NORDIC JOURNAL OF INFORMATION LITERACY IN HIGHER EDUCATION

2018, vol. 10, issue 1, 18–36 10.15845/noril.v10i1.270

"More than Meets the Eye" – Analysing the Success of User Queries in a Library Discovery System

Peer Reviewed

Hugo C. Huurdeman, Mikaela Aamodt*, Dan Michael Heggø*

University of Oslo Library

Abstract

Discovery systems allow academic library users to locate a wider range of resources than previous OPACs. However, actual usage of these systems may still be challenging. The main aim of this research is to get a better understanding of the hurdles users face while searching contemporary library systems.This study utilizes a transaction log analysis approach, using popular and zero result queries datasets gathered from the statistics module of a library discovery system. It explores what types of queries users perform, how successful the queries are, and examines underlying reasons for unsuccessful queries. To our knowledge, this is the first academic paper to use data originating from built-in transaction logs of the Oria library discovery system. The analysis shows that queries are often curriculum-related: we could pinpoint a relation with curriculum for 58% of the popular queries, and 28% for the zero result searches. A vast majority of popular queries refer to books, databases and journals, and over half of the queries used the title to locate a resource. 20% of the popular queries turned out to be unsuccessful. Zero result queries typically involve long queries, and in many cases consist of pasted reference citations. Our conclusion is that the examined discovery system is rather sensitive. Whilst this suggests the importance of increasing users' information search skills, it also points to the need for enhancing discovery systems and their underlying metadata. Furthermore, due to the prominence of curriculum-related queries, a better integration of curriculum materials ought to be achieved.

Keywords: user queries; query log analysis; user experience; oria; primo

***Contact:**
e-mail: mikaela.aamodt@ub.uio.no, d.m.heggo@ub.uio.no

Introduction

A wide variety of information resources is available through academic libraries, and can be located using user-facing information retrieval (IR) systems, formerly manifested by online public online catalogues (OPACs), and in more recent times increasingly by discovery systems (Hofmann & Yang, 2012). Discovery systems allow users to directly access library materials from a wider variety of sources than previous OPACs, since they contain a central index of both local and remote resources. In parallel with the evolvement of library systems, user expectations have simultaneously evolved. Web search engines such as Google have increasingly shaped users' expectations of search systems (Griffiths & Brophy, 2005; Fast & Campbell, 2004), and it has been found that users apply typical web search strategies in library systems (Shiv, 2012; Willson & Given, 2010). Discovery systems still have their roots in the online catalogue, which led us to believe that traditionally acknowledged challenges users face, as studied by e.g. Borgman (1986, 1996) and Kani-Zabihi, Ghinea & Chen (2008) could still remain. Challenges include difficulties performing subject searches, difficulties to increase and decrease the number of search results, incorrect use of Boolean logic and not utilizing the more sophisticated capabilities of the system.

To get a better understanding of hurdles in contemporary library catalogues, we study user interactions with the discovery system of the University of Oslo Library. This is done within the context of the Visual Navigation Project. Based on newly available transactional log data, we harness new opportunities to analyse user interactions with current library catalogues, and to potentially improve digital access to library collections based on lessons learned. We explore what types of queries users perform, determine how successful user queries are, and look at underlying reasons for unsuccessful queries. This leads to the following research questions:

RQ 1. Which insights can we gain from classifying user queries within Oria by popularity, specificity and target resources?
RQ 2. To what extent are the most popular user queries successful, and how do the most popular queries evolve over time?
RQ 3. What underlying reasons for zero result queries can be determined?

Even though transaction log analysis has been readily applied to logs originating from library catalogues in the past years (see e.g. Chapman et al., 2011; Harnath & Kottman, 2015; Lown, Sierra, & Boyer, 2013; Niu, Zhang, & Chen, 2014) our analysis extends previous analyses in different ways. First of all, we intend to determine the successfulness of user queries, and to pinpoint underlying reasons for unsuccessful queries. Second, we look at the relation of popular as well as unsuccessful queries with the curriculum of a large university. Moreover, to our knowledge, we are the first academic paper to use data originating from built-in transaction logs from Primo, a popular library discovery system. Our findings can have implications for the design of library systems, cataloguing and metadata, the role of information literacy, as well as future improvement in transactional data logging in the context of libraries' discovery systems.

Previous Literature

To evaluate current use of library systems, and to design future solutions, it is essential to understand the actual use of library systems. For this purpose, a variety of ways may be employed, including user studies in a lab setting, and transaction log studies. Transaction logs "arise from the activities recorded when people interact with computer systems and services" (Dumais, Jeffries, Russell, Tang & Teevan, 2014, p. 350). More specifically, they capture "the communications (i.e., transactions) between a system and the users of that system." (Jansen, 2006, p. 408). Logging can include search queries in search applications, clicks on search results, browsing patterns, mouse movements, and so forth. There are advantages to utilizing logs: Dumais et al. (2014, p. 351) characterize log analysis as "the most natural observations of people as they use systems in whatever ways they typically do", as opposed to lab and field studies, which are more controlled, but potentially have a larger impact on user behaviour. However, there are also drawbacks to the use of transaction log analysis. Jansen (2006, p. 411) indicates that transaction logs may be incomplete, and do not "capture the underlying situational, cognitive, or affective elements of the searching process". Thus, as Markey (2007a) suggests, the use of transaction logs imposes a limit on the types of research questions which can be asked.

Despite potential limitations, transaction log analysis has been performed frequently in the fields of information science and information retrieval, since they allow researchers to study "real users" and "real needs" (Jansen, Spink, & Saracevic, 2000). Jansen et al. (2000), performed a log analysis of user queries on the web, using a large-scale dataset from the Excite search engine. In particular, they looked at user *sessions* ("changes in queries during a sessions, number of pages viewed, and use of relevance feedback"), *queries* ("number of search terms, and the use of logic and modifiers"), and *terms*, ("their rank/frequency distribution and the most highly used search terms."). In a follow-up study, Spink, Wolfram, Jansen, & Saracevic (2001) extended their analysis by studying one million queries of the same search engine. They found that searchers used few search terms (2.4 terms per query), visit few pages and generally only look at the first (and sometimes second) result page. Advanced search features were used sparingly, and Spink et al. (2001) even found that the use of these were erroneous in half of the cases. In a literature study, Markey (2007a) summarized 25 years of end-user searching studies using transaction logs. She concludes that user sessions are generally short – in most studies, the mean number of queries in a session is "between two and four queries per session". In her follow-up paper, Markey (2007b) proposes extensions to the types of studies referring to user queries, for instance by studying multi-session searches.

In recent years, a variety of transaction log studies have been done in a library context (e.g. Han, Jeong, & Wolfram, 2014; Lown et el., 2013; Meadow & Meadow, 2012; Moulaison, 2008; Niu et al., 2014). Most of these studies found, in concordance with earlier log-based studies in a web context, that users issue two to three query words on average, have short sessions, and made little use of Boolean logic in their queries. Lown et al. (2013) specifically looked at user behaviour in single search box environments, based on a set of nearly 1.4 million transactions. They found that the many of the most frequent queries were "not well served" by

the catalogue and articles modules of the search system, as the top 20 queries mainly referred to database titles, journal titles and support information searches. Hanrath and Kottman (2015) investigated the Ex Libris Primo system, using a combination of usability testing, web analytics and further usage statistics from Primo. In their discussion, they underline the importance of distinguishing discovery and actual delivery, the discrepancy between user confidence based on self-assessment and actual usage statistics, as well as the importance of monitoring usage data to improve services. Chapman et al. (2011) reviewed the 100 most frequently occurring searches during a month in the unified search box of an academic library website. They categorized these as database, library service, publication, keyword/subject, and ambiguous searches. In addition, they investigated the "long tail" of specific searches using a sample of 992 searches from a larger set. Their results showed that 44% of all analysed queries were known-item queries, and 28% non known-item queries (i.e. exploratory, creator and library website).

Methodology

This study focuses on transaction log analysis of a large dataset with popular and zero result user queries. As indicated by Jansen (2006), transaction log analysis usually consists of three composite steps: *collection*, *data preparation* and *analysis*. This section describes the collection and data preparation, while our analysis is described in the *Results* section.

Collection

ExLibris Primo is an example of a discovery system, which embodies a single interface for discovering bibliographic records, digital collection materials and items in institutional repositories (Vaughan, 2011, p. 39). Primo Central, an extension of Primo, in addition provides a "central index of article-level content from a variety of publishers and aggregators". Hence, it is possible to search not just traditional library material like books and journals, but also book chapters, journal articles and much more. In terms of the ranking of results "Primo's proprietary relevancy-ranking algorithm includes but is not limited to factors such as term frequency, field weighting, number of times a record has been accessed, and currency." (Vaughan, 2011, p. 40).

The discovery system *Oria*, an installation of Primo, is a web-based tool that aims at providing a single point of entry into the collections of a consortium involving the majority of Norwegian academic libraries since January 2015. Within Oria, miscellaneous user actions are recorded, and aggregated statistical data can be accessed via Primo Analytics. To our knowledge, data originating from Primo Analytics has not been utilized and discussed in earlier academic papers, and no extensive studies have been done yet in a Norwegian context. To address our research questions, we obtained two datasets from Primo Analytics at the University of Oslo, Norway: *Popular Searches*, and *Zero Result Searches*, i.e. the queries for which no results were obtained. Data used in this paper is available via: https://github.com/uio-library/virak_paper_data.

The acquired *Popular Searches* dataset covers the period January to June 2015, and November 2015 to September 2016. For unknown reasons, no data existed for the period July

to October 2015. The dataset is aggregated by Primo Analytics with a monthly granularity. According to the Primo Analytics documentation (ExLibris, 2017), a search is seen as popular "if it has been performed at least 10 times within a month". Up to 500 of such searches are stored for each month. If fewer than 200 of such searches occur, however, Primo Analytics will store up to 200 popular searches with less than 10 occurrences. The second dataset, the *Zero Result Searches*, ranges from August 7 2015 to September 30 2016, and is available at a daily granularity (i.e., we get a list of zero result queries for each day in this time period).

Excluding single letter queries, the popular queries dataset includes 3,729 distinct queries performed in the University of Oslo's Oria instance – these queries have been sought 115,590 times (see Table 1). The mean number of words per query in the whole dataset indicate that user queries include few terms: 2.6 query words on average. This shows some similarities with earlier research using log studies (Spink et al., 2001), which showed a mean number of 2.16 query words, or Markey (2007a)'s review of previous log studies pointing to an average of 2 to 4 words per query. It should be noted that our number is not directly comparable with previous studies since our dataset only include popular queries, and the chance of a query to be repeated (i.e. popular) could increase with decreasing query length.

Table 1:
Datasets used in this study. Repeated queries are included in the calculation of means/medians.

Dataset	**Popular searches**	**Zero result searches**
Time period	Jan–Jun'15 + Nov'15–Sep'16	Aug '15–Sep '16
Number of distinct queries	3,729	36,221
Number of queries	115,590	52,257
Percentage of the total number of search actions in this time period	4.9%	2.2%
Mean (median) number of terms per query	2.6 (2)	5.6 (4)
Mean (median) number of characters per query	20 (16)	40 (29)

To get an overview of how the number of searches in our dataset relates to the total number of searches in the same period, we performed a cross-validation using the *Actions* dataset in Primo Analytics, containing statistics on the frequencies of user actions. This dataset, which does not contain any details about the performed queries themselves, indicates that in this timeframe 2,357,503 search actions were done in the system (of which 2,153,404 basic and 204,099 advanced searches). This means that the 115,590 searches in our dataset reflect 4.9% of all Oria search actions in the same period.

The Zero Result Searches dataset points to a variety of problematic queries leading to a dead-end, since no results were returned. The number of distinct queries (36,221) available in this set is higher than the Popular searches, but the zero result queries are issued less often than the Popular Searches: 52,257 times in total. This is likely caused by the larger variation in terms of the formulation of searches without results, which also is reflected in the higher average number of query words (5.6). In the same time period, 2,410,632 search actions (2,202,647 basic and 207,985 advanced) were registered in Primo Analytics.

Data Preparation

Before initiating the coding of our data, a series of processing steps was performed using Microsoft Excel. First, we normalized all queries, while retaining the raw queries as a reference. We did this by removing non-alphanumeric characters, such as ", *, < and >. This was done to obtain better statistics on the number of searches for a certain target resource. For instance, "*stanislav andreski"* and *stanislav andreski* were considered as being the same query, since they refer to the same target resource. There are a remarkable number of single letter queries that appears to stem from a navigational feature such as the alphabetical journal list. Since we were only interested in looking at the actual query formulations, we removed these single letter queries (*a-z*).

From the *Popular Searches* dataset, we excluded two queries in the query top 50 from our analysis, namely *direktekrav* [direct claim], a query verified to be used in a law library skills course (queried in various forms, e.g. *direkte krav, "direkte krav" and direktekrav**), and a query consisting of an ID from Alma, the used library resource management system, *998903677444702000*. This ID could not be connected to any valid resource, and was also excluded from our analysis.

After processing the data, we manually coded the 50 most popular queries in the dataset, accounting for a considerable degree of popular queries in terms of the number of searches, as well as a random set of 50 zero-result queries. Those two sets of queries were initially coded by one of the paper authors beginning October 2016. These codes were revisited by the two other authors on April 18–20 2017, and the decisions taken during the process were subsequently discussed in a group setting. The following elements were coded for both datasets. If derivable from the query, we noted the *nature of the query*, i.e. whether it was for a title, topic, author, ISBN, etc. Second, we looked at which is *the resource type* being sought, if derivable from the query, for instance a book, database or journal (called "perceived intent" by Chapman et al. (2013)). To help us in this process, we made use of a list of the top 500 most frequently loaned items. Using this data, we cross-checked if a query could refer to a commonly used book. Next, we checked whether a query was related to curriculum lists at UiO, where all reading lists are publicly available. We checked this by querying Google using the *site:uio.no* filter, and by checking the results for the occurrence of curriculum ("pensum") pages.

For the popular searches dataset, we then verified, if possible, whether a query was a success or not. We issued the searches in the Oria search box, and replicated the query settings

(e.g. search scope) that we could observe from the query logs. To prevent personalization from influencing the retrieved results, we performed the searches as a guest user. We defined a successful query as a query for which the initial result page (i.e., the first 10 results) included the resource likely sought for. This was used as the success criterion since many previous studies (e.g. Spink et al., 2001) have shown that searchers do rarely look beyond the first search results page. The process involved some interpretation of the query, since the anonymous nature of the query logs prevented us from verifying this with the actual users. In cases there was no obvious candidate for the intended result, we coded this as an "unknown" resource. For the zero result searches dataset, we also tried to determine *why* a query did not return any results, for instance because of using the wrong field in the advanced search feature, because of a misspelling, or because the item simply was not available in Oria. Further checks were performed on 18–20 April 2017: whether the query at that point still resulted in zero items, and if so, whether the current spelling corrections feature suggested a valid query reformulation.

Results

Classifying User Queries

First, we address our first research question: *Which insights can we gain from classifying user queries within Oria by their popularity, specificity and target resources?* For this question, we look both at the Popular Searches and Zero Results dataset.

The queries in the top 50 of the **Popular Searches dataset** together have been issued 19,797 times (17.1% of the total number of searches in the popular searches dataset). The top 5 of most popular queries consists of *atekst* (issued 1,425 times), *pubmed* (1.221), *exphil* (719), *det kvalitative forskningsintervju* [the qualitative research interview] (711) and *nature* (669). Thus, we observe a wide diversity of requested sources including databases (*atekst* and *pubmed*), curriculum books (*det kvalitative forskningsintervju*), and a journal (*nature*).

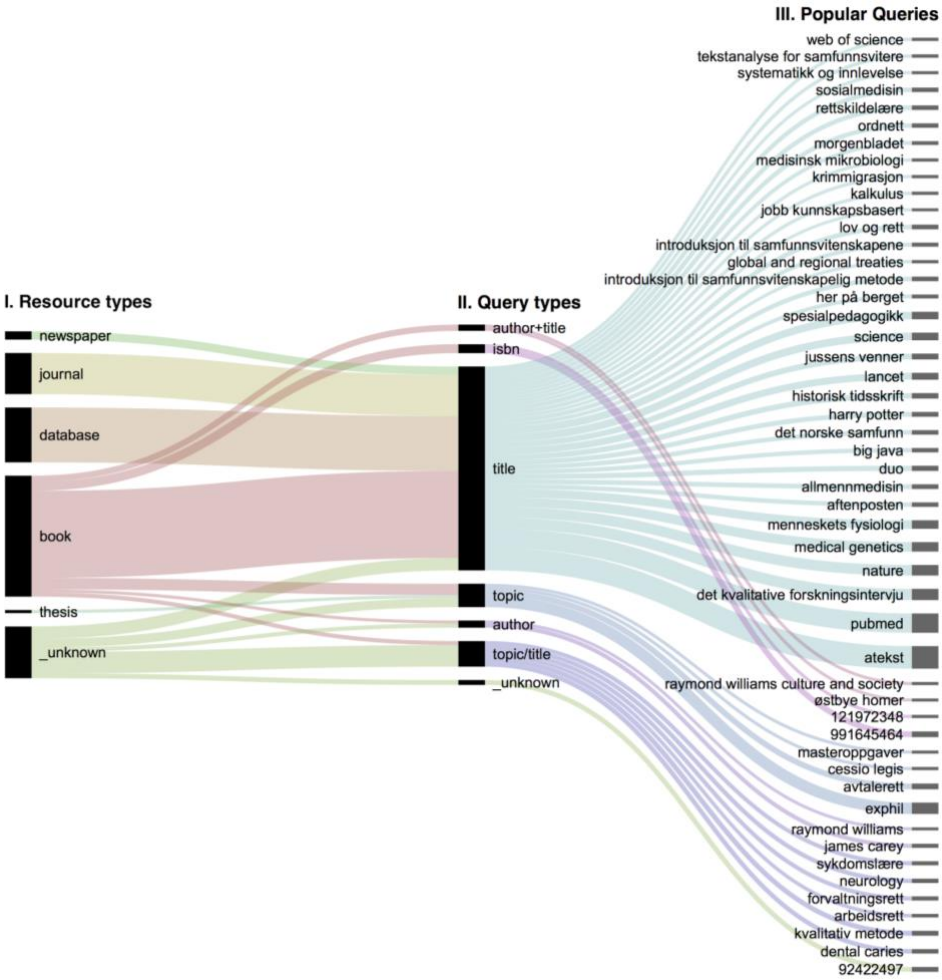

Figure 1. Alluvial diagram, depicting the relation between specified resource, query type and the query top 50. The size of the lines represent the number of issued searches corresponding to a query.

The alluvial diagram in Figure 1 visualizes the relations between types of resources, issued queries, and the actual query strings. **Column I** of the figure shows that in terms of resource types, half of the queries referred to books, for instance *det norske samfunn* [Norwegian society]. In 12% of the cases, users referred to journals, such as *Science* or the *Lancet*. 10% of the queries were related to databases, such as *atekst* (a Norwegian database with newspapers, among other things), *pubmed*, and *duo* (the university's repository). This can cause issues, since databases should normally be accessed via a link on the library homepage or via a link within the search interface. **Column II** of Figure 1 shows what types of queries were most popular in the surveyed time period. The majority of queries (66%) consisted of titles of books, journals and databases. For instance, some users simply entered book titles such as *det kvalitative forskningsintervju*. In 12% of the cases, searches consisted of a broader topic, for instance *spesialpedagogikk* [special needs education], which also matched with commonly loaned book titles. Sometimes (8% of the cases), a topic which does not match a regularly loaned book title was entered, such as *cessio legis* (an expression from law studies). **Column III** of Figure 1 depicts the full list of queries, as well as their relative popularity.

The full **Zero Results Searches dataset** includes a wide variety of queries. Many of the queries appear to be known-item queries, i.e. searches "for a 'known' object or an object 'known to exist'" (Lee, Renear, & Smith, 2006) – this shows similarities with the findings of Chapman et al. (2011). For this dataset, looking at the most commonly issued queries does not provide many insights, because of the sheer variation in the failed queries. Therefore, we coded a random sample of 50 queries.

Figure 2 summarizes the resource types, query types and query strings for the zero result searches. **Column I** shows that the most common identified resource type is a book (28%), for instance represented by the misspelled query *Prcopius Secret History*. In addition, specific book chapters are referred to in 12% of the cases (e.g. *Solhaug, (2006). Kapittel 13: Strategisk læring i samfunnsfag. I*) [Solhaug, (2006). Chapter 13: Strategic learning in social science]. For almost a quarter of the queries (24%), the specified resource could be classified as a specific (journal) article. This often appears to have been done by directly pasting a citation reference. In 8% of the cases the intended result was a journal. For 26% of the queries, the search terms were not clear enough to derive which target resource type was meant, and thus the target resource was categorized as 'unknown'. **Column II** depicts the types of entered queries. The largest query category consists of directly pasted citation references, accounting for 40% of the failed queries (for instance, the query *Browning, N. (2015). The ethics of two-way symmetry and the dilemmas of dialogic kantianism. Journal of Media Ethics*). Titles were also common (30%), such as *Sentralbankens oppgaver i dag og i fremtiden* [The central bank's tasks today and in the future]. This is lower than for the most popular queries dataset, but still considerable. In 16% of the cases, users referred to an author, e.g. *Christopher Hotchens* (likely intending Christopher Hitchens). Also, a few unsuccessful queries were for an ISBN number. Finally, **Column III** shows the actual zero result query strings for our sample.

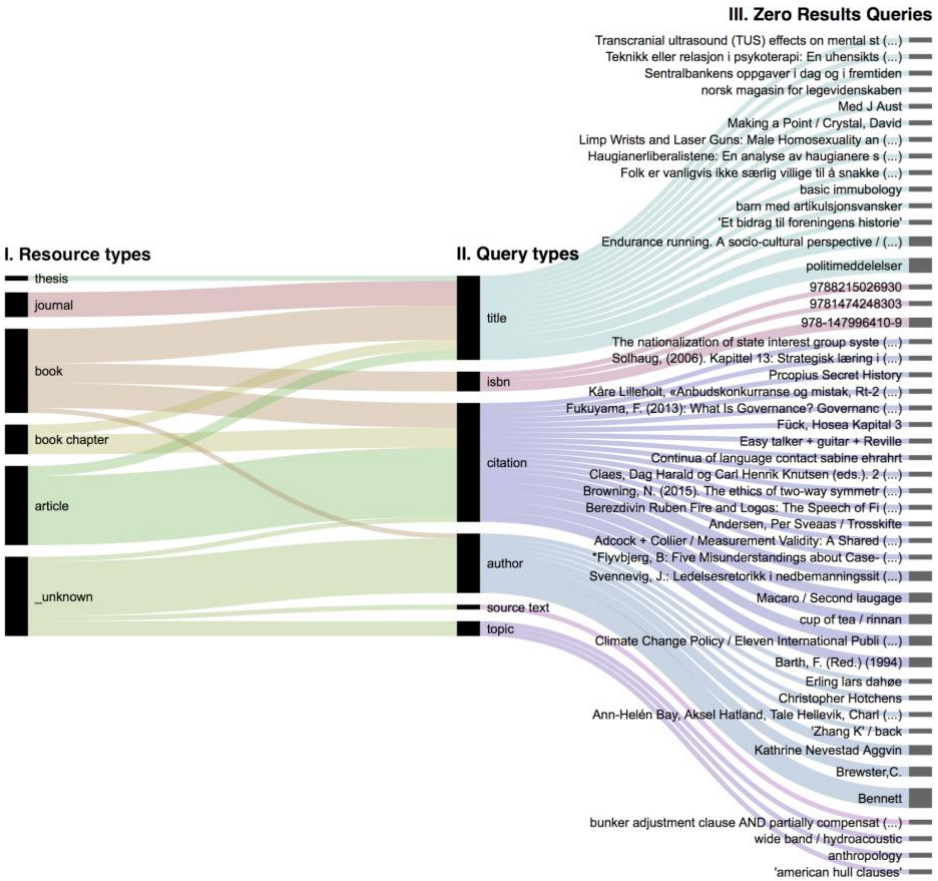

Figure 2. Alluvial diagram of intended resource and query type (if derivable) of all 50 analyzed zero result queries. The size of the lines represent the number of issued searches corresponding to a query. Queries longer than 50 characters are abbreviated.

In addition, we surveyed the relationship of queries with the materials in the curriculums of different studies. As described in Methodology section, finding out the relation of queries with the curriculum was done by searching curriculum lists on the website of the University of Oslo.

Table 2 (**a**) summarizes the results for the popular searches dataset. More than half of the entered queries (58%) could be connected to an actual curriculum list, for instance *menneskets fysiologi*, [human physiology] and *det kvalitative forskningsintervju* [the qualitative research interview]. In one third of the cases (34%), it could be determined that a query was not related to the curriculum. For the remainder of queries (8%), the nature of the query prevented determining with certitude whether it was related to the curriculum (e.g. the query *james carey*).

Table 2 (**b**) shows the relation between zero result searches and the curriculum. We could pinpoint almost one third of the unsuccessful queries (28%) to be related to reading list material, which is lower than for the popular queries, but still considerable. For instance, *Fukuyama, F. (2013): What Is Governance? Governance, Vol. 26, No. 3, July 2013 (s. 347–368)*, or the misspelled book query *basic immubology*. Due to the nature of the queries, we could not determine the connections with the curriculum for the remainder of inspected zero result searches (32%).

Table 2:
Curriculum-related searches in the top 50 of Popular searches and the sample of 50 Zero result searches

	Curriculum-related	**Not curriculum-related**	**Unknown**
a. Popular searches	58%	34%	8%
b. Zero result searches	28%	40%	32%

The success of user queries

Our next step is to investigate the successfulness and temporal distribution of the most popular queries in our dataset, addressing our second research question: *To what extent are the most popular user queries successful, and how are they distributed over time?*

We initiate by looking at the "success" of user queries. We define a query as a success if the target resource can be found on the first search results page, i.e. in the first 10 results (see *Data Preparation* section). First, Table 3 (**a**) summarizes the success and failure of the totality of 50 queries in the surveyed set. In over half of the cases (58%), the query is successful, but for 20% the resources cannot easily be found. In 22% of the cases (e.g. *james carey* or *neurology*) we could not determine with some certainty the perceived intent of the user and thus whether the search was successful.

A number of causes for the lack of success for certain queries can be determined. At the time of our analysis, there were no entries for some databases in the 50 most popular queries, meaning that these databases could not be found via the regular Oria search box. Also, ambiguous names are an issue: for example, *nature* and *science* result in a large number of hits, and the relevance ranking in Primo does not resolve this by putting the "right" results first. Table 3 (**b**) illustrates the average number of results retrieved for the queries. The average number of results per search measurement field is only available from Nov. 2015, so we looked at the 11 subsequent months (Nov. 2015–Sep 2016). From this information, we can tentatively derive that queries with a higher number of returned results seem to have more issues: for the 10 unsuccessful queries in the dataset of 50 most popular queries, the average number of results is 6.5 million, while for the 29 successful queries this is much lower at 38,486 (and 38,949 for the unknown queries).

Table 3:
Success and failure of the top 50 of searches (Popular Searches dataset)

	Successful	**Unsuccessful**	**Unknown**
a. Number of queries (perc.)	29 (58%)	10 (20%)	11 (22%)
b. Average number of retrieved results p/query	38,486	6,483,123	38,949

Second, we zoom in to the actual queries. Figure 3 summarizes the success of the top 15 of all Oria queries, as well as their temporal distribution. We can observe that quite a number of queries in the top 15 are not successful, or cannot be determined to be successful. For instance, a search for the database *pubmed* did not result in a link to the database, and the *nature* journal could not be found among the first 10 results. Hence, ranking issues appear to be causing some queries to fail. Naturally, there are many dimensions and potential causes of this problem; this has to be researched further in future work, also using more extensive query logs. From the figure, we can also derive the temporal differences in the occurrence of successful and unsuccessful queries. We can observe that some queries (e.g. *medical genetics*, and *menneskets fysiologi* [human physiology]) are most common in certain months. These peaks could be related to teaching, for instance at the medical faculty in those cases. Other queries, for instance for databases such as *atekst* or *pubmed*, are more commonly issued, and do not have large variations over time.

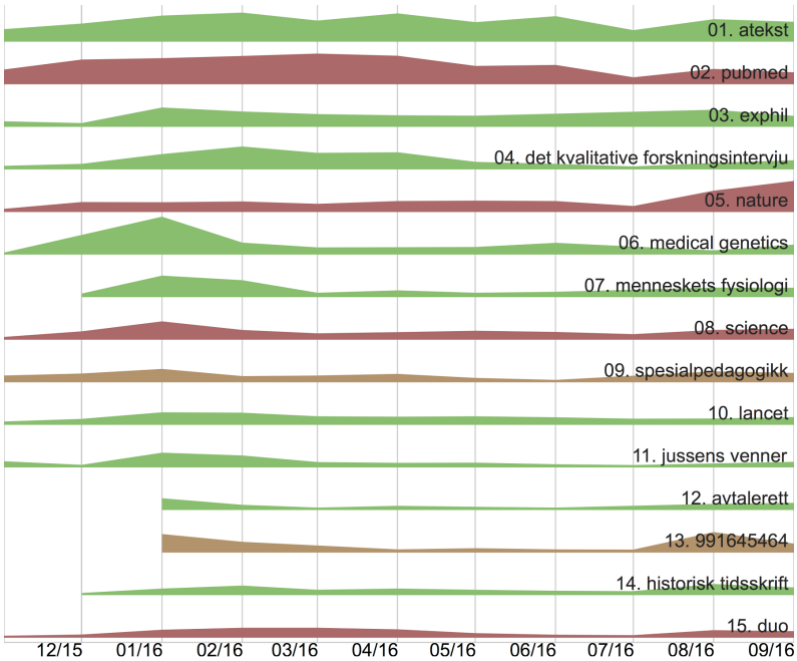

Figure 3. Proportional evolution of the query top 15 in the period November 2015 to September 2016. Green queries: successful, orange queries (09, 13): unknown, red queries (02, 05, 08, 15): unsuccessful.

The underlying reasons of failure

Next, we move on to the zero result queries, and try to determine the cause of their failure. In doing so, we look at our third research question (RQ3): *What underlying reasons for zero result searches can be determined?*

Figure 4 summarizes the reasons for acquiring zero results in a search. The most common source of zero results in our analysis (11 out of 50 times, or 22%) is that directly pasted citation references fail to locate the cited document, even if the document is actually available. An example is *Browning, N. (2015). The ethics of two-way symmetry and the dilemmas of*

dialogic kantianism. Journal of Media Ethics. Misspellings or reference mistakes are also a common cause of zero results (20%). For instance, the misspelled *why students underacheive* returns 0 results, while *why students underachieve* returns 463 results, but no spelling suggestion is given.

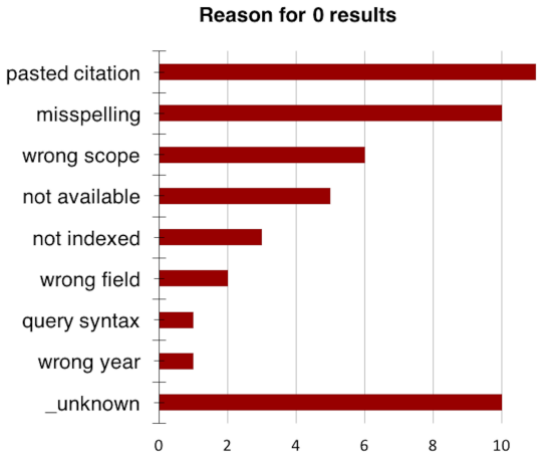

Figure 4 Reasons for zero results in our sample (number of occurrences)

At the time of this study, the Oria search interface had two scopes: *Bøker + artikler ved UiO* [books + articles at University of Oslo], and *Bøker i norske fagbibliotek* [books in Norwegian academic libraries]. From our analysis we can observe that this distinction can cause issues: for 12% of the zero result cases, there were no results in the currently selected scope, but there actually were some in another scope. To take a practical example, some users searched for a research article in the *Bøker i norske fagbibliotek* [books in Norwegian academic libraries] scope. In 10% of the cases, a resource is not available at all in Oria, and naturally would lead to no results – thus perhaps being the only "true" zero result searches. For another 6%, the resource is not indexed in Oria. This means the user searched for a journal article or book chapter, but only the host journal or book was indexed. In few cases, users searched in the wrong field (e.g. the author field for a title), or incorrectly used the query syntax (2%). Curious is the single case of the reference *svennevig j.: ledelsesretorikk i nedbemanningssituasjoner 2009* [management rhetoric in downsizing situations], which contains the wrong year and promptly leads to no results. For 20% of the queries, we were not able to determine exactly why the search system did not return any results.

Naturally, a search system such as Oria evolves over time since it is in constant development. This might mean that at a later point of time, zero result queries may actually return results. Therefore, we revisited the problematic queries within the zero results dataset in April 2017. We replicated the settings of the queries (e.g. the used field), and performed the queries again in the then-current system. This way, we observed that 36 (72%) of all zero results queries still returned zero results in the contemporary version of Oria. In 14 cases (28%), improvement was seen: the queries now returned (some) results. However, citations that include too many extra words, such as *og* [and], or 'bind' / 'volume' still frequently caused searches to fail. In addition, we observed that the zero results set contained many queries with misspellings, even though Oria has a built-in query correction (*"Did you mean…?"*) feature. To test the

effectiveness of the correction feature, we also checked if suggestions were provided. For only half of the misspellings, a correction was proposed. Hence, in some cases the spell-checker is helpful, but it also fails to catch various misspellings, for example in author names. As we will discuss further in the next section, improvement of these types of features could be helpful.

Discussion

The results of our analysis have shown that the Oria discovery system contains more than meets the eye: even though requested materials are available in the library, they do not always show. Our findings have various implications, which we discuss next.

In our analysis of both popular and zero result queries, we observed that users often do not find the resource they most likely intended to acquire. Aiding users in the **process of formulating queries** and executing their searches is thus essential. This could be achieved by providing more helpful features in the search system, and by enhancing current metadata. More helpful features could entail improved spelling suggestions, more specific and contextual suggestions (including autocomplete (Berget & Sandnes, 2016) and query suggestion features), automatic links to helpful functionality (e.g. the *request material* form), and information about potential results in other available scopes. Naturally, it would be best to prevent these dead-ends altogether, likely requiring changes in terms of the utilized search and retrieval algorithms by the vendor of the library system. In our study, we observed that queries for databases, but also for book chapters often fail in the Oria search system, which could also be remedied by enhancing the metadata using more in-depth cataloguing of these types of resources.

The availability and extension of usage data also implies various opportunities **to increase insights for library staff**. One possibility is to monitor and detect frequent issues in library searches. Library staff could receive alerts on sudden spikes of zero result queries, and take action if needed – besides query formulation issues, this paper shows that zero result queries can also originate from errors in curriculum lists and omissions in the library collections. We also encountered a number of issues with the used data. First, due to issues in *data quality*, cleaning and filtering is necessary. For instance, it is hard to distinguish in the current data between regular user queries, library staff queries, and queries in the context of library skills courses. Therefore, a better division between the regular search system and "test" system should be envisioned. In the surveyed discovery system, a sandbox/testing environment is already available, but at the moment of writing the retrieved result sets in general do not match the ones from the production system, reducing its usefulness. Second, there are limitations in terms of the *quantity* and scope of the current data. The coarse granularity of popular user searches in Primo Analytics makes it only possible to do broad analyses of the performed queries, preventing the ability to analyse series of query formulations during user sessions. Thus, the functionality of Primo Analytics could be extended to cater for these options, or custom additional logging has to be done, as for instance done by Hanrath and Kottman (2015).

An important way to potentially improve search success is the **integration of curriculum materials**. An important finding of our paper is that a considerable degree of the 50 most popular queries as well as the zero result searches sample can be traced to students' curriculums. We observed that users often appear to be copying and pasting references from curriculum lists into the Oria search box: full references constitute 22% of the analysed failed queries. To improve matters, the cataloguing of curriculum books could be improved by adding as much info as possible to make them searchable, including their table of contents. It would also be helpful to enable search for course materials using the distinct course codes assigned by the university to each course, which students are already familiar to using in other contexts.

Finally, there are implications for the **role of information literacy.** There is the aspect of "responsibility": proper searching is not just the responsibility of the system and of interface designers and content creators, but also of the user. Users, especially undergraduates just entering academia (Novotny, 2004), appear to be adopting similar search behaviour in library catalogues as in commercial search engines (Shiv, 2012; Willson & Given, 2010). The frequent failure of pasted citation references in Oria has shown that practical search skills, an important aspect of information literacy (Lloyd, 2010), are still essential for searching library discovery systems. Searching in a library system is not the same as searching in a popular online search engine, for instance due to the nature of bibliographic metadata, which is quite different from the full-text documents indexed by search engines (Blenkle, Ellis & Haake, 2015). In the context of library search systems such as Oria, it is important to emphasize to end-users that they may have to put more effort in actually receiving the "right" materials, including the need to do additional query reformulations and inspection of result pages. This additional effort may be worthwhile, though, as evidenced by Brophy & Bawden (2005), who conclude based on four case studies that Google was superior in terms of coverage and accessibility, but that the library system was superior in terms of the quality of the results.

The used data has some limitations, which can be addressed in future work. First, transaction log analysis has limitations, as for instance summarized by Dumais et al. (2014). While providing unobtrusive impressions of user behaviour, it is impossible to discern population statistics, such as age, education level or (academic) search experience from anonymous transaction logs. In future work, we plan to combine more qualitative data (for instance using interviews) with the more quantitative data from the transaction logs. Second, the process of coding involves decisions, which can be sometimes of a subjective nature, for instance in terms of the perceived intent of a user. Similar to Chapman et al. (2013), we tried to minimize this by involving multiple annotators. Third, limitations exist in direct relation to our dataset. The granularity is rather coarse (daily for zero results queries, and monthly for popular queries), and the reasoning behind having a limitation of 200–500 popular queries in Primo's analytics module is also unclear. Furthermore, session information is unavailable, so it is currently not possible to perform some types of transaction analysis (Jansen, 2006), specifically analysing the user sessions and query reformulations, as well as determining success at a session level. For further studies, we are currently gathering more specific anonymous statistics locally in the local Oria instance at the University of Oslo.

Conclusion

The first part of this paper's analysis examined the nature of popular and zero result searches in the discovery system at the University of Oslo by classifying them. Popular user queries typically contain few words, and there is a prevalence of queries for titles of specific resources, including books, journals and databases. Zero result queries, on the other hand, typically involve long queries, pointing to a more diverse set of materials, including journal articles, and in many cases consist of pasted reference citations. Many of the analysed queries were curriculum-related: 58% of the popular queries, as well as 28% of the zero result searches.

Our subsequent analysis of the success of popular queries showed that query success is not guaranteed: we observed a failure rate of 20% in the query top 50. Queries resulting in a higher number of returned items were more often unsuccessful, pointing to a need to improve ranking and potentially to "boost" important items in the result-set (so they appear higher in the list). The occurrence of popular queries varies over time, and we detected a distinction between queries popular at specific moments, for instance for reading list materials, and more regular queries occurring across time, including database queries.

Finally, our investigation of the reasons behind failed zero result searches suggested that together 42% of these user queries consisted of pasted citations and minor spelling errors, for which the system failed to retrieve items or suggest corrections. In our analysis, we rechecked the zero result queries in the contemporary system six months after the initial analysis, and saw that in 28% of the cases some improvement was seen, as the queries now did not return zero results anymore.

As our research has shown, not encountering a book or article in the analysed library discovery system often does not mean that it is not available, but just that the system has trouble finding it with the current search terms. This points to the need for increasing users' information literacy, but also for improvement of discovery system features, metadata and curriculum material integration. Besides the direct results of our analysis, this paper has also demonstrated the range of opportunities in using aggregated usage statistics for insights into user behaviour, and for potential improvement of the user experience. The paper also shows the need for refining currently available aggregated statistics for further research opportunities.

Acknowledgements

Our acknowledgements go out to the Digital Services department of the University of Oslo Library for support in acquiring and understanding Oria data. Also, to the National Library of Norway for co-funding the Visual Navigation Project (http://bit.ly/visualnavigationproject).

References

Berget, G., & Sandnes, F. E. (2016). Do autocomplete functions reduce the impact of dyslexia on information-searching behavior? The case of Google. *Journal of the Association for Information Science and Technology*, *67*(10), 2320–2328. https://doi.org/10.1002/asi.23572

Blenkle, M., Ellis, R., & Haake, E. (2015). Only the first results count: user-feedback-modified relevance ranking in E-LIB Bremen. *Insights*, *28*(2), 75–80. https://doi.org/10.1629/uksg.235

Borgman, C. L. (1986). Why are online catalogs hard to use? Lessons learned from information-retrieval studies. *Journal of the American Society for Information Science*, *37*(6), 387–400. https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28198611%2937%3A6%3C387%3A%3AAID-ASI3%3E3.0.CO%3B2-8

Borgman, C. L. (1996). Why are online catalogs *still* hard to use? *Journal of the American Society for Information Science (1986–1998)*, *47*(7), 493–503.

Brophy, J., & Bawden, D. (2005). Is Google enough? Comparison of an internet search engine with academic library resources. *Aslib Proceedings*, *57*(6), 498–512. https://doi.org/10.1108/00012530510634235

Chapman, S., Desai, S., Hagedorn, K., Varnum, K., Mishra, S., & Piacentine, J. (2013). Manually Classifying User Search Queries on an Academic Library Web Site. *Journal of Web Librarianship*, *7*(4), 401–421. https://doi.org/10.1080/19322909.2013.842096

Dumais, S., Jeffries, R., Russell, D. M., Tang, D., & Teevan, J. (2014). Understanding User Behavior Through Log Data and Analysis. In J. S. Olson & W. A. Kellogg (eds.), *Ways of Knowing in HCI* (p. 349–372). New York: Springer International Publishing.

ExLibris (2017, February 6). *Primo Popular searches.* Retrieved from https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Analytics/040Subject_Areas#Primo_Popular_Searches

Fast, K. V., & Campbell, D. G. (2004). "I still like Google": University student perceptions of searching OPACs and the web. *Proceedings of the Association for Information Science and Technology*, *41*(1), 138–146. https://doi.org/10.1080/10875301.2014.894955

Griffiths, J. R., & Brophy, P. (2005). Student searching behavior and the Web: Use of academic resources and Google. *Library Trends*, *53*(4), 539–554. Retrieved from http://hdl.handle.net/2142/1749

Han, H., Jeong, W., & Wolfram, D. (2014). Log Analysis of Academic Digital Library: User Query Patterns. In *iConference 2014 Proceedings* (p. 1002–1008). iSchools. https://doi.org/10.9776/14346

Hanrath, S., & Kottman, M. (2015). Use and Usability of a Discovery Tool in an Academic Library. *Journal of Web Librarianship*, *9*(1), 1–21. https://doi.org/10.1080/19322909.2014.983259

Hofmann, M. A., & Yang, S. Q. (2012). "Discovering" what's changed: a revisit of the OPACs of 260 academic libraries. *Library Hi Tech*, *30*(2), 253–274. https://doi.org/10.1108/07378831211239942

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management*, *36*(2), 207–227. https://doi.org/10.1016/S0306-4573(99)00056-4

Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research*, *28*(3), 407–432. https://doi.org/10.1016/j.lisr.2006.06.005

Kani-Zabihi, E., Ghinea, G., & Chen, S. Y. (2008). User perceptions of online public library catalogues. *International Journal of Information Management*, *28*(6), 492–502. https://doi.org/10.1016/j.ijinfomgt.2008.01.007

Lee, J. H., Renear, A., & Smith L. C. (2006). Known-Item Search: Variations on a Concept. *P Am Soc Inform Sci*, *43*(1), 1–17. https://doi:10.1002/meet.14504301126

Lloyd, A. (2010). *Information Literacy Landscapes: information literacy in education, workplace and everyday contexts*. Oxford: Chandos Publishing.

Lown, C., Sierra, T., & Boyer, J. (2013). How users search the library from a single search box. *College & Research Libraries*, *74*(3), 227–241. https://doi.org/10.5860/crl-321

Markey, K. (2007a). Twenty-five years of end-user searching, Part 1: Research findings. *Journal of the American Society for Information Science and Technology*, *58*(8), 1071–1081. https://doi:10.1002/asi.20462

Markey, K. (2007b). Twenty-five years of end-user searching, Part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, *58*(8), 1123–1130. https://doi.org/10.1002/asi.20601

Meadow, K., & Meadow, J. (2012). Search Query Quality and Web-Scale Discovery: A Qualitative and Quantitative Analysis. *College & Undergraduate Libraries*, *19*(2–4), 163–175. https://doi.org/10.1080/10691316.2012.693434

Moulaison, H. L. (2008). OPAC Queries at a Medium-Sized Academic Library. *Library Resources & Technical Services*, *52*(4), 230–237. https://doi.org/10.5860/lrts.52n4.230

Novotny, E. (2004). I don't think I click: a protocol analysis study of the use of a library online catalog in the internet age. *College and Research Libraries*, *65*(6), 525–537. https://doi.org/10.5860/crl.65.6.525

Niu, X., Zhang, T., & Chen, H. (2014). Study of User Search Activities With Two Discovery Tools at an Academic Library. *International Journal of Human–Computer Interaction*, *30*(5), 422–433. https://doi.org/10.1080/10447318.2013.873281

Shiv, K. (2012). Impact of internet search engines on OPAC users: a study of Punjabi University Patiala (India). *Program*, *46*(1), 56–70. https://doi:10.1108/00330331211204566

Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). Searching the web: the public and their queries. *Journal of the American Society for Information Science*, *53*(2), 226–234. Retrieved from https://pdfs.semanticscholar.org/01cc/4c3b2934d756f87cbc75c16c8716454b053b.pdf

Vaughan, J. (2011). Web Scale Discovery What and Why? *Library Technology Reports*, *47*(1), 5–11. Retrieved from https://journals.ala.org/index.php/ltr/article/view/4380/5065

Willson, R., & Given, L. M. (2010). The effect of spelling and retrieval system familiarity on search behavior in online public access catalogs: a mixed methods study. *Journal of the American Society for Information Science and Technology*, *61*(12), 2461–2476. https://doi.org/10.1002/asi.21433