermText>Refine>ViewRecord>ViewRecord
T*F+V*T+F*V*	Refine>ViewRecord>MultipleTermText>Refine MultipleTermText>Refine>SingleTermText
B+N*V*	Browse>Browse>Browse>Browse>ViewRecord Browse>Browse>NextPage>NextPage

F denotes facet operations; V denotes viewing an item; T denotes text search; B denotes browsing; N denotes viewing next page
+ denotes occurring one or more times; * denotes occurring any time

Of these four families of patterns, the first three are the same with those of the UNC data. The last one is the unique pattern for the PPL data. The PPL searchers take advantage of the offered browsing features and tend to continue browsing before getting to a reasonable result set.

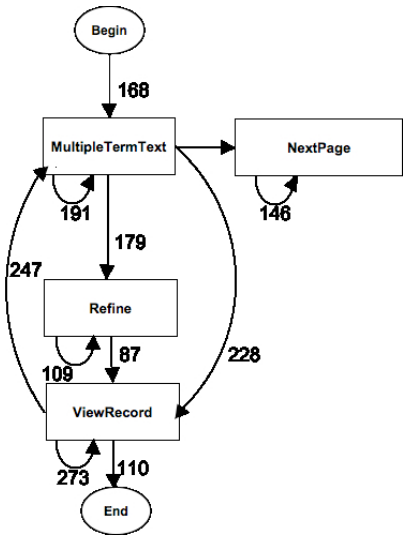

Figure 3. State transition network of search tactics for PPL (428 search sessions, 6987 moves), the numbers only include the transitions > 1% of total.

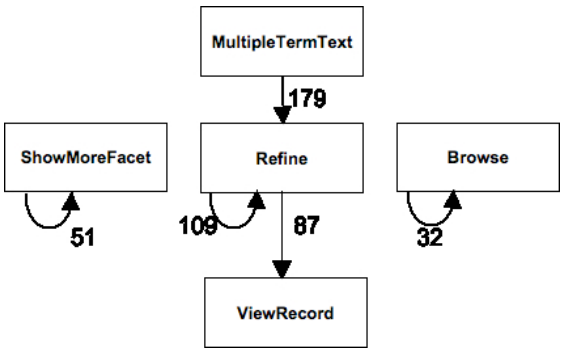

Figure 4. state transition network of facet operations for PPL (only illustrates the top 5 transitions).

6. CONCLUSION

A catalog search is made up of a sequence of temporally and semantically related moves as a search tactic. This study examines a large number of searches conducted when the searchers are interacting with two faceted library catalogs. An analysis of the search moves indicates that the most common tactics across the two environments are text querying followed by viewing result items. This is similar to what the searchers have done through the traditional OPAC without facet features. MRP analysis suggests that search tactics are rather idiosyncratic and users do not have much in common in terms of their search moves. To some extend, a number of searchers are conservative in using facets. They just adopt simple and naïve search strategies. It is unlikely that facets are useful in all types of search situations and for all types of tasks. Facets used by people are primarily for two purposes: refining a search and browsing the collection. Facet usage in PPL is much higher than that of UNC. It might be the better support of facet browsing of PPL that cause the boost of its facet usage. Another striking thing about search tactics is that people are likely to repeat what they did in the previous run. For example, adding a facet is high likely to be followed by adding another facet. Browsing is most likely to repeat during the search process.

REFERENCES

[1] Antelman, K., et al. (2006) Toward a Twenty-First Century Library Catalog. Information Technology and Libraries. 25(3) 128-139.

[2] Bates, M. J. (1979). Information search tactics. Journal of the American Society for Information Science, 30(4), 205-214.

[3] Breeding, M. (2007). Introduction [next-generation catalogs] Library Technology Reports 43(4), 5-14.

[4] Chen, H. M., & Cooper, M. D. (2002). Stochastic modeling of usage patterns in a web-based information system. Journal of the American Society for Information Science and Technology, 53(7), 536-548.

[5] 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 & Management, 36(2), 207–228.

[6] Kiestra, M.D., Stokmans, M.J.W., & Kamphuis, J. (1994). End-users searching the online catalogue: The influence of domain and systemknowledge on search patterns. The Electronic Library, 12(6), 335–343.

[7] Lown, C. (2008). A transaction log analysis of NCSU's faceted navigation OPAC. Master's paper.

[8] Niu, X., Lown, C., & Hemminger, B. M. (2009). Log based analysis of how faceted and text based search interact in a library catalog interface. In Proceedings of Third Workshop on Human-Computer Interaction and Information Retrieval.

[9] Qiu, L. (1993). Markov models of search state patterns in a hypertext information retrieval system. Journal of the American Society for Information Science, 44(7)

[10] Siochi, A.C., & Ehrich, R.W. (1991). Computer analysis of user interfaces based on repetition in transcripts of user sessions. ACM Transaction on Information Systems, 9(4), 309 – 335.

[11] Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. Journal of the American Society for Information Science and Technology, 55(3), 246-258.

[12] Wildemuth, B. (2009). Applications of social research methods to questions in information and library science. Santa Barbara, CA :Libraries Unlimited.