Known-item searches resulting in zero hits: Considerations for discovery systems

Christiane Behnert & Dirk Lewandowski
Hamburg University of Applied Sciences
Department of Information
christiane.behnert/dirk.lewandowski@haw-hamburg.de

This is a preprint of an article accepted for publication in The Journal of Academic Librarianship.

Please cite as:
Behnert, C., & Lewandowski, D. (2017). Known-item searches resulting in zero hits: Considerations for discovery systems. The Journal of Academic Librarianship, 43(2), 128–134. https://doi.org/10.1016/j.acalib.2016.12.002

Abstract:

The goal of this article is to understand the reasons why known-item search queries entered in a discovery system return zero hits. We analyze a sample of 708 known-item queries and classify them into four categories of zero hits with regard to whether the item is held by the library and whether the query is formulated correctly: (1) Item in stock, but query incorrect, (2) Item not in stock, (3) Item in stock, but incomplete or erroneous metadata, (4) Query is ambiguous or not understandable. The main reasons for zero hits are caused by acquisition and erroneous search queries. We discuss possible solutions for known-item queries resulting in zero hits from the side of the system and show that 30% of zero hits could easily be avoided by applying automatic spelling correction. We argue that libraries can improve their discovery systems or online catalogs by applying strategies to avoid or cope with zero hits inspired by web search engines and commercial search web sites.

Keywords: Discovery system; known items; known-item search; query analysis; zero hits

Introduction

Web search engines in general and Google, in particular, are popular tools in information seeking today. Users perceive web search engines to be successful in providing satisfactory results and therefore to be successful in satisfying their information needs (Purcell, Brenner, & Rainie, 2012). One of the major reason for that might be that a significant number of search queries can be identified as being navigational, i.e., queries intended for a known website or a website assumed to exist (Broder, 2002). Studies show that proportions of navigational queries entered into web search engines range from 11.7 to 15.3 percent (Rose & Levinson, 2004) to 20-24.5 per cent (Broder, 2002), even to an average proportion of 40 per cent (Lewandowski, 2006) or from 27 to 42 percent (Lewandowski, Drechsler, & Mach, 2012). Such queries are answered rather easily because they demand the one correct result only, such as the URL to one's Facebook profile or to an online banking website. In contrast, topical or informational search queries usually produce a set of results ranked according to the system's ranking algorithm because there usually are several documents needed in order to help users solve their search tasks. When navigational queries are entered, the user's search intention is driven by a *concrete* information need (CIN); in contrast, informational queries are intended to satisfy a *problem-oriented* information need (POIN) (Frants, Shapiro, & Voiskunskii, 1997).

Users expect library catalogs to be as efficient as web search engines. Their information behavior today is strongly influenced by the comfortable search features and easily accessible information that search engines provide. If CINs are, in fact, being successfully satisfied by web search engines to a large degree (Lewandowski, 2011, 2015), i.e., the success rate at a particular position as the appropriate measure is large (Craswell & Hawking, 2005), then libraries must improve their retrieval systems accordingly to give information seekers a similar experience in helping to satisfy their information needs.

The library's ability to provide the (one) correct information object has been a measure in library evaluation for decades. Several availability studies published from 1934 until 2006 were reviewed by Mansbridge (1986) and Nisonger (2007). The availability study by Kantor (1976) has led to a commonly accepted technique by the name of "the branching method". The author identifies four "principal categories of dissatisfaction" (Kantor, 1976): *acquisition, circulation, library error* and *user error*. He also provides a "check list [sic] to guide the investigators in tracking down the books which the user is unable to find" (Kantor, 1976, p. 313):

1. *Not owned by the library,*
2. *Incorrect call number,*
3. *Book located in a special area identified on catalog card,*
4. *Book properly shelved,*
5. *Book misshelved,*
6. *Book recently used or in use in library,*
7. *Book in preshelving,*
8. *Book located in (or listed in) an area not identified on card catalog,*
9. *Book circulating,*
10. *Other.*

Although originally developed with respect to printed books, both the categories and checklist can be adapted to digital materials accessed through libraries nowadays. As Kaske (1994) states, the concept of availability is not bound to physical walls anymore because modern library information systems use internet technologies that allow search in and access to multiple, heterogeneous information sources. In fact, considering user satisfaction with item availability is an important issue still reflected by new library information systems with implemented search engine technologies. Not only do such systems, known as discovery systems, claim to provide *discovery* of diverse resources, they also offer *delivery* of resources.

Studies on the availability of electronic materials revealed that, based on Kantor's branching method, non-availability is often caused by lacks in licensing or linking problems (Nisonger, 2009), or incomplete or inaccurate metadata (Crum, 2011). The error types of access are covered by Kantor's four categories of dissatisfaction (acquisition problems, circulation, library error, user error). Although circulation problems with digital materials are assumed to be excluded, there are licensing models that do not allow simultaneous "use" of the same e-book. In this case, a user may be frustrated because he or she would not be able to access the information object.

Non-availability of items is not only demonstrated at that moment a search result is selected from a list in order to access it but also at the moment when the entered search query yields (a) not the correct result or (b) no results at all. In both cases, the user would be frustrated with zero hits, which is understandable considering web search engines' ability to provide any partially matched result or to detect and correct possible user errors. In order to improve library information systems, we need a better understanding of the problems with known-item searches resulting in zero hits.

In this paper, we show what errors occur in known-item searches and give possible solutions to the zero hits problem from the side of the system. We analyzed original search queries entered into a discovery system (n=1,981), identified known-item queries, and investigated the reasons for zero hits. We demonstrate that a significant proportion of known-item search queries could be successfully processed by the information retrieval (IR) system by applying syntactic reformulation strategies used by web search engines.

The rest of the paper is structured as follows: In the next section, we review the literature on identifying known-item searches and on how to cope with zero search results. We describe the sample of search queries we used for our analysis and how the queries were classified. After that, we present and discuss the reasons for known-item queries yielding zero hits, and propose possible solutions. We conclude with a summary of the findings.

Literature review

There is no standard definition of known-item searches and interpretations of the concept behind this type of search vary. In general, the overall proportion of known-item search queries in library information systems is as significant as the number of navigational queries entered in web search engines, as we describe below. Libraries must consider retrieving the correct search result for a known-item search query in order to improve their modern information systems' retrieval effectiveness and thus, to satisfy users' CINs. They also must provide immediate access (delivery) to digital information objects.

Identifying known-item searches

Studies that involve definitions of known-item searches show that most researchers define a known-item search intuitively. This is a problem because it compromises the comparability and validity of findings. Lee, Renear, & Smith (2007) found that the concept of known-item search is rather complex and varies. They discuss the concept of known-item search in library and information science systematically in comparison with operational definitions; they state that there is a difference between the concept behind known-item search and the applied search strategy to retrieve the item sought. For example, Kan & Poo (2005) identified known-item searches as follows: "When users seek to find specific resources in a digital library, they often use the library catalog to locate them" (Kan & Poo, 2005). In their study, they extracted queries from Online Public Access Catalog (OPAC) log files and determined characteristics to identify known-item queries in contrast to other query types (e.g., the query length that is above average, the presence of determiners, and matching orthographic cases of titles). They then integrated these characteristics into an automatic approach that can distinguish known-item queries from other queries. Building on that approach, Khabsa, Wu, & Giles (2016) applied a similar method to classify navigational queries in the context of academic search. They identified a query as navigational according to seven features, including the number of query terms, the presence of stop words and punctuation as indicators for a title search. These features are only partially in consistency with the ones analyzed by Jansen, Booth, & Spink (2008), who developed an automatic approach to classifying informational, navigational and transactional web search queries. This may partly be because the studies do not consider search in an academic context. For example, one characteristic of navigational queries is that the queries are short with usually less than three words. This can be explained with navigational web queries that are assumed to be intended for going to, for example, a company's homepage; in academic search, a large number of query terms usually would indicate a full citation or at least a title. Such characteristics help us to identify a certain type of search query, but they cannot provide a definition of a known-item search from a conceptual perspective.

Besides a differentiation between the concept of known-item search and an operational definition, another issue in defining known-item searches is the condition of existence a known item is assumed to have, regardless of whether the item sought is actually held by the particular library: the user knows the item exists because she may have used the item before or she just assumes the item exists (e.g., a textbook introducing statistics to students). Allen (1989) defines a known-item search as a search for an item the user has read (or accessed) before, thus, knows that it exists. Thus, "the user's memory of the item is of consequently primary importance" (Allen, 1989). When Wildemuth & O'Neill (1995) interviewed 160 catalog users in an academic library, they found that 64% of them conducted a search for a known item, which they specified as an item the participants knew existed (they were prepared for searching with some bibliographic data) but had not necessarily read it before. They also observed errors in known-item searches mainly being typographical errors or misspellings. Other common errors occurred because of the user's failure (1) to specify a search mode, (2) to not invert the author names (instead of "surname, given name", they entered "given name surname"), and (3) to include initial articles in queries entered in title search fields (e.g., "The", "An"). These errors usually would not cause any problems in today's search engines or modern library information systems, as these systems follow best or partial match approaches and offer a single search interface.

Proportion of known-item searches

Considering the different definitions of known-item search queries and the different approaches in identifying them, it is not surprising that proportions of known-item queries differ accordingly. Slone (2000) interviewed and observed public library patrons using the online catalog, which led them to categorize three different types of searches: "[The patrons] who appeared to know, prior to the search, the specific item(s) they sought were identified as known-item searchers. Conversely, those who did not know the specific items they sought were classified as either unknown-item searchers or area searchers. […] Area searchers […] used the on-line system to find a library section or area" (Slone, 2000). The searches for known items and areas accounted for 8 searches each of the observed 36 searches in total, while 20 were identified as non-known-item searches.

In contrast to this small sample and the then technological state of OPACs, we can consider the findings of the study by Chapman et al. (2013) as more applicable to today's library information systems and their users. They classified a final sample of 992 queries extracted from a library website's integrated search box: within the final classification scheme known-item searches, i.e., queries "in which the patron entered the title of something that was recognized as a unique, identifiable object" (Chapman et al., 2013, p. 411), yielded 44%. Other queries were categorized as non-known-item searches (28%), i.e., "topics, authors, and anything too vague to categorize as a known-item search" (Chapman et al., 2013, p. 412). 25% of all queries were marked as non-categorized because of disagreement between the jurors, leaving 3% as non-classifiable due to lack of information. Interestingly, the queries were classified considering a differentiation between *item type* (what was searched for, e.g., article or book as known-item searches) and *query type* (how was searched for, e.g., title or ISBN, only applicable to known-item searches). A search query that only consists of an author name was classified as a non-known-item search because sole author names do not provide a reliable source for being clearly and consistently classifiable as a known-item search query; instead, the type *creator* was used for author queries.

How to cope with zero hits?

In general, reasons for search queries resulting in zero hits can be seen from two different points of view – either the user is not able to formulate the query correctly or the system is not able to retrieve results that match the query because of operational limits, e.g., a limited number of multiple data sources or failures in metadata. As an IR system's purpose is to help users in their information seeking, both views on the problem of zero hits need to be taken into account.

To avoid zero hits the user can choose advanced search features such as selecting the correct search mode or field, e.g., the title field for entering title keywords or the full title copied from a reference list instead of the search in the full text. Common query reformulation strategies may be helpful with known-item searches. However, if we assume that a known-item search query is based on correct bibliographic or location data, then reformulating a query seems to be more effective with informational queries or topic searches, respectively. Nonetheless, discovery systems should provide certain features to correct errors in queries, for example spellchecking, "Did you mean…?"

suggestions, and an auto-complete algorithm (Chickering & Yang, 2014). These functions are important for retrieving search results according to users' expectations and their search behavior, which is highly influenced by web search engines.

In addition to implemented features for query (re-)formulation to avoid zero hits, with regard to user needs "the zero results page is an opportunity to provide support in the form of advice and tools for query reformulation" (Russell-Rose & Tate, 2013, p. 148). When it comes to search design issues regarding user experience, we can find inspiration in the context of e-commerce and commercial search websites. How to deal with zero hits pages is an important issue in order to face the competition. In his guide on user experience strategies, Nudelman (2011) presents four design principles for implementing a zero search results page (Nudelman, 2011, p. 5):

(1) Don't be afraid to say *I did not understand.*
(2) Focus on providing a way out.
(3) Create a robust partial match strategy.
(4) Employ multiple content strategies.

Although these principles do not specify the search intents or query types, they do provide a general understanding of how to deal with zero hits when they actually occur (principles 1, 2), regardless of what can be done to avoid zero hits at all (principles 3, 4).

A major purpose of discovery systems is to employ searching in multiple information sources, including to access local library holdings as well as licensed documents. Electronic documents, for example, can be accessed via link resolvers. How known items can be effectively retrieved was studied by Bailey & Back (2008). They evaluated retrieving known items using the link resolver LibX using Google Scholar's service. They found that the item sought could be accessed by the user with a single click in 81% in the sample because results that match the query are displayed to the user including a link to the particular OpenURL link resolver. LibX is an open source Firefox extension that reacts to zero hits before displaying the results page to the user as follows: a) removing the term that could not be found, b) reformulating after identified spelling mistakes, c) searching for terms as a phrase using double-quotes instead of default AND operators. Further, Mann (2015) observed an increasing electronic resource availability with the creation of standards for OpenURL linking. A "branching" diagram inspired by Kantor was applied to compare the results of two studies – before and after fixing access errors the initial studies had revealed, whereas the availability rate could be improved by 24%.

To sum up, the problem of known-item availability or accessibility in libraries has been recognized. Despite the fact that there is no standard definition of what characterizes a known-item search query, past studies describe successful approaches in increasing item availability. Further, commercial web search tools cope with queries that yield zero hits with respect to user experience and search design. Considering that the information-seeking behavior of today's library users is highly influenced by their behavior during web search, it seems appropriate to gain a deeper understanding of the reasons why known-item queries yield zero hits.

Methods

We extracted 2,000 search queries from transaction log files that were entered in the single search interface of EconBiz, in July 2014. EconBiz is an information portal for economic literature provided by the German National Library for Economics (ZBW – Zentralbibliothek für Wirtschaftswissenschaften)[1]. It has implemented the open source discovery software VuFind and combines data from multiple heterogeneous sources. Based on the queries from the log files we built a sample of 708 known-item search queries which we used to evaluate the ability of the system to produce correct results.

Sample building
We classified 1,981 search queries[2] as topic search queries, known-item searches, and person searches (only author names), respectively. Each of the queries was restricted to one category (see Figure 1). The fourth category of non-classifiable queries was added because a small number of queries (2%) could be identified as neither topic nor known-item search nor person search, for example, unidentifiable abbreviations. The majority of queries (51%) could be identified as topic searches; the second largest number of queries (38%) was assigned to the category of known-item queries. The number of searches for an author (9%) is still large in relation to queries consisting of author names in combination with other types of metadata; we assigned those to known-item search queries.

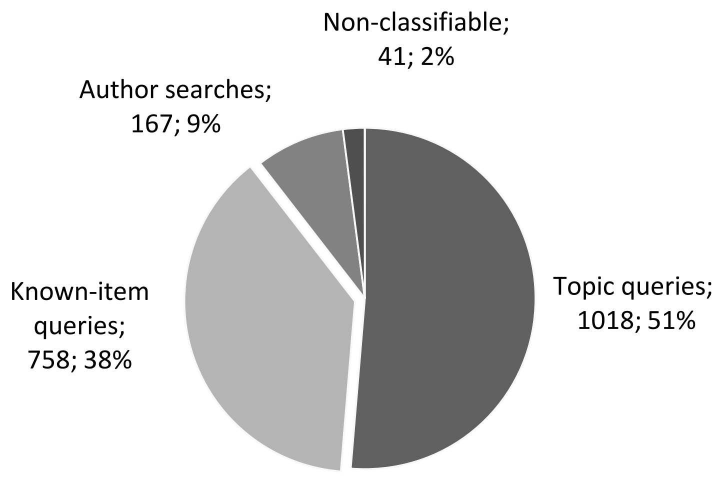

Figure 1. The proportion of query types.

Since we were only able to extract the queries from the transaction logs without session data, we did not have information about which results the user clicked. Thus, we could only assume a query to be a known-item search based on the terms entered into the search box. This means that we identified known-item queries from the operational point of view rather than a conceptual view (Lee et al., 2007). Queries were categorized as known-item queries if the search terms could be identified as (a) author name in combination with title or title keywords, (b) author name in combination with a year of publication, (c) author name in combination with title or title keywords and year of publication, (d) author name in combination with title or title keywords and a number of edition, (e) full title or title keywords, (f) title or title keywords in combination with a year of publication, (g) journal title, and (h) journal title in combination with a paper's title or title keywords. Another category was needed for queries classifiable as known-item queries but without assigning those to one of the other categories because they included searches for library internal record ID that we assumed had been performed by library professionals (see Figure 2). These 50 known-item queries in the Category "Other known-item searches" were excluded from the final sample. We performed searches with 708 of the 758 classified known-item queries in EconBiz.

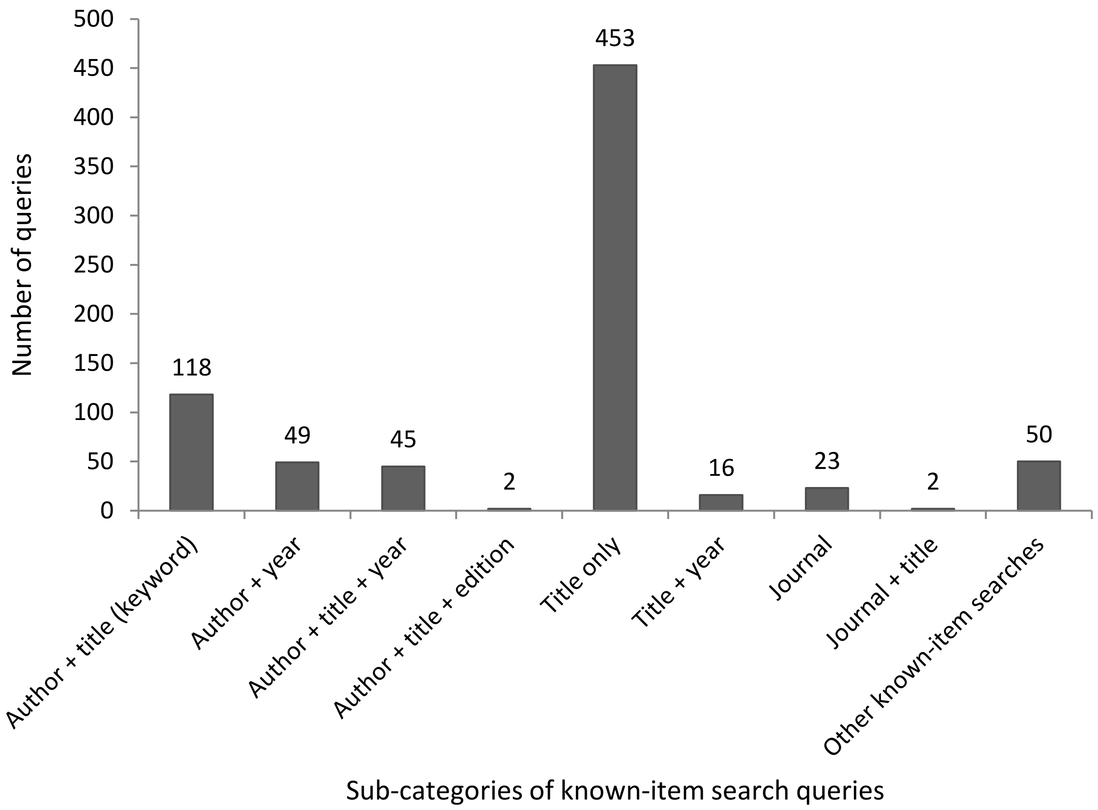

Figure 2. Sub-categories of known-item search queries.

Testing known-item queries

We analyzed the results of known-item queries within two test runs. In the first run, all 708 queries were entered into the search interface of EconBiz, whereas every query URL was copied and pasted into an Excel sheet. In this table sheet, the number of all search results each query yielded was added. Then the queries that produced at least one result were analyzed by two student assistants following these instructions: (a) the URL of the assumed to be the correct result was copied into the table sheet; (b) in the case of no result, a zero was coded. A result was judged to be correct if it matched all query terms or a number of terms were assumed to be suitable. For queries that yielded more than one matching result, such as different editions of a book, the URL of the most recent item was judged as the correct result. We considered the top 30 results because we assumed that possible matches for a query would not be ranked below that position.

In the second run, we analyzed all queries that yielded zero hits and determined the reasons for queries yielding no results at all, or queries yielding no correct result within the top 30 results of the result set, respectively. To every one of those queries, it was noted why no results could be found. For example, an item sought could not be found because it is not in stock or the item sought is published in a journal that is not licensed by the library. Thus, we built categories of reasons for zero hits, which lead us to propose possible solutions for the IR system.

Results and Discussion

Proportion of zero hits

Only 394 (56%) of the 708 queries in the sample yielded correct results, while in 314 cases (44%) zero hits occurred. Of these zero hits queries, 172 did not return any results at all, leaving 142 queries that yielded at least one result but none that would match enough query terms (see Figure 3).

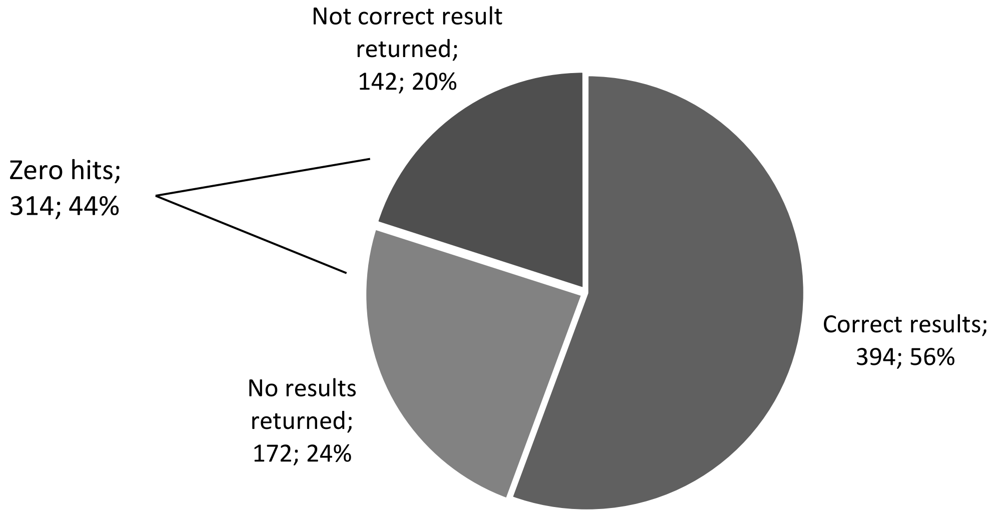

Figure 3. The proportion of known-item queries yielding correct results and zero hits (n=708).

With regard to this quite large proportion of queries yielding no results at all the main question that arose was, whether the assumed to be known items could not be found simply because they had not been purchased by the ZBW (because of its acquisition profile focusing on economic literature), or whether there were other reasons.

Reasons for zero hits

In order to analyze the reasons for 314 known-item queries resulting in zero hits, each of them was researched in additional information systems, e.g., library catalogs (to verify books) or academic search engines (to verify articles and other types of documents). Thus, query errors and items not in stock could be detected; they were marked in the table sheet. Within this zero hits analysis, each of the 314 queries was inductively assigned to one category of reasons why they yielded zero hits. In the end, we determined ten individual reasons for zero hits queries that could be grouped into four categories. Figure 4 shows the four categories of reasons for zero hits and the number of queries assigned to each category: (1) With regard to 95 queries (30%) the item is in stock, but the query was erroneous, (2) in 125 cases (40%) the item is not in stock, (3) for 62 (20%) queries the assumed to be known item is in stock, but because of incomplete or erroneous records they cannot match the query, (4) 32 queries (10%) are ambiguous and it is not clear, which of the possible items is the one that is sought.

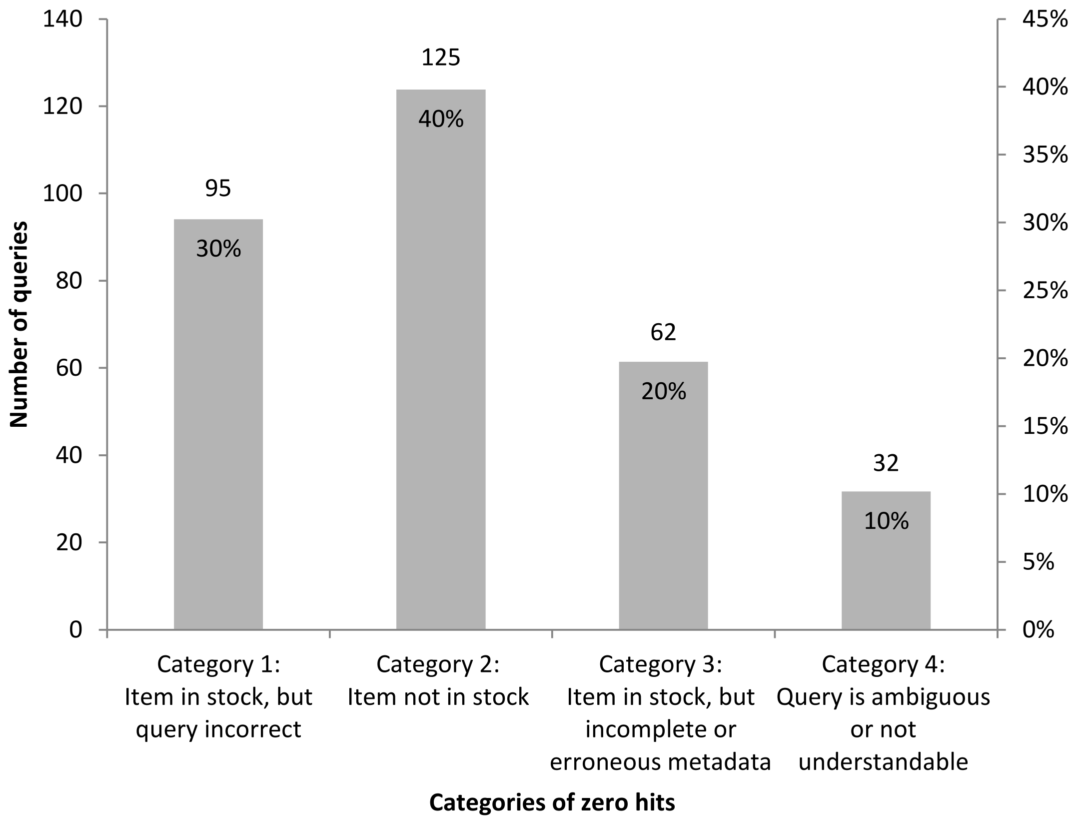

Figure 4. The number of queries resulting in zero hits with regard to each category (n=314).

In Table 1 the reasons why known-item queries resulted in zero hits are presented. While every main category combines three sub-categories, Category 4 includes all queries that are assumed to be known-item queries but that are not precise: Although several records match the query, it is not clear, which one of them is supposed to be the sought item, i.e., the query is ambiguous. Other queries contain abbreviations that require either intellectual efforts researching the item the user supposedly sought, or the user would be required to provide some context information. This category, however, has the lowest number of queries.

Table 1.
Reasons behind queries resulting in zero hits.

Category	Sub-category	Definition	Number of queries
1) Item in stock, but query incorrect	query_too_extensive	Some search terms cannot be processed (e.g., page numbers, edition).	23
	detectable_without_x	Some characters cannot be processed (e.g., punctuation marks, brackets).	32
	typo	The query includes typos or words that are not part of the title (small deviation).	40
2) Item not in stock	different_subject	Item not in stock, as it does not comply with acquisition profile (economics).	34
	not_EB	Item is not (yet) in stock.	87
	only_older_edition ; only_newer_edition	Another edition of the work is available.	4
3) Item in stock, but incomplete or erroneous metadata	only_compilation	No single record of the item available.	12
	only_periodical		42
	error_in_EB	False or missing data in the record (e.g., another editor).	8
4) Query ambiguous or not understandable	imprecise_query	It is unclear, which item is sought, as false terms or abbreviations are used by mistake.	32

The main reason why queries yielded zero hits is because the item was not acquired by the library (Category 2), i.e., either a book was not purchased or a journal is not licensed. In 87 cases, the item plainly is not in stock, though not because its purchase would not be compliant with the library's acquisition profile. Thirty-four queries that would return items of a non-economic subject were assigned to another sub-category. The second-largest category contains queries that are erroneous (Category 1). Most of the errors were caused by typos, while other queries consist of (a) characters (e.g., question marks) the IR system could not process, or (b) too many, unnecessary terms (e.g., page numbers or edition information) presumably taken from citations. In contrast, Category 3 comprises queries that are well formulated but cannot return the sought item. One reason for this is that the record that would be the correct one is erroneous (e.g., typos) or not available (e.g., not all editors listed), as detected in 8 records. Another reason is that the queries aimed at articles (42 queries) or books (12 queries) for which no single record was available. This is caused by (inherited) traditional cataloging rules, i.e., records of books and journals are listed as collective volumes or periodicals, respectively. Thus, queries in search for an article or part of a collective volume require another bibliographic database or academic search engine for finding bibliographic information in order to use the library catalog for finding location information.

Approaches to avoid and cope with zero hits

The reasons behind known-item queries yielding zero hits are not very diverse. In fact, the four categories reflect three of Kantor's categories of dissatisfaction, which are library error (Categories 1 and 3), acquisition (Category 2) and user error (Category 4), whereas dissatisfaction due to circulation is not an issue here, since we focus on the ability of the IR system to retrieve known items, i.e., from the retrieval perspective, rather than consider acquisition or licensing issues.

In Table 2 we present the actions required to each category and give possible solutions from the side of the IR system, in order to increase the number of (correct) known-item search results.

Table 2.
Possible solutions for queries resulting in zero hits.

Category	Kantor's categories of dissatisfaction	Required action	Responsibility	Possible solution by IR system
1) Item in stock, but query incorrect	Library / user error	Query reformulation (syntactical)	Information system or user	Spell checking / spelling correction; partial match approach
2) Item not in stock	Acquisition	Alternative offer	Library (acquisition)	Link to local or national library network (consortium); link to alternative full-text sources; display acquisition request form
3) Item in stock, but incomplete or erroneous metadata	Library error	Correction / adjustment of metadata	Library (cataloging)	Link to alternative bibliographic database or academic search engine; partial match approach
4) Query ambiguous or not understandable	User error	Query reformulation (semantic)	User	Display search interface and link to help page

Although queries in Category 1 could easily be reformulated by the user himself, a web-scale library information system, which is expected to provide discovery and delivery of information objects, must be able to detect and correct all syntactical query errors. In general, focusing on a "robust partial match strategy" (Nudelman, 2011, p. 5) is necessary to cope with problems due to query errors.

Another of the library's responsibilities lies within alternative ways to provide access to electronic materials or information on the local availability of printed books and journals, regardless of whether the item is not in stock (Category 2) or the library records are not correct (Category 3). As we described in the literature review section, studies showed that availability (delivery), which also requires the retrievability (discovery) of results, increases by employing link resolvers. Again, we can relate this to Nudelman's design principles, as the fourth one is to "employ multiple content strategies" (Nudelman, 2011, p. 5). However, these three categories indicate that libraries must deal with zero hits by combining different strategies.

In contrast, queries in Category 4 must be semantically reformulated, and it would be the user's responsibility to re-enter the corrected query. The library information system should offer support, such as presenting FAQ, a link to a help page, and displaying the search interface. This complies with the "providing a way out" search design strategy followed in e-commerce search (Nudelman, 2011, p. 5). Another useful approach for the system is to suggest similar queries entered by other users.

In the case that sought items are in stock but cannot be found because of query errors (Category 1), we can conclude that the IR system's best or partial match approach is not developed to its full potential with regard to known-item search queries. Solving this problem would lead to an increased number of successful known-item searches, as the following example shows: We manually reformulated the 95 queries in Category 1 (13% of the sample), i.e., queries that are too extensive for the system to process (e.g., "Research methods for business : a skill-building approach Uma Sekaran and Roger Bougie (sixth edition)"), queries that would yield results with certain characters cut off (e.g., "A rating agency for Europe - A good idea?"), and queries containing typos (e.g., "A Dynamie Attribute Satiation Model of Variety-Seeking Behavior"). After removing misspellings or unnecessary characters, and correcting typos, 219 zero hits (31%) remain (see Figure 5). We note that the number of queries returning no results at all (14%) in proportion to the number of queries that did not return the correct result (17%) is inversely proportional to the numbers before the queries had been reformulated (see Figure 3), where 24% of queries did not return any results and 20% of queries did not return the correct one within the result set.

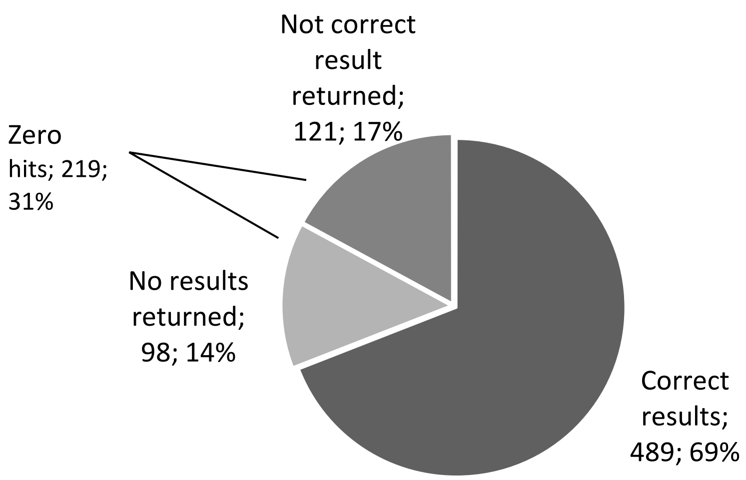

Figure 5. Proportion of known-item queries returning correct results and zero hits after re-analysis of queries in Category 1 (n=708).

Conclusion

The goal of this article was to understand the reasons why known-item queries result in zero hits. We built a sample of 708 known-item queries and analyzed their retrieved results using an information portal for economic literature, EconBiz. Of the 708 queries, 314 queries (44%) yielded zero hits, i.e., queries that did not yield any results and queries that did return some results but not the correct one. We identified ten different categories of reasons for zero hits that we combined into four categories. The main reasons for zero hits are caused by acquisition problems (125 queries or 40% of all zero hits) and erroneous queries entered by the users (95 queries or 30% of all zero hits). We showed that queries of the latter category could be corrected and retrieve the sought item by 13% by applying syntactical reformulation strategies. Although we manually re-entered 95 erroneous queries, we are convinced that modern libraries could improve their information systems with retrieval methods and zero hits strategies adapted from web search engines to a large degree. This would help today's library users satisfy their concrete information needs more effectively.

The results of our research go well beyond the library context. Information retrieval systems of any sort could profit from analyzing and improving on zero-hits queries. Users would profit from a reduced number of queries yielding no results, and also from the improved results for their known-item queries. Further research should focus on ways of how to help users experiencing zero-hits results pages, e.g., through making explicit when an item is definitely not in the collection vs. an item probably not being found due to an ill query.

Acknowledgements

Findings presented in this article were partly examined within the research project *LibRank – New Approaches to Relevance Ranking in Library Information Systems*. The project was funded by the German Research Foundation (DFG - Deutsche Forschungsgemeinschaft) from March 2014 until February 2016.

References

Allen, B. (1989). Recall cues in known-item retrieval. *Journal of the American Society for Information Science*, *40*(4), 246–252. https://doi.org/10.1002/(SICI)1097-4571(198907)40:4<246::AID-ASI4>3.0.CO;2-Z

Bailey, A., & Back, G. (2008). Retrieving known items with LibX. *The Serials Librarian*, *53*(4), 125–140. https://doi.org/10.1300/J123v53n04_09

Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum*, *36*(2), 3–10. https://doi.org/10.1145/792550.792552

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

Chickering, F. W., & Yang, S. Q. (2014). Evaluation and Comparison of Discovery Tools: An Update. *Information Technology and Libraries*, *33*(2), 5–30. https://doi.org/10.6017/ital.v33i2.3471

Craswell, N., & Hawking, D. (2005). Overview of the TREC-2004 Web Track. In *The Thirteenth Text Retrieval Conference (TREC 2004)*. National Institute of Standards and Technology. Retrieved from http://trec.nist.gov/pubs/trec13/papers/WEB.OVERVIEW.pdf

Crum, J. A. (2011). An availability study of electronic articles in an academic health sciences library. *Journal of the Medical Library Association : JMLA*, *99*(4), 290–296. https://doi.org/10.3163/1536-5050.99.4.006

Frants, V., Shapiro, J., & Voiskunskii, V. (1997). *Automated information retrieval: theory and methods*. San Diego: Academic Press.

Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management*, *44*(3), 1251–1266. https://doi.org/10.1016/j.ipm.2007.07.015

Kan, M.-Y., & Poo, D. C. C. (2005). Detecting and supporting known item queries in online public access catalogs. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries - JCDL '05* (pp. 91–99). New York, New York, USA: ACM Press. https://doi.org/10.1145/1065385.1065406

Kantor, P. B. (1976). Availability analysis. *Journal of the American Society for Information Science*, *27*(5), 311–319. https://doi.org/10.1002/asi.4630270507

Kaske, N. K. (1994). On My Mind: Materials Availability Model and the Internet. *The Journal of Academic Librarianship*, *11*(5–6), 317–318.

Khabsa, M., Wu, Z., & Giles, C. L. (2016). Towards better understanding of academic search. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries - JCDL '16* (pp. 111–114). New York, New York, USA: ACM Press. https://doi.org/10.1145/2910896.2910922

Lee, J. H., Renear, A., & Smith, L. C. (2007). Known-Item Search: Variations on a Concept. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1–17. https://doi.org/10.1002/meet.14504301126

Lewandowski, D. (2006). Query types and search topics of German web search engine users. *Information Services & Use*, *26*, 261–269.

Lewandowski, D. (2011). The retrieval effectiveness of search engines on navigational queries. *Aslib Proceedings*, *63*(4), 354–363. https://doi.org/10.1108/00012531111148949

Lewandowski, D. (2015). Evaluating the retrieval effectiveness of web search engines using a representative query sample. *Journal of the Association for Information Science and Technology*, *66*(9), 1763–1775. https://doi.org/10.1002/asi.23304

Lewandowski, D., Drechsler, J., & Mach, S. (2012). Deriving query intents from web search engine queries. *Journal of the American Society for Information Science and Technology*, *63*(9), 1773–1788. https://doi.org/10.1002/asi.22706

Mann, S. (2015). Electronic resource availability studies: An effective way to discover access errors. *Evidence Based Library and Information Practice*, *10*(3), 30–49. https://doi.org/10.18438/B88C82

Mansbridge, J. (1986). Availability studies in libraries. *Library and Information Science Research*, *8*(4), 299–314.

Nisonger, T. E. (2007). A review and analysis of library availability studies. *Library Resources & Technical Services*, *51*(1), 30–49. https://doi.org/10.5860/lrts.51n1.30

Nisonger, T. E. (2009). A simulated electronic availability study of serial articles through a university library web page. *College & Research Libraries*, *70*(5), 422–445. https://doi.org/10.5860/crl.70.5.422

Nudelman, G. (2011). *Designing search: UX strategies for ecommerce success*. Indianapolis, IN: Wiley.

Purcell, K., Brenner, J., & Rainie, L. (2012). Search Engine Use 2012. Retrieved from http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf

Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th conference on World Wide Web - WWW '04* (pp. 13–19). New York, New York, USA: ACM Press. https://doi.org/10.1145/988672.988675

Russell-Rose, T., & Tate, T. (2013). *Designing the search experience: The information architecture of discovery*. Amsterdam: Elsevier.

Slone, D. J. (2000). Encounters with the OPAC: On-line searching in public libraries. *Journal of the American Society for Information Science*, *51*(8), 757–773. https://doi.org/10.1002/(SICI)1097-4571(2000)51:8<757::AID-ASI80>3.0.CO;2-T

Wildemuth, B. M., & O'Neill, A. L. (1995). The "known" in known-item searches: Empirical support for user-centered design (research note). *College & Research Libraries*, *56*(3), 265–281. https://doi.org/10.5860/crl_56_03_265

[1] http://www.econbiz.de/

[2] Due to a technical error 19 queries from the log files had been irreversibly removed.