

Query formulation and reformulation in various information retrieval systems for academic purposes \ Shelly Shen-Aridor

Abstract

Information seeking, even though being part of our ordinary daily life routine, is a complex process.

Research demonstrates that successful search depends on several elements, including data structure parameters (e.g. full text, thin metadata, thick metadata), the information retrieval system and the end-user's capabilities.

Research in the field of information searching involves exploring various methods and techniques designed to improve information retrieval. The ultimate goal is to improve the way the search is performed (e.g., algorithms, advanced and specialized search options) and to design improved user interfaces (e.g., intuitive access to indices, easy to use search boxes). Information retrieval systems will continue to improve and enable more efficient search, based on the semantics of search terms and associative contexts. However, in the foreseeable future, the end-user will continue to be a pivotal player in the information search process. It is only the end-user who is able to define his information needs and evaluate the search results (e.g. relevancy, coverage, authority).

Clever use of information systems (e.g. correct phrasing of information needs, compatible with the information retrieval system, usage of unique or advanced search options, usage of effective search tactics) is expected to yield better success in the search process (e.g. results relevancy, time dedicated to search). It is therefore important to evaluate the impact of the end user's behavior on the search process in general, and on

successful queries in particular. This study belongs to the field of research concerned with analysis of user queries and the relevance of their results, from the user's perspective.

This research is focused on information searching for the purpose of writing academic papers among graduate students. The study population included sixteen graduate students from the Department of Information Science at Bar-Ilan University, who conducted searches for their papers. Given their education and professional background, these students were expected to have better than average information searching skills. The rationale was that this study population could provide insights into the difficulties of the information search process. In addition, possible avenues to improving the process would be identified. Any problems apparent within this population would likely to be present in other populations.

In this study qualitative methodology was employed to analyze and match between queries and specific students, considering their academic research question/topic and relevancy of the results. In addition, we compared the search progress and query success rate per user, across different databases. Each user conducted a single topic search by using three different information retrieval systems. The databases used were: Bar-Ilan University library catalog, Google-Scholar and Proquest databases. This comparison enabled the assessment of the impact of the retrieval system and its special search options, on the success rate. To date, there are few such qualitative studies.

Based on the informants queries (using screen capture software), the think aloud recordings and data from the questionnaires, we sought to answer the following questions:



- Are there clear characteristics of failed queries?
- What is the evolutionary process of the queries reformulation?
- Are there differences in attitude of the users to the various databases?

Addressing the first research question, several characteristics of failed queries (defined as those queries that did not generate any items the user expressed interest in, or has chosen to keep) have been found:

- Inadequate command of English.
- Difficulties in navigating the system to find the full text of items.
- Lack of skills in the use of unique interface search options.
- Difficulty in translating the research topic and research question to appropriate search terms, and in finding synonymous, narrower or wider terms reflecting the various aspects of the research question.
- Lack of a systematic search strategy (manifested by, for example, repeated queries, not using metadata). Conducting searches and reformulating queries without planning ahead and without learning from results of previous queries.

With respect to the second research question, this study did not identify any clear patterns of query evolution during the search process. On the contrary, in addition to technical difficulties (e.g., language barriers, interface proficiency) a clear lack of systematic search strategy and no usage of feedback from previous queries, were identified. This behavior was found across the three databases.

Comparing informant's disposition towards each of the three databases, the following differences were found:

- Satisfaction with database's search results: satisfaction from the catalog was low compared to other databases.
- Time: time invested in the catalog was shorter compared with that invested in Google-Scholar and Proquest databases. It was difficult to determine whether the shorter time spent searching the catalog is one of the reasons for dissatisfaction with the results or vice versa.
- Search language: the students tended to search in Hebrew (their mother tongue) in databases with a Hebrew interface. Despite the appeal of an interface in their native language, and the ease of use and comfort it may provide, in many cases, their search results were less successful. Developers of information retrieval systems and interfaces should take into account this user's tendency when providing the interface in languages for which the databases does not have items in these specific languages.
- Queries and search strategies:
 - Search queries performed in the library catalog have properties that distinguish them from the other two databases: 1. Using metadata in the first queries. 2. Searching only one aspect of the research topic in the first query. 3. Three times lower query success rates, compared to Google-Scholar.
 - Vocabulary of search terms used in the search session:
 - The vocabulary was divided into three groups: core terms, used in all three databases; search terms that were used in two databases; search terms that were used in one database only.



- Use of unique terms (terms used only in one database), which in part led to relevant results to the user, suggesting the use of the search terms was non-systematic.
- Google-Scholar and Proquest shared more search terms than either one shared with the catalog.
- The search terms have mixed effectiveness retrieving relevant items in the same database (using different search tactics) and in different databases. In other words, using the interface in an educated, planned manner is pivotal.

This study highlights a number of difficulties in the search process. We introduce ideas and a possible course of action to improve the information search process. Users need to understand the structure of information and how information retrieval systems work, for example, by distinguishing between full-text search and a search using metadata. Users must be provided with specific knowledge of information retrieval techniques (e.g. using nesting, metadata) and skills of using varied search tactics (e.g. specification, generalization, using narrower/wider/synonymous search terms), including knowledge of various operators (especially the OR) and their mode of operation. Likewise, users need to learn systematic use of search terms, combining them within well planned search strategies of an interactive process designed to improve search results based on previous queries.

The end-users must be aware that different information retrieval systems have different operation modes. Consequently, the search strategy should be adjusted to the possibilities and capabilities of the system. It is important to be familiar with the specifics of the system being used and its unique indices and limiting options. It is also important to address the form of information storage, and type of information the search in

being performed on (full text, thin metadata, thick metadata). Educated use of the information retrieval interface will raise the query success rate.

It is necessary to teach searching skills on the way to accessing the information source itself, specifically the full text of the source (for example by learning the differences between free online sources and paid sources provided freely to students by the library). Understanding the routes of access to information sources will reduce the time spent locating them and enable reaching a greater number of relevant information items.

Hebrew (mother tongue) speakers should strive to enrich their vocabulary in the field of their information need (professional terms, narrower/wider, synonyms) to enhance effective queries. The end-user should be taught to use specific language tools (such as dictionaries and thesauri). The importance of using relevant terms from the search results must be emphasized.

We cannot and should not rely solely on improvements in the retrieval systems. Our findings point to a pressing need for information literacy education to enable the self-reliant user to perform effective information search, without the mediation of information specialists. Information literacy covers many skills, including the ability to search and locate the necessary information efficiently, the ability to evaluate the information and sources of information critically, and the ability to use information effectively to achieve a specific goal.

Teaching these skills will address the issues raised and discussed in this study. Information specialists, librarians, whose expertise relies on theoretical background, professional experience and unique knowledge, can provide end-users with necessary tools needed to perform effective searches and develop information literacy. This recommendation is supported by empirical findings indicating that information literacy is a



significant factor in academic success. Moreover, it is our belief that the efficiency of searches can and should be improved significantly. We find this to be true not only for students of Information Science, but also for the broader population of academic researchers and for the general population. Everyone needs effective search skills to meet an ever growing diverse need for information in everyday life, work, and leisure.

System number:

1206012