Understanding Book Search Behavior on the Web

Jin Young Kim
Dept. of Computer Science
University of Massachusetts
Amherst, MA 01003
jykim@cs.umass.edu

Henry Feild
Dept. of Computer Science
University of Massachusetts
Amherst, MA 01003
hfeild@cs.umass.edu

Marc-Allen Cartright
Dept. of Computer Science
University of Massachusetts
Amherst, MA 01003
irmarc@cs.umass.edu

ABSTRACT

With the increased availability of e-books and digitized book collections, more users are searching the web for information about books. There are many online digital libraries containing book, author and subject data, which are accessed via internal search services as well as external web sites, such as Google. Although this is a common yet complex information-seeking behavior involving multiple search systems with different characteristics, little is known about how users find information in this scenario.

In this work, we analyze web-based book search behavior using three months of logs from the Open Library, a globally accessible digital library. Our study encompasses the user behavior on web search engines and the digital library, unlike previous work which focused on institution-level digital libraries. Among our findings are (1) query characteristics and session-level behaviors are drastically different between internal and external searchers; (2) the field usage is different based on the modes of interaction—keyword search, advanced search interface and faceted filtering; (3) users go through with more iterations of faceted filtering than query reformulation. To facilitate future research on book search, we also create a book search test collection based on the log data. We then perform an evaluation of several retrieval methods, finding that field-based retrieval models have advantages over document-based models.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*search process*

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

Query log analysis, user modeling, book search

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.

1. INTRODUCTION

Books are the most widely used archival form of knowledge and entertainment, and searching over them and related information like authors, subjects, and publishers is a popular activity on the web. There are many online digital libraries that collect, organize, and offer information about books and if a user wants to access this content, they can do it in several ways. One way is to visit the digital library website directly and browse or search using the internal search interface. Alternatively, provided the digital library's content can be crawled, then web search engines (WSEs) are another avenue for finding book-related content. We consider web-based book search that pertains to digital libraries that are searchable by WSEs.

There are several characteristics that make book search on the web particularly interesting from an information retrieval perspective. First, users can use a domain-specific search engine (i.e., the internal digital library search service), a generic WSE, or a combination of both to find relevant information. Second, there is rich metadata associated with books, authors, and other book-related entities, which have implications for the design of a ranking algorithm that uses such structural evidence. Third, the existence of metadata also allows the implementation of interaction methods such as field operators, advanced search interfaces, and faceted filtering for book search, all of which are offered by most digital libraries.

Given these characteristics, there are many complexities regarding users' book search behavior on the web. While previous work [4, 12] investigated user behavior within a digital library or within a web search engine, we know of no work that studies the behavior of users searching for books across both WSEs and digital libraries. Scale is another aspect by which previous analyses of book search behavior are limited, in that most work studied institution-level digital libraries with limited user bases.

To better understand the complex task of web-based book search, we present a large-scale analysis of book search behavior across WSEs and the Open Library,[1] a freely available digital library that serves users from across the world. The log consists of 18 million records spanning 8 million user sessions and the last three months of 2010.

In our analysis, we focus on internal and external search behavior separately. Our decision is motivated by the extreme differences in behavior we found between them, as well as the number of variables unique to each, such as the search interface. We also study the use of various field-based inter-

[1] http://www.openlibrary.org

action methods within the library website, finding that the field usage is different based on the modes of interaction—keyword search, advanced search, and faceted filtering.

Based on these findings, we provide recommendations that can significantly influence the design of a book search system from both the interface and ranking perspectives. While we focus on book search behavior in this work, our findings are generalizable to most web catalogs that rely on WSEs as a primary entry point, and provide various interaction methods based on rich metadata.

To evaluate retrieval models for book search in a realistic setting, we then create a test collection based on a year-long sample of Open Library log data. We perform a comprehensive evaluation of relevant retrieval methods, and find that the effective use of field structure is critical for retrieval performance. We plan to release the data set for future research on book search.

The rest of this paper is organized as follows. We provide an overview of related work in the next section. Then we introduce the search environment and data set methods we used in Section 3. We then present the results of our analysis in Section 4. Finally, the process of test collection generation and the results of retrieval experiments are presented in Section 5.

2. RELATED WORK

There are three primary areas of related work: user behavior analysis on digital libraries, search behavior analysis on the web and other domains, and book retrieval development.

Analyzing user behavior over digital libraries is an important problem in the library and information science community. Blummer et al. [2] provide a detailed survey. Jones et al. [5] analyzed a digital library log, focusing on the use of operators and search options in queries. They also examined patterns in query construction and refinement. Kules et al. [9] performed a task-based evaluation of faceted search interfaces in a library catalog using eye-tracking devices, finding that faceted search plays a major role in exploratory search and affects searcher tactics, including the reduction of query reformulations.

More recently, Niu et al. [12] investigated people's faceted search behavior over two library catalogs, concluding that users incorporate faceted filtering as part of their information-seeking process, and that the use of facets depend on how facets are ordered, and whether faceted filtering is allowed before entering a query. Guo et al. [4] proposed a Markovian user interaction model, which we adapt to model user session behavior.

Domain-specific search behavior has been studied in the context of searching for people on the web [16], biomedical information seeking [10], blog search [11], and so on. Weerkamp et al. [16] analyzed a large-scale log of people search on the web, providing detailed statistics at the level of queries, sessions and users. Lin et al. [10] examined the relationship between initial search results and the overall utility of interaction, finding that similarity-based browsing tool helps in the case of poor initial retrieval results. Mishne et al. [11] focused on the analysis of blog search behavior, finding that blog searches typically include named entities of interests.

The INEX Book Track workshop[2] has held a number of book retrieval tasks in recent years, with the most relevant to this work being the 2010 Best Books to Reference task [6] and the 2011 Social Search for Best Books task [7]. The Social Search task dataset includes 2.8 million book metadata records from Amazon Books, augmented with user generated information from LibraryThing.[3] Search topics were extracted from LibraryThing forums in which users ask for recommendations. In contrast, our test collection aims to evaluate retrieval models using actual user queries and judgments generated from click data.

3. ENVIRONMENT AND DATA SET

In this section, we describe the search environment and the data set we used. We first describe the Open Library—the environment in which we observe user behavior—followed by several definitions used throughout this paper. We then describe the data set collected from the Open Library users, how we prepared the data, and the limitations of our analysis.

3.1 Open Library

The Open Library (OL) is a community-curated, open access digital library. An Internet Archive[4] initiative, the goal of the OL is to have one web page per book ever published along with the corresponding subject and author data.

The OL provides several features for searching and exploring the stored book records. An example book search result page is shown in Figure 1. At the top right of every page is a search box with an advanced search interface that appears on page at the click of a link

The search box, labeled (1) in the figure, supports several search operators, most notably quotes and field operators.[5] While WSEs also provide several operators including field operators, the OL supports field operators based on book metadata (e.g., *title:*, *author:*), unlike WSEs, which provide domain-independent operators (e.g., *site:*, *inurl:*).

The advanced search interface, labeled (2), includes: title, author, ISBN, subject, place, person, and publisher. The interface also provides an option to perform a full-text search. This search box and associated advanced search interface only searches over books, although users can also access author names linked in the results list. As with the search operators, the OL's advanced search interface features are domain specific, unlike the advanced options of general WSEs.

At the top of the search page is a set of sorting options, labeled (3): sort by relevance (the default), the number of editions a book has, and publication date (ascending and descending order). When a book search is conducted, a set of filters are presented on the right side of the screen, labeled (4) in the figure.

Filters available in the OL include: e-books versus non-e-books, authors, subjects, people, places, times, first published dates, publishers, and languages. These narrow the results by filtering them based on the selected criteria. For example, clicking the "History" facet under the subjects filter will restrict the results list to only books that fall into the "History" category. Note that filters can only be applied

[2] `https://inex.mmci.uni-saarland.de/tracks/books/`
[3] `http://www.librarything.com/`
[4] `http://www.archive.org`
[5] `http://openlibrary.org/search/howto`

Figure 1: An example book search. The red circles are annotations highlight, in order, (1) the book search box, (2) the advanced book search interface (displays when "More search options" is clicked), (3) the available sorting options, (4) the filters available for book search, (5) the links available for each result, and (6) additional verticals.

after an initial book search by clicking one of the links on the right side of the results page.

A book search retrieves a list of *work* pages, where a work is a collection of book editions. Each work in the results (an example is labeled (5) in the figure) includes a link to that work's page and a link to the author. If there is an electronic version of the book available, a link titled "Read" appears to the left of the result. A work page, e.g., "Moby Dick", consists of a description of the book and a list of the known editions. If any of the editions are available in full text, as e-books, or via library loan, the appropriate links are displayed.

On the top left of every page is a set of verticals, labeled (6): subjects, authors, and lists. Subjects contain sets of books with shared heading information and are broken into several categories: subjects, places, people, and times. Authors are self explanatory. Lists are user-generated and contain a set of books.

3.2 Definitions

We will use the following definitions throughout this work:
Object types: books, works, authors, and subjects. The OL differentiates between a *work* and a *book*. A *book* represents a specific edition. A *work* refers to all of the editions of a given book. The OL's book search actually searches over works. The page for a given work shows a listing of all the editions for that work. A *subject*, as described in Section 3.1, is a collection of works that share common heading information. In addition, *authors* can be searched and are linked with both books and works.
SERP-clicks. A *SERP-click*, short for search engine results page click, signifies that a user arrived at the current page via a link on a search result page, either from an internal or external search engine.
User. We consider the combination of an IP address and the access date (e.g., *111.111.111.111/December 15, 2010*) to identify a user. Due to the instability of IP address assignments, we do not group actions from a single IP address over more than a 24 hour period.
Session. A session contains all actions from a user such that the duration between any two adjacent events is no greater than 30 minutes.
External vs. Internal session. An *external* session is a session in which the first event is an external query. Any session that is not an external session is internal.
Query. We use the term *query* to refer to a set of user inputs to search for information. A query issued to a WSE is a *keyword query* that can optionally include search operators. A query in the OL can be any combination of three elements described in Section 3.1: a *keyword query*, which refers to the text a user entered into the search box including search operators; a *structured query*, which refers to user input to one or more advanced search fields; and lastly, a set of *filters*.
External vs. Internal query. A search that is initiated from outside of the OL domain, e.g., from the Google search engine, is called an *external query*. A search that was initiated from an OL search interface is called an *internal search*.
Search operator. A search operator is any portion of a keyword query that invokes special instructions to the retrieval system. For example, using quotes around phrases or specifying an advanced search feature like *title:moby dick*.

3.3 Data Set & Preparation

The analysis presented here is based on the Apache server logs from the OL web servers from October through December 2010. The log contains roughly 18 million records. The log format[6] includes, among other information, the URL re-

[6]The full Apache log format is a slight modification of the NCSA extended/combined log format, the description of which can be found here: `http://httpd.apache.org/docs/2.2/mod/mod_log_config.html`

quested, the time of a request, and the referrer URL (when available).

In preparing the data for analysis, we converted the chronologically ordered Apache logs into a more readable user session format. User identifiers were created by concatenating the IP address and the date (day, month, and year)—as mentioned before, this helps avoid the ambiguity of associating the same user with an IP address over a prolonged period of time, since the IP address may have been reassigned. We demarcate session breaks with 30 minutes of inactivity between consecutive events [15]. All events were grouped by user and session.

4. ANALYSIS OF SEARCH BEHAVIOR

In this section, we present our analysis of book search behavior. We first introduce how we structure our analysis exploring behavior at the query, session, and user levels, and major goals within each. We then present analysis results for each level.

4.1 Framework of Analysis

The overarching goal of our analysis is to understand how users find books and book-related information on the web. Figure 2 illustrates the major steps of user interactions for book-related search assumed in this paper. One way users can begin their interaction is from within the digital library (e.g., by typing the URL into the address bar), and using the internal search engine to find the relevant content, possibly with multiple iterations of query formulation and faceted filtering.

Alternatively, users can arrive from an external search engine that leads to either the digital library homepage or a content page directly, from which users can continue their interaction with the digital library. Since it is plausible that users entering from WSEs might have different intents from the those starting their interaction from within, we hypothesize that user behaviors between the former group and the latter group are different.

Compared to the user interaction models assumed in previous work, this model allows users to start their interaction from both the inside and outside the digital library, and multiple modes of interaction once they enter the library. While we focus on book search behavior with the Open Library in this work, the proposed interaction model is generalizable to most web catalogs that rely on WSEs as a primary entry point, and provides various interaction methods based on rich domain-specific metadata.

Since *users* search books over one or more *sessions*, each of which is composed of one or more *queries* or steps of information-seeking, we structure our analysis into three levels of granularity: queries, sessions and users. We outline our expectations for each of these next.

4.1.1 Query-level Analysis

Queries form a unit of a user's expression of an information need, and therefore it is critical to understand the characteristics of queries in our analysis. As we assume that users can search from inside or outside of the digital library, each presenting a different search environment, we distinguish between internal and external queries. As digital libraries provide a rich set of interaction methods, internal queries can be a keyword query, a structured query (using search operators or the advanced search interface), a set of filtering conditions, or any combination thereof.

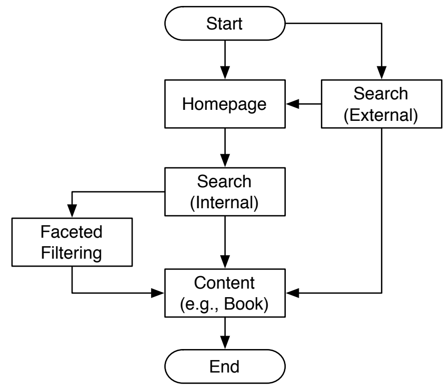

Figure 2: The sequence of user interaction for book-related search assumed in this paper.

Given these diverse classes of queries, the goal of our query-level analysis is to understand how internal and external queries differ in terms of length, the use of operators, and other aspects. For internal queries, we want to understand the use of search operators, the advanced search interface, and filtering conditions; we would like to know how frequently these different query types are used. Finally, since we hypothesize that field-level metadata plays a critical role in book-related searches, we analyze the usage of different metadata fields for each interaction method.

4.1.2 Session-level Analysis

While queries provide a glimpse of how users express their information needs, a more complete understanding of search behavior requires the inspection of the interaction between the user and the system as a whole. Analyzing user behavior at the level of sessions, defined as a series of system-user interaction with a single goal,[7] allows us to look at how users combine internal and external search and interaction methods, like filtering, to accomplish their goal.

Since we hypothesize that users starting from within the digital library have different intents from the those arriving via a WSE, we make a distinction between internal sessions (starting from within the digital library) and external sessions (starting from the referral of a WSE). We first compare the statistics of internal and external sessions, and analyze how users' navigation behavior in internal sessions differ from external sessions. Finally, since the internal search engine provides both advanced search and filtering interfaces, we analyze users' query reformulation behavior within internal sessions.

4.1.3 User-level Analysis

Moving beyond stand-alone sessions, we want to understand the habits of users across multiple sessions. As defined in Section 3.2, we identify users by a combination of IP address and the date on which activity is logged. Thus, activity from the same IP address on two different days is interpreted as coming from two distinct users. This means

[7] As noted in the definition of *session* in Section 3.2, we approximate user goal boundaries using a 30 minute timeout between adjacent user interactions.

Table 1: The distribution of SERP click counts for different object types in internal vs. external queries.

Source	works	books	authors	subjects	Total
External	575,291	8,634,914	1,094,050	46,737	10,350,992
Internal	413,775	8,123	133,829	69,260	624,987

Table 2: The distribution of search types for internal queries.

works (metadata)	works (full-text)	authors	subjects
1,006,487	18,113	75,141	80,096

that user-level statistics are only ever computed over a single day. For user-level analysis, we are interested in how often users search internally and externally, how consistent their use of internal versus external search is between sessions, and general statistics, like the average number of sessions, queries, and SERP clicks per user.

4.2 Query-level Analysis Results

Here we present the results for our query-level analysis. We first present the overall query statistics, and the results on search operator and method usage. We then provide a detailed analysis of how fields are used in keyword queries (both implicitly and explicitly by operators), queries from the advanced search interface, and faceted filtering conditions.

4.2.1 Query Statistics

We first compare the volume of internal and external searches. Note that we compare over SERP clicks as opposed to search requests here in following analyses, since we only have SERP clicks and the corresponding queries for external searches.

Table 1 shows the distribution of SERP click counts for different object types. Clicks from external searches account for 94% of all SERP clicks, and it is clear that clicks from internal searches exhibit very different patterns from external searches. While it is natural that *work* clicks overwhelm *book* clicks for internal searches, since internal SERPs link to *works*, it is interesting that *book* clicks substantiate a vast majority of external clicks. It might be due to the fact that search engines favor indexing book over work pages, or users looking for books tend to click on results with 'books' rather than 'works' in the URL.

The OL provides vertical search over books, authors, and subjects and provides the option to search over the full-text or only metadata (the default) for book search. Table 2 shows the distribution over these search verticals and options for internal queries. The two columns on the left denote the count of book searches based on metadata vs. full-text contents—the option which is accessible from advanced search interface, and the right columns show statistics about the other verticals.

It is interesting that full-text search accounts for only 1.8% of metadata-based search, even after considering that users need to enter the advanced interface to do so. We revisit this issue in the following section. Also, the search volume over authors and subjects each is less than 10% of book searches. Based on the statistics above, we focus on the analysis of metadata-based book search in what follows.

Table 3: Statistics of query lengths and operator usage for internal and external searches.

Source	length	quote	plus	minus	field	count
External	4.24	2.95%	0.09%	0.08%	0.02%	8,634,914
Internal	2.31	0.55%	0.01%	0.01%	0.27%	413,775

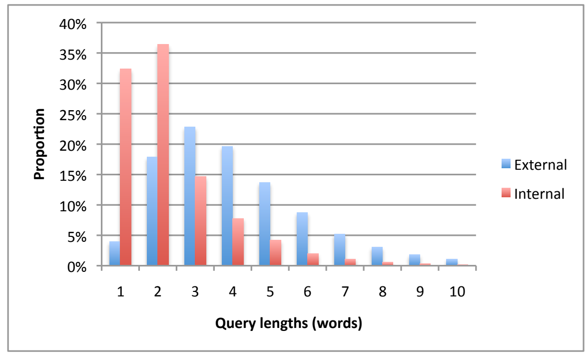

Figure 3: The distribution of query length (number of terms) based on source.

4.2.2 Search Feature Usage

Here we analyze how various search features supported by the internal and external search engines are used by book searchers. We first compare query length and the operator usage statistics for internal versus external searches in Table 3. Internal search queries are significantly shorter (2.31 words) than those from external search engines (4.24 words). The distribution of query lengths is shown in Figure 3. It is clear that the bulk of external queries are in the 2–4 word range, while internal queries contain 1–2 words.

Analyzing queries from both sources suggests the increase in query length for external searches is due in part to an abundance of book title searches, though we will see additional explanations in the coming sections. As for the operator usage shown in Table 3, while most operators are used more frequently for external queries, field operators are more common in internal searches. Note that the internal search engine provides field operators corresponding to book-level metadata, which may account for this difference.

While the user can only submit typed queries to external search engines, the internal search engine provides additional field-based search features via the advanced search interface and faceted filtering. It also supports options to specify whether users want results that only contain works with available e-books, or to change the sorting order of initial results (by the number of editions, oldest or newest in terms of publication date). Here, we looked at how frequently these search features are used for internal queries.

The results in Table 4 show that other search features such as the advanced search interface (advIF), faceted filtering and sorting by other criteria are all used more frequently than any search operators, including field operators (we include field operator usage for comparison). The use of these rich interaction methods may attribute to the discrepancy in query lengths between internal and external queries, since

Table 4: Statistics of search feature use in internal searches.

field	advIF	facets	e-book	sort
0.27%	2.41%	9.59%	32.59%	0.76%

Table 5: Correlation of keyword query length and search feature use for internal searches.

quote	field	advIF	facets
0.044	-0.011	-0.269	-0.082

these methods can be used instead of entering additional terms in a query.

As for search options, it turns out that filtering results to show only books with available e-books is a very popular activity (33% of book searches), which stands in stark contrast with the dearth of full-text search (1.7% of total book searches). Sorting of initial search results was used for 0.8% of internal queries (3,592 queries), and the most popular sorting orders were by the most recent first (64%), the least recent first (27%) and the number of editions (9%). From this we can infer that recency can be a important feature for the book search ranking algorithm.

To verify the hypothesis that the use of advanced search methods reduces the length of queries, we investigate the Spearman's rank correlation between query length and search method use for internal searches. The results in Table 5 show that the use of search features have a negative correlation with the length of keyword queries, except in the case of the *quote* operator. The trend is most conspicuous with the use of the advanced search interface and faceted filtering, which provide field-based interactions that complement keyword queries. The positive correlation of query length and the quote operator can be explained by considering that many people submit queries consisting of quoted book titles or author names, which are typically longer than average keyword queries.

4.2.3 Field Usage

Users can specify field-specific search terms in their keyword query or via the advanced search interface, but they can also restrict results to those with specific field values via filtering. In the previous section, we found that the field operators and field-based interaction methods are often used for internal searches. Here we investigate the use of fields more closely.

Table 6 compares the field use in search operators and the advanced search interface across all internal queries. While *title* and *authors* are popular in both search methods, *isbn* and *subject* are used mostly with the advanced search interface.

Previously in Table 4, we found that faceted filtering is by far the most frequently used search method. Table 7 shows the frequency of field usage for faceted filtering. While *title* and *authors* are still popular, *publisher* and *language* are used with disproportionately high frequency compared to other methods. The high use of the *language* filter is partly explained by the fact that corresponding search operator and advance interface field do not exist, whereas *publisher* field is available for other search methods.

Overall, the statistics reveals an interesting trend: the use of metadata fields is different based on the mode of interaction. For instance, it is plausible that people do not realize that they can use operators such as *isbn:* and *subject:*. Also, some field values (e.g., *publisher*) may be hard to type in, yet easy to click on. As will be discussed further in Section 4.5, this finding has implications for the search interface design,

Table 8: Average duration and action count for sessions from external and internal entry point.

Entry Point	duration (sec.)	actions	total sessions
External	69	1.76	7,516,990
Internal	262	3.79	286,587

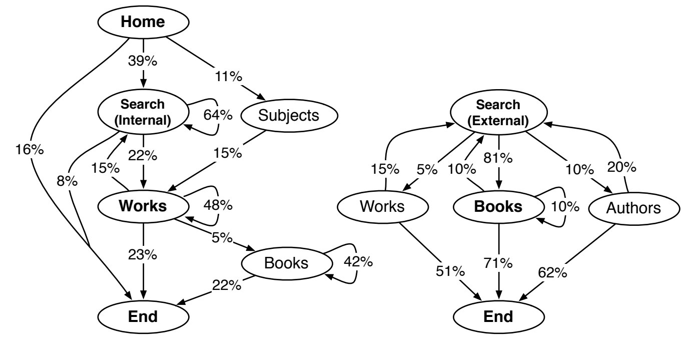

Figure 4: Markov model of navigation behavior for sessions started from the homepage (left), and started from the Google search engine (right).

many of which present fields in the same order regardless of the modes of interaction.

4.3 Session-level Analysis Results

Here we present the results of our session-level analysis. We first present overall session statistics from our data set and analyze user navigation behavior captured in page-to-page transitions conditioned on entry point. We then shift our focus to query reformulation behavior, which includes query reformulation and faceted filtering.

4.3.1 Session Statistics

Table 8 shows the session statistics from our data according to the point of entry. Session lengths are measured in duration and the count of user actions recorded. We find that external sessions are only half as long compared to internal sessions, and the difference is more significant in duration than in the action count.

We also looked at the distribution of action counts for external and internal sessions, where we found that about 90% of external sessions have fewer than 3 actions, whereas 70% of internal sessions consist of up to 3 actions. Consistent with previous results, users from external search engines tend to engage in brief search activities than those who start the interaction from within the website.

4.3.2 Analysis of Navigation Behavior

Previous work used Markov models to characterize user behavior in the context of email systems [3] and institution-level library systems [4]. To analyze the difference in user behavior between between externally and internally started sessions, we built a Markov model of navigation behavior, mostly following the methodology described in [4]. We first filtered sessions by a representative starting point for both internal sessions (homepage) and external sessions (Google), and then calculated the transition probability from each of page type to other page types.

The results in Figure 4 show the navigation patterns for

Table 6: Statistics of field use via search operators and the advanced search interface, as a ratio within all internal queries.

	title	author	isbn	subject	place	person	publisher
Operator	0.110%	0.110%	0.020%	0.030%	0.005%	0.001%	0.010%
Adv. I/F	0.490%	0.800%	0.890%	0.300%	0.030%	0.006%	0.060%

Table 7: Statistics of field usage for faceted filtering, as a ratio within all internal queries.

author	subject	person	place	time	publish_year	publisher	language
3.16%	2.75%	0.35%	0.59%	0.23%	0.43%	2.28%	1.34%

internal and external sessions, where the edge labels denote the probability that users move from one page to another. For instance, users starting from the *Home* page move into a *search* page with a probability of 39%. Note that we only show transitions with a minimum occurrence of 5%.

For internal sessions, a typical user enters the *search* page, and then visits a *work* page before leaving the website. It is typical that users re-visit the *search* page to reformulate a query or visit another *work* page. Some users diverge into a *subject* page to reach a *work* page, or visit a *book* page to find the information about a specific edition, though this happened rather infrequently. For external sessions, most users directly enter a *book* page, and then exit the website. Fewer than 20% of users enter the site via an *author* or *work* page.

There are several noticeable differences in navigation patterns between internal and external sessions. First, interactions in internal sessions show greater diversity with lower transition probabilities between major pages, whereas external sessions mostly include a single *book* page visit. This explains why external sessions are substantially shorter than internal sessions. Second, the most visited page types are vastly different between internal and external sessions and reflect the dominate result type for each source: Internal sessions mostly finish on a *work* page, whereas external sessions typically only include visits to *book* pages.

4.3.3 Analysis of Query Reformulation Behavior

The previous section explored user behavior as a whole, focusing on the difference between internal and external sessions. Since we found that internal sessions typically involve many query reformulations, filtering and other operations, here we investigate the patterns of query reformulation. Note that our query reformation include adding and removing a filtering condition. We built a Markov model of query reformulation by calculating the transition probability from one query behavior to another. The difference from the previous section is that each state of the model represents the form of reformulation taken by the user.

The results in Figure 5 illustrates the transitions between major reformulation actions. Note that we conflated the state of a session ending and a new query formulation to reduce visual clutter. While most users follow up a new query by ending their session or issuing a completely different query (87%), about 5% of users add a filtering condition, and another 7% reformulate the original query by adding or removing terms.

Once users add a filtering condition or reformulate their query, they tend to engage in the interaction by adding or deleting filtering conditions, or by reformulating the query again. The transition probabilities between faceted filtering actions were greater than those between query reformulation

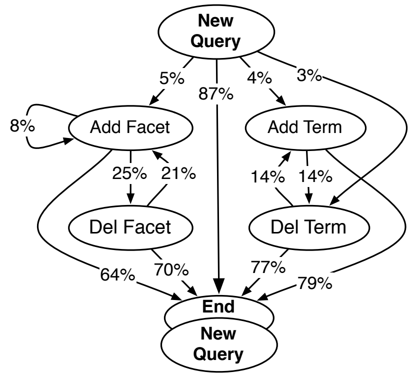

Figure 5: Markov model of query behavior for sessions started from the internal pages.

Table 9: User-level statistics from the October–December 2010 Open Library logs.

User counts	**All Users**	**Users with >1 query**	**Users with >1 session**
Total	8,728,077	2,015,013	582,947
Start all sessions from the OL homepage	322,441	135,117	8,765
Start *no* sessions from the OL homepage	8,373,609	1,857,640	542,155
Conduct only internal searches	1,014,483	159,013	50,945
Conduct only external searches	6,758,948	1,371,633	267,840
Average counts/user			
Actions	2.03	4.25	5.48
Queries	1.34	2.90	3.11
Internal queries	0.14	0.57	0.44
External queries	1.21	2.33	2.67
SERP clicks	1.35	2.88	3.13
Sessions	1.10	1.37	2.54

actions, which shows that users go through more iterations of faceted filtering than query reformulation.

4.4 User-level Analysis

In this section, we present the results of our user-level analysis. We first characterize the behavior of Open Library users in terms of the session starting point and the type of searches they perform. We then explore the generation of book suggestions using user-level session data.

4.4.1 User Statistics

Table 9 shows a summary of the user-level statistics. Each of the three columns describes a different subset of the data. The *All users* column covers all users. The *Users with >1 query* column is the subset of users that entered more than one internal or external search, which includes 23% of all users. Finally, the *Users with >1 session* column describes the subset of users that engaged in two or more sessions and includes approximately 7% of all users.

The first several rows give counts of users that meet the specified criteria. The second and third rows describe how consistent users are across sessions in terms of how they begin their session—either from starting from the OL homepage or a different page, including SERP clicks for external searches. Only 4% of all users begin all of their sessions from the OL homepage, while almost all other users begin all their sessions from somewhere else. Of users conducting at least two sessions, the vast majority—93%—start from somewhere other than the homepage in every session, while only 1.5% consistently start from the homepage, and 5.5% start at least one session from the homepage, and one from a different page.

The next two rows indicate the number of users that exclusively submitted queries either internally or externally across their sessions. Users are more consistent in searching externally, though this drops significantly for users engaging in more than one session, while inconsistency dominates. This suggests that many users rely on a mix of internal and external searching.

The bottom half of the table describes average event counts. As expected, users conducting multiple sessions tend to enter more queries. However, the ratio of internal to external searches is the highest for users who enter multiple queries, but who do not necessarily conduct multiple sessions: 0.24 versus 0.16 (users with multiple sessions) and 0.12 (all users). This is possibility explained by users entering the site via an external search and then performing a follow-up internal query.

Most users seem interested in the OL's content and not its search service. It is difficult to discern by only looking at the log data why users with multiple sessions prefer external searching—it could be that they are looking for other media in addition to book records, like Wikipedia pages, book reviews, or shopping sites like Amazon. In this case, using a WSE makes more sense. However, it may also be due to other aspects of users' online experiences. For example, external WSEs may be used as a shortcut—plugging a query into the default search box built into the browser and clicking on a link takes much less effort than navigating to the OL web site and entering the query there. Without an in depth user study, we can only speculate.

4.4.2 Extracting Book Suggestions

In addition to user-level statistics, we investigate the possibility of extracting book suggestions from the user-level. Table 10 shows a sample of four frequently clicked books (in bold) and the five most commonly co-occurring book clicks associated with each. For example, users who click on the book *1984* also click on *Animal Farm* and *Brave New World*. Surprisingly, *Pride and Prejudice* co-occurs frequently with slightly more antiquated classics, a grouping that would be unlikely using genre and publication era meta data alone. These initial results, based on a limited sample of data, suggest recommendations are a promising application of co-occurring book clicks.

4.5 Implications of Findings

The analyses presented in earlier sections have many implications for improving book search on the web. First, since search sessions typically span both internal and external search engines, a coordination between internal and external search is necessary to maximize user benefit. Second, the importance of metadata was stressed in both query formulation and the use of advanced search features, like filtering. In what follows, we provide specific design recommendations for both external WSEs and the Open Library's internal search engine.

Table 10: A sample of some of the most frequently co-occurring book clicks at the user-level.

Pride and Prejudice	**1984**
Chapman's Homer	Animal Farm
Odyssey	Brave New World
Don Quixote	The Hobbit
Divina commedia	The Call of the Wild
Faust: A Tragedy	The Great Gatsby

The Right Stuff	**Little women**
The right stuff	Wuthering Heights
Ascent	Pride and Prejudice
Omon Ra	Through the looking-glass
Man on the Moon	Gulliver's travels
Now and Forever	A Christmas Carol

4.5.1 External Search Engines

The proliferation of web search means that it functions as a entry point for many types of information-seeking and our analysis demonstrates that the same holds true of book search. However, the distinctive characteristics of book search do not seem to be taken into account by WSEs. Assuming that book-related queries can be classified with reasonable accuracy, there are two major ways in which we believe WSEs can better serve book searchers.

First, WSEs could allow field-based access of library catalogs. While services such as the Open Library use metadata with many fields, supporting the interaction (field operator, advanced UI or faceted filtering) based on the key fields across common across most library sites would be beneficial for users. And second, since a large fraction of internal searches included filtering results to only show those with available e-books, WSEs might consider including shortcuts to e-books next to Open Library results when available.

4.5.2 Internal Search Engines

Our analysis has several implications for design choices to the Open Library search engine and the site as a whole. While we only analyzed data from the Open Library, there is no reason these design recommendation could not be applied to similar digital libraries, or online catalogues in general. First, given the importance of fields in user's information-seeking process, field-level metadata should be exploited for ranking, query suggestions, and other purposes. The results from our retrieval experiments in Section 5.3 also support this argument. The analysis in Section 4.2.2 also showed that users consider recency an important ranking criteria.

Second, consider that most users are directed to an Open Library book page from an external WSE, after which they leave the site, sometimes performing another external search. To prevent users from leaving, the Open Library needs to present an easy-to-use feature that external WSEs do not have. We believe one such feature is a list of book recommendations presented on each book and work page. As we saw in Section 4.4.2, even with three months of data and a simple recommendation algorithm, the suggestions seem

Table 11: Statistics of the queries in the training and test collections.

Collection	Train Count	Train Avg. Length	Test Count	Test Avg. Length
Internal-Works	704	2.00	686	2.02
Internal-Authors	697	1.70	646	1.65
External-Books	9,758	3.44	9,911	3.45
External-Authors	115	2.84	105	2.65

useful. Extending this to multiple years of data and some additional information, such as dwell time and whether users tried to download the e-book or buy it from a linked vendor, should produce higher quality recommendations.

5. RETRIEVAL EXPERIMENTS

In this section, we present retrieval experiments using a test collection created from the Open Library log data. We first describe how we created test collection, and then experimental settings including the methods we compared, followed by the experimental results.

5.1 Building a Test Collection

Here we describe the process of building a test collection out of queries and click data. We first describe how we collected a reasonable set of queries, and then how we eliminated the position biases from click data.

5.1.1 Query Cleaning

To build a training and test collection, we first cleaned all the queries and aggregated their clicks over a year's worth of data (1.9 billion records). Cleaning entailed removing consecutive whitespace, non alpha-numeric characters, and converting all characters to lower case. Two queries were considered the same if they shared the same source (i.e., internal versus external searches), general key word phrase, advanced interface queries, and search facets. We ignored all queries with user-defined operators in the general keyword query, e.g., "title: moby dick" and queries with facets. We then broke these into two groups a training set Tr and a testing set Te.

Within each set, we broke queries and document clicks down into one of the following groups: internal searches over works (IW), internal searches over authors (IA), external searches over books (EB), and external searches over authors (EA). A query may belong to one or more of these groups, though the relevance judgments are specific to the object type (books or author).

We further filtered out queries without a sufficient number of clicks so as to not be overwhelmed with the data tail (too few total clicks) or the uninteresting queries where the same result is always clicked. Sufficiency required there be at least two distinct clicked documents of that object type (e.g., books) and a total of 30 or more clicks between the documents within that group. Table 11 shows the statistics of the queries built by the procedure described above.

5.1.2 Click De-biasing

Previous research has shown clicks to be biased by the presentation rank, and thus are not well suited for relevance judgments. To mitigate this for the internal queries, we followed the background subtraction model introduced by Agichtein et al. [1].

The process is as follows: first, we established an expected relative click frequency model over the queries—for each query, we calculated the proportion of clicks made at each rank, then averaged each rank's relative click frequency across all queries. Because rankings are not available in the Apache logs, we used the rankings of the current system as a surrogate. For each group in the training set, we generated a background model using all the queries with a sufficient number of clicks and a random sample of queries without. For each group in the testing set, we took the corresponding training background model and added the relative click frequencies the test queries with sufficient clicks.

Once the background models were established, we de-biased queries in the training and testing sets by subtracting the background model from each query's relative click frequencies. After the subtraction, a result with a de-biased click frequency greater than 0 was normalized on a graded scale from 1 to 5. This was then used as the relevance judgment for that query-document pair.

To quantify the effectiveness of the de-biasing, we randomly sampled 20 queries from the internal book and author test sets, gathered the records that were marked relevant in either the biased or de-biased implicit relevance qrels, pooled the top 10 results from each of the three retrieval systems, and manually assessed each record. Using the explicit relevance judgments as the gold standard, we calculated the accuracy of the binarized biased and de-biased implicit relevance judgments. We found that de-biasing greatly increases the the accuracy for internal book search from 0.29 to 0.54.

Note that this de-biasing model assumes that users view internal search results, which are available by crawling the OL search results for the queries within our set. For external queries, due to the difficulty in obtaining ranking information for the variety of external search engines, we decided not to de-bias their clicks. Instead, if a document is clicked, it is counted as relevant.

Table 12: The list of fields indexed.

Category	Field
title	title, series, edition
author	authors, bio
subject	subjects, subject-people, subject-places
time	publish-date, death-date
place	publish-places, publish-country
person	personal-name, alternate-names
publisher	publishers

Table 13: Record counts for the training and test collections.

Collection	Record Count
Works	15,032,815
Books	24,359,019
Authors	6,730,812

5.2 Experimental Settings

Here we describe settings for our retrieval experiments. During indexing, no stemming was applied, and no stopwords were eliminated. Table 12 lists the indexed fields and their categories corresponding to the fields in Table 6. Table 13 provides the number of records contained in each of the collections. All experiments were conducted with a modified version of the Galago retrieval system.[8]

Since our goal in retrieval experiments is to provide results for well-established baseline retrieval models, we exper-

[8]http://www.lemurproject.org/galago.php

Table 14: Retrieval results (MAP).

Internal queries	*QL*	*MFLM*	*BM25F*
Works	0.1866	0.1913	0.2041
Authors	0.4114	0.4311	0.4277

External queries	*QL*	*MFLM*	*BM25F*
Books	0.5568	0.5268	0.5924
Authors	0.3151	0.3224	0.3282

Table 15: Retrieval results (P@5).

Internal queries	*QL*	*MFLM*	*BM25F*
Works	0.3000	0.2040	0.2120
Authors	0.1867	0.1972	0.2012

External queries	*QL*	*MFLM*	*BM25F*
Books	0.1880	0.2360	0.2840
Authors	0.1644	0.1584	0.1624

iment with the query-likelihood language model (QL) [14], the mixture of field language models (MFLM) by Ogilvie and Callan [13], and the 2004 TREC Hard track version of the BM25F model by Robertson et al. [17]. A detailed descriptions of these models can be found in [8].

Each model is tuned using training queries within each group. MFLM has each language model weight tuned independently, and for BM25F, the order of tuning for the parameters is: $K_1 \rightarrow B_f \rightarrow W_f$. All field-level parameters are tuned by iterative line search.

5.3 Experimental Results

Tables 14 and 15 list the mean average precision (MAP) and precision at rank 5 (P@5) of the systems, respectively. Internal book queries were run over *works* while the external book queries were run over *books*—this reflects the click-through behavior we observed in the logs. The results are consistent with a recent evaluation of structured document retrieval models [8], where field-based models showed advantages over document-based models.

Another interpretation is that the results tend to improve as the amount of parameterization increases (from left-to-right in the tables). We hypothesize that the high number of fields in these sets provides opportunity for fine-grained tuning, which is what the additional parameters provide in a model such as BM25F.

Finally, MAP results tend to be higher than the corresponding P@5 values, indicating that while the experimental models can effectively retrieve relevant documents, they have difficulties in getting them consistently into the top 5 results.

6. CONCLUSIONS

In this work, we analyze web-based book search behavior using three months of logs from the Open Library. We find that the query characteristics and session-level behaviors are drastically different between internal and external searchers. We also study the use of various field-based interaction methods within library website, finding that the field usage is different based on the modes of interaction—keyword search, advanced search interface, faceted filtering. We make numerous design recommendations for both web search engines and digital libraries.

In addition to the behavior analysis, we also took the first steps towards developing a viable test collection based on the Open Library metadata record and query log data. We also perform evaluations of several well-known retrieval methods. We are in the process of preparing to release this test collection, so that it can be used by researchers to further improve systems that provide book search capabilities.

An avenue of future work includes implementing the recommendations we made here in a large-scale digital library service. By selectively exposing the new features to user traffic in a controlled manner, we expect to uncover the degree to which our suggestions make a difference in the digital library user experience.

7. ACKNOWLEDGMENTS

This work was supported in part by the Center for Intelligent Information Retrieval, in part by UMass NEAGAP fellowship, and in part by NSF grant #IIS-0910884. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

References

[1] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proc. of SIGIR '06*, pages 3–10, 2006.

[2] B. A. Blummer. A literature review of academic library web page studies. *Journal of Web Librarianship*, 1(1):45, 2007.

[3] D. Elsweiler, S. Mandl, and L. Azzopardi. Evaluating personal information management behaviour using markov models. In *Wissensmanagement*, pages 280–288, 2009.

[4] Q. Guo. Modeling user interactions for automatic library search evaluation: Preliminary results. In *Proc. of the JCDL'08*, 2008.

[5] S. Jones, S. J. Cunningham, R. Mcnab, and S. Boddie. A transaction log analysis of a digital library. *Int. Journal on Digital Libraries*, 3:152–169, 2000.

[6] G. Kazai, M. Koolean, J. Kamps, A. Doucet, and M. Landoni. Overview of the inex 2010 book track. In *INEX 2010 Workshop Proceedings*, 2011.

[7] G. Kazai, M. Koolean, J. Kamps, A. Doucet, and M. Landoni. Overview of the inex 2011 book track. In *INEX 2011 Workshop Proceedings*, 2012. To appear.

[8] J. Kim and W. B. Croft. Retrieval experiments using pseudo-desktop collections. In *Proc. of CIKM*, CIKM '09, pages 1297–1306, 2009.

[9] B. Kules, R. Capra, M. Banta, and T. Sierra. What do exploratory searchers look at in a faceted search interface? In *Proc. of JCDL*, JCDL '09, pages 313–322, 2009.

[10] J. Lin and M. D. Smucker. How do users find things with pubmed?: towards automatic utility evaluation with user simulations. In *Proc. of SIGIR*, SIGIR '08, pages 19–26, 2008.

[11] G. Mishne and M. Rijke. A study of blog search. In *Proc. of ECIR '06*, pages 289–301, 2006.

[12] X. Niu and B. M. Hemminger. Beyond text querying and ranking list: how people are searching through faceted catalogs in two library environments. In *Proc. of ASIS&T*, ASIS&T '10, pages 29:1–29:9, 2010.

[13] P. Ogilvie and J. Callan. Combining document representations for known-item search. In *Proc. of SIGIR*, SIGIR '03, pages 143–150, New York, NY, USA, 2003. ACM.

[14] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *Proc. of SIGIR*, SIGIR '98, pages 275–281, New York, NY, USA, 1998. ACM.

[15] F. Radlinski and T. Joachims. Query chains: learning to rank from implicit feedback. In *Proc. of KDD '05*, pages 239–248. ACM, 2005.

[16] W. Weerkamp, R. Berendsen, B. Kovachev, E. Meij, K. Balog, and M. de Rijke. People searching for people: Analysis of a people search engine log. In *Proc. of SIGIR*, pages 45–54. ACM, ACM, July 2011.

[17] H. Zaragoza, N. Craswell, M. Taylor, S. Saria, and S. Robertson. Microsoft Cambridge at TREC-13: Web and HARD Tracks. In *Proc. of TREC*, TREC '04, 2004.