Abstract

In this empirical study, search tactics and strategies of the users of the German Education Index, a bibliographic database for the domain of educational research, were analyzed. With the help of logfile analyses, the use of 15 different term tactics, 4 search formulation tactics and 6 distinct search strategies could be identified. As a first step towards gaining a deeper understanding of complex subject search processes, the quantitative occurrences of these tactics and strategies are presented in this paper, augmented with illustrative examples for each of them.

1 Motivation

Although much research exists on the search behavior of users of web search engines, these results are not necessarily transferable to the search in specialized search engines. Consequently, the question arises if the findings for web search behavior also apply to specialized search engines, one of which is the German Education Index.

For example, a well-known study by [Spink et al. 2001] states that advanced search features are only scarcely used and that few query reformulations occur. The study at hand analyzed if this also holds true for the German Education Index.

Moreover, the study aimed to get an insight into the users’ subject search behavior in general and to understand commonly applied search tactics and strategies. In order to analyze user’s reformulation tactics and to examine if they show a tendency for certain search strategies, the study’s focus was on complex search sessions. Based on these results, search support functions may be deducted and implemented in the future to optimally assist the users during the query formulation process.

2 Related Work

2.1 Studies on search behavior

To assess the search behavior of users of a web search engine, [Spink et al., 2001] analyzed the logfiles of one day of the Excite search engine. They came to the conclusion that “most people use few search terms, few modified queries, view few Web pages, and rarely use advanced search features”. But they also state that this may be characteristic of web searching and that this does not necessarily apply to other retrieval systems. They found out that 48,4% of the users submitted a single query, 20,8% two queries and 31% three or more queries. If queries were reformulated, modifications appeared in small increments, preferably by changing, adding or deleting one term at a time. The average number of terms per query was 2,4 and less than 5% of the queries used Boolean operators. On average, users posted 4,86 queries per search session.

In a previous study on the Excite data set, [Jansen et al., 2000] reported an average of 2,8 queries per search session but also stated that most users use just one query per search. [Lau and Horvitz, 1999] further analyzed the semantics of query refinements in Excite searches and state that few users refined their searches by specialization, generalization or reformulation.

[Rieh and Xie, 2006] analyzed Excite searches not only quantitatively but traced query reformulation sequences with the help of query logs. They focused on 313 sessions with six or more unique queries. In this excerpt, they could distinguish eight reformulation patterns - specified, generalized, parallel, building-block, dynamic, multitasking, recurrent and format reformulation patterns, for which they present illustrative examples. Furthermore, they analyzed the use and frequency of the following content related reformulation tactics - specification (29,1%), generalization (15,8%), replacement with synonym (3,7%) and parallel movements (51,4%).

[Silverstein and Henzinger, 1999] analyzed query logs of the AltaVista Search Engine and also came to the conclusion that users mostly defined short queries of an average length of 2,35 and scarcely use query modifications.

Apart from studies focusing on search behavior in web search engines, several studies exist that refer to domain-specific retrieval systems. They thus have a similar application domain as the study at hand.

For example, [Sutcliffe et al., 2000] conducted an empirical study on the MEDLINE database and assessed the search behavior of information novices and experts. They found out that the average recall of all subjects was low (13,9%), compared to a gold standard. On the whole, novices used fewer query iterations than experts. For example, experts used cycles of narrowing and broadening, whereas novices focused on trial and error approaches. Moreover, experts used facilities like term suggestions, thesaurus and term exploration (truncation
and wild cards) more frequently than novices and also made more use of Boolean operators. On average, experts used 9,1 terms per query, while novices used 6,6 terms.  

[Wildemuth and Moore, 1995] also analyzed search behavior on the MEDLINE database, coming to the conclusion that the query formulations could be improved by a more frequent use of synonyms, the correct use of Boolean operators and the more frequent consultation of controlled vocabulary resources such as an online thesaurus.

2.2 Search tactics and strategies

Basic concepts for the description and analysis of search behavior were defined by [Bates, 1979] and [Harter, 1986]. [Bates, 1979] describes search processes in terms of search tactics and search strategies. While the search strategy designates the overall search plan, a tactic is defined as “a move made to further a search”, thus serving to realize the superordinate strategy.

According to [Bates, 1979], four types of search tactics can be distinguished, two of which are term tactics and search formulation tactics. She lists the following different search formulation tactics - specify, exhaust, reduce, parallel, pinpoint and block, which all describe actions to design or reformulate a query, for example by using Boolean operators.

Term tactics, by contrast, apply specifically to certain terms in the query, which may be added, replaced or deleted, for example by the following tactics - super, sub, relate, neighbor, trace, vary, fix, rearrange, contrary, respell, respace, as listed by [Bates, 1979].

A combination of search tactics may be used to pursue a certain search strategy. [Harter, 1986] draws a distinction between subject searches, that focus on the document contents, and non-subject searches, which query for non-semantic characteristics of documents. To achieve the latter goal, non-subject searches can make use of certain query fields such as document type, year of publication, language, author or source.

Compared to non-subject searches, subject searches show a big variety of complexity, ranging from simple searches following the quick approach to very sophisticated approaches such as in the pairwise facets strategy.

As defined by [Chu 2003], the quick approach is the simplest way of searching. The user enters one or more terms without using any operators.

The briefsearch strategy [Harter, 1986] adds one level of complexity to the quick approach as Boolean operators are used. It primarily serves to get an overview of the documents in the retrieval system. For this reason, it is often used as an entry point to a more complex search. If several retrieval systems are queried at the same time, [Harter, 1986] speaks of a multiple briefsearch strategy.

For example, a briefsearch can serve as an entry point to the building blocks approach. In this strategy, the search is split up into equivalent facets, each of which can be represented by several terms. While the terms in each facet are connected by the Boolean OR operator, the distinct facets may be combined by both OR or AND operators.

[Harter, 1986] further enumerates different kinds of successive facet strategies. Depending on the nature of the search facet that is used in the first query step, he distinguishes the most specific concept first strategy and the fewest postings first strategy.

If the search facets are considered to be equally important, the pairwise facets strategy [Harter, 1986] can be employed. In this case, two facets at a time are intersected, and finally the result sets of all facet combinations are merged.

While the above mentioned strategies start with a high recall, the citation pearl growing approach [Harter, 1986] starts with a precision-oriented search to identify a few relevant documents from which new search terms can be inferred.

In interactive scanning, the user starts with a high-recall search and scans the possibly long result lists, which is a very time-expensive approach.

If a domain-specific search system is used, facets that represent the domain focus should not be used. [Harter, 1986] calls this strategy implied facets.

Other strategies that take the result documents’ citations into account are subsumed as citation indexing strategies by [Harter, 1986]. They are based on the assumption that cited publications, authors or cocited authors lead to relevant documents.

3 The German Education Index

The German Education Index is a bibliographic database for the domain of educational research. In April 2009, it comprised 657720 documents, primarily in German language (more than 80%). The corpus can be searched by a retrieval system that is based on the Lucene search engine.

The quick search mode (figure 1) allows to dispatch free text queries which apply to the document fields index terms, title, author/editor, institutions, abstract and source.

![Quick Search in the German Education Index](image)

**Figure 1: Quick search of the German Education Index**

As illustrated in figure 2, the advanced search mode offers more sophisticated search functionalities than the quick search. The user can define searches in one or more of the different fields free text, index terms, title, year, author/editor, journal, institutions, source or update (1) and combine the search in different fields by the Boolean operators AND, OR and NOT (2). In every field, multiple terms are by default combined by AND, but a configuration by the user is possible (3). To define his/her query, the user can use an index term register (4) and a person register (5). Further restrictions are possible by document type (6) and by language (7).
Capitalization is disregarded by the system. If a user enters a term that can be looked up in a synonym list, the synonym(s) is (are) automatically added to the query, connected by the OR-operator. Furthermore, the system provides a “Did you mean” – functionality. It is based on the Levenshtein algorithm and generates search suggestions if a query has invoked an empty result set. Depending on the search field, the entered query is compared to a register over this field. When a user searches for a last name that several authors have in common, the respective full author names are listed for specification, for example.

4 Method

To analyze users’ search tactics and strategies, logfile analysis is a promising method [Jansen, 2009]. One of its advantages is the availability of large amounts of data that do not need to be collected especially for analysis purposes. Moreover, logfiles document real user queries and no artificial situation is created, such as in formal retrieval experiments.

Nevertheless, logfile analyses do not allow to get into direct contact with the users. Consequently, the users’ information needs are unknown and can only be inferred from the available data.

For this study, the logfiles of one day, April 1st 2008, were analyzed. They were ordered both chronologically and by IP address. This way, the logfiles of each user were accumulated in separate text files.

In the next step, the files were filtered by size with the aim of restricting the analyses to files that were likely to comprise complex searches. For this purpose, a minimum size of 10 kb was defined, which was fulfilled by 153 files. The discarded 512 smaller files were not expected to comprise complex search processes that were in the focus of this study.

Afterwards, the text files were split up into search sessions whose distinct query steps were timely and thematically related. If the time between two user actions was more than 30 minutes and/or if a query clearly addressed a new topic, this was considered as an indicator for the start of a new search session. This way, 235 search sessions were identified. These may comprise several distinct queries, each of which may consist of one or more query steps.

The query reformulation steps throughout the search sessions could be described by term tactics. Partly based on the classification by [Bates, 1979] presented in section 2.2, several tactics were identified, as listed in figures 3 and 4. Whenever a term in a query was deleted, replaced, altered or added, these actions were classified as one of these tactics.

The term tactics ranged from semantic relations such as broader terms, narrower terms, synonyms, related terms, antonyms and compounds over the use of affixes, singular or plural forms and truncations to changes in word order, the use of spacing characters between terms, phrase searches, translations (e.g. from English to German) and term conversions (e.g. from adjective to substantive). If none of these tactics was applied, the action was classified as an unrelated term.

Figure 2: Advanced Search of the German Education Index

Search formulation tactics that could be identified by logfile analyses in this study refer to the use of Boolean operators, as illustrated in figure 4. If elements were added by one of the operators or if one such element was deleted, this was considered as a distinct search formulation tactic.

Figure 3: Identifiable term tactics

Out of the strategies presented in section 2.2, the quick approach, briefsearch, building blocks and pairwise facets could be identified by logfile analyses, as shown in figure 5.

Figure 4: Identifiable search formulation tactics

<table>
<thead>
<tr>
<th>Search Formulation Tactics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition of an AND-element</td>
</tr>
<tr>
<td>Deletion of an AND-element</td>
</tr>
<tr>
<td>Addition of an OR-element</td>
</tr>
<tr>
<td>Deletion of an OR-element</td>
</tr>
</tbody>
</table>

Figure 5: Identifiable subject search strategies

Unfortunately, the logfiles did not give information about the number of result documents for each query. For this reason, the most specific concept first and the fewest postings first strategies could not be identified. Instead, two new strategies were defined, inspired by Harter’s strategy definitions. The most general query first strategy starts with a general query which is afterwards specified, for example by adding new facets with the Boolean AND-operator. The most specific query first strategy proceeds in the contrary way, starting with a specific query, consisting of several facets, and deleting or generalizing parts of the query step by step.

As the study focused on the analysis of a single search system, the multiple briefsearch strategy was not identifiable. Furthermore, the logfile analyses did not allow to deduce the consequences that users drew from
the viewing of documents, which impeded the detection of citation indexing strategies and interactive scanning. As the result documents were not inspected in this study, the citation pearl growing approach could not be identified either.

5 Results

5.1 Search statistics

On average, the users reformulated their query 7.8 times per search session, which accounted for an average session length of 8.8 steps. The mean query length was 2.2 terms per query.

The major part of the search sessions comprised queries with an average query length between one and two terms, as shown in figure 6. It categorizes search sessions by their average query lengths. For example, a search session with several one-term queries would have an average query length of 1, while a session with two two-term queries and two three-term queries would have an average query length of 2.25. In 8.1% of the search sessions only one query term was used on average, in 50.6% of the search sessions, the average query length was between 1 and 2 terms, in 29.4% of the sessions it was between 2 and 3 terms, and in 11.9% the average query length was higher than 3.

Results

The majority of the 235 search sessions was conducted in the advanced search mode (78.3%), which is illustrated in figure 7. In 17.9% of the search sessions, a combination of the quick and advanced search masks was used and in another 3.8% of the sessions only the quick search was employed. This may be due to the fact that the analyses mainly focused on complex searches.

If a users followed a link in the result set, he/she was presented the bibliographic data of a document, which is defined as a document view here. The average number of visited documents was only 2.4 per search session. In 11.91% of the search sessions, no result document at all was visited, in 43.4% of the sessions only one document was visited and in 34.89% of the searches, 2-5 documents were visited, which is illustrated in figure 8.

5.2 Overview of search strategies

The following figure 9 illustrates how the strategies can be split up into subject search strategies and non-subject search strategies. On the whole, 440 strategies were identified in the 235 search sessions. Subject search strategies constituted the biggest part (56.8%), while non-subject search strategies made up for the remaining 43.2%.

Figure 8: Percentage of sessions with x documents viewed

Rather than being directly consulted by the users, the registers were preferably used indirectly when the system provided register based search suggestions. These suggestions were accepted by the users in 113 cases whereas they autonomously consulted the available person and index term registers in only 14 cases. Unfortunately, the logfiles did not give evidence of the number of suggestions that were disregarded by the users. Users addressed the registers most frequently for looking up index terms (92.9%) while the consultation of the person register accounted for only 7.1% of the register uses. By contrast, the system generated suggestions accepted by the users were mainly based on the person register (62%).

Figure 9: Percentage of subject and non-subject searches

Figure 10 shows how the 250 subject search strategies could be classified. The biggest amount of strategies were briefsearches (47.6%), followed by the quick approach (23.2%), the pairwise facets strategy (11.6%), the most general query first strategy (10.4%) and the most specific query first strategy (5.6%). In 1.2% of the cases, a combination of the two latter strategies could be
identified. The building blocks approach was used in only one of the cases.

The following sections present examples for each of the identified strategies. Furthermore, they analyze which search tactics were applied in each of these search strategies.

5.3 Briefsearch

The following figure 11 shows how often each of the term tactics listed in section 2.2 was applied in the total amount of 119 briefsearches. Obviously, one of the most frequently used tactics was the replacement of a term with a completely unrelated term (24x). The second group of tactics was made up by thesaurus relations that were applied for search reformulation by the users: related terms were used in 18 of the cases, broader terms 11x, narrower terms 10x and synonyms 13x. Syntactic variants such as plural or singular forms (9x) and truncations (9x), compounds (7x), translations (6x), conversions (5x) and phrases (4x) made up the third group of tactics in briefsearches.

The following figure 12 traces an example of a briefsearch. The example queries given for the distinct search strategies were originally defined in German language. To increase the readability of this paper, they were translated into English.

5.4 Quick Approach

In the 58 quick approach searches, fewer different term tactics could be identified than in the briefsearches. This is due to the fact that the quick approach searches comprised only one word queries in this study because query terms were by default combined by the AND operator.

Related terms made up the most frequently used term tactic in the quick approach (16x), followed by the use of broader terms (7x), narrower terms (4x), synonyms (4x) and syntactic variants such as translations (3x), affixes (1x) and singular and plural forms (1x). The remaining term tactics could not be detected in the quick approach searches.

5.5 Pairwise Facets

In the 29 searches adhering to the pairwise facets strategy, a big variety of different term tactics could be identified, such as in the briefsearches. But in contrast to the term tactics in briefsearch, the use of related terms accounted for the highest number of tactics in the pairwise facets strategy.
Figure 16: Example of a pairwise facets strategy

In this example, the user started with a query consisting of the three facets “intercultural”, “competence” and “pedagogics”, which were consecutively combined in pairs in steps 2 to 5. This example also illustrates the tactic of using related terms (“pedagogics” and “learning”), which was identified as the most frequently applied tactic for this strategy.

5.6 Most General Query First

Term tactics in the 26 most general query first searches were dominated by the use of narrower terms (5x) and broader terms (4x). Unrelated terms, spacing characters, compounds and singular and plural forms were all applied in two of the cases.

Figure 17: Most general query first term tactics

The following example in figure 18 illustrates the use of the most general query first strategy.

Figure 18: Example of a most general query first strategy

In this example, the user started with a general term that was consecutively specialized with new facets that were combined with the AND operator. Starting with the general query “education”, the new facet “fundamentals” was added in step 2, followed by another facet in step 3.

5.7 Most Specific Query First

In the 14 most specific concept first searches, only few term tactics could be identified - the use of a broader term in one case, the use of affixes in one case and the use of a conversion in one case.

The following example in figure 19 illustrates the use of the most specific query first strategy.

Figure 19: Example of a most specific query first strategy

In this example, the user started with a very specific query and then deleted the different facets consecutively.

5.8 Combination of Most General Query First and Most Specific Query First

In three cases, users applied a combination of the most general concept first and the most specific concept first strategies. They either started with a very specific query which was then generalized and afterwards specialized again with different terms, or the other way round. An example of such a strategy combination is given in figure 20.

Figure 20: Example of a combination of the most general query first and most specific query first strategies

In this excerpt of a search, the user started with a general query which was then consecutively specialized in steps 2 and 3 before generalizing it again in step 4.

5.9 Building Blocks

Although the building blocks approach could be identified only once, an example is given in figure 21.

Figure 21: Example of a building blocks strategy

In this example, the user defined a query with two facets which were combined by the AND-operator. The second facet was expressed by a combination of two terms combined by the OR-operator.

5.10 Overview of Term Tactics

The following figure 22 gives an overview of the percentages of the different term tactics throughout all subject search strategies. It illustrates that related terms were most frequently used (19,1%), followed by the use of broader terms (17,5%), completely unrelated terms (12,2%), narrower terms (8,9%) and synonyms (9,4%). Singular and plural forms made up 6,5% of the term tactics, followed by compounds (5,3%), truncations, conversions (each 4,5%) and translations (4,1%). The tactics affix, word order, antonym, spacing and phrase search each had a share of less than 4%.
be determined so that strategies such as lowest postings first and most specific concept first may be identifiable.

Throughout the analyzed complex searches, the use of the advanced search mask was high. This means that users were aware of and familiar with the advanced search functionality. This is a main difference from the use of web search engines, as discussed in section 2.1. Nevertheless, the use of Boolean operators in the advanced search mode was rare as users tended to maintain the default configurations.

But they used a wide variety of term tactics throughout their searches, of which the semantic relations such as broader terms, narrower terms, synonyms and related terms made up the biggest part. While further investigations still need to verify if these tactics also prove to be effective, the use of such tactics can be supported by the system, such as by the suggestion of semantically related query expansion terms. In this context, the possibility of using an ontology for query expansion support in the German Education Index is currently examined [Carstens, 2009].

On the whole, users made frequent use of the suggestions the system already provided, which speaks for their general acceptance of search suggestions. The autonomous use of registers, by contrast, was very rare. Consequently, suggestions or possible expansion terms should be proactively offered to the user.

Compared to the users of web search engines, the users of the German Education Index made frequent reformulations in the analyzed searches. But it has to be stated that these focused on mainly complex searches. This also accounts for the comparatively high average session length of 8,8 query steps. The average query length of 2,2, by contrast, was even slightly lower than in the studies mentioned in section 2.1.

Throughout the search sessions, most users viewed less than 5 documents. These results indicate a tendency for selective searches in the analyzed sessions, while recall-oriented overview searches scarcely occurred. But it still needs to be examined if this also applies to the remaining less complex searches that were not inspected in this study. If this is the case, future enhancements of the retrieval system should focus on the support of precision-oriented processes.

While this study primarily examined subject searches, non-subject searches are planned to be analyzed in a similar way in the near future, shedding light on the way users search for documents based on already known facts.

6 Conclusion and Outlook

This study focused on the identification of tactics and strategies in complex subject searches in the German Education Index. With the help of logfile analyses, the quick approach, briefsearch, building blocks, most specific query first, most general query first and pairwise facet approaches could be identified. Briefsearches and the quick approach were proven to make up for the biggest part of queries, even in the sample of mainly complex searches to which this study was restricted.

Nevertheless, more sophisticated strategies such as building blocks, pairwise facets, most specific query first, most general query first and a combination of the latter two could also be detected. Future research will have to investigate if these sophisticated strategies are also successful and effective in terms of recall and precision. If this is the case, the system may be designed to offer more support in the application of such strategies.

For deeper analysis in the future, the logged searches may be reconstructed. This way, the result lists can be analyzed in depth and the number of result documents can be determined so that strategies such as lowest postings first and most specific concept first may be identifiable.

Throughout the analyzed complex searches, the use of the advanced search mask was high. This means that users were aware of and familiar with the advanced search functionality. This is a main difference from the use of web search engines, as discussed in section 2.1. Nevertheless, the use of Boolean operators in the advanced search mode was rare as users tended to maintain the default configurations.

But they used a wide variety of term tactics throughout their searches, of which the semantic relations such as broader terms, narrower terms, synonyms and related terms made up the biggest part. While further investigations still need to verify if these tactics also prove to be effective, the use of such tactics can be supported by the system, such as by the suggestion of semantically related query expansion terms. In this context, the possibility of using an ontology for query expansion support in the German Education Index is currently examined [Carstens, 2009].

On the whole, users made frequent use of the suggestions the system already provided, which speaks for their general acceptance of search suggestions. The autonomous use of registers, by contrast, was very rare. Consequently, suggestions or possible expansion terms should be proactively offered to the user.

Compared to the users of web search engines, the users of the German Education Index made frequent reformulations in the analyzed searches. But it has to be stated that these focused on mainly complex searches. This also accounts for the comparatively high average session length of 8,8 query steps. The average query length of 2,2, by contrast, was even slightly lower than in the studies mentioned in section 2.1.

Throughout the search sessions, most users viewed less than 5 documents. These results indicate a tendency for selective searches in the analyzed sessions, while recall-oriented overview searches scarcely occurred. But it still needs to be examined if this also applies to the remaining less complex searches that were not inspected in this study. If this is the case, future enhancements of the retrieval system should focus on the support of precision-oriented processes.

While this study primarily examined subject searches, non-subject searches are planned to be analyzed in a similar way in the near future, shedding light on the way users search for documents based on already known facts.

6 Conclusion and Outlook

This study focused on the identification of tactics and strategies in complex subject searches in the German Education Index. With the help of logfile analyses, the quick approach, briefsearch, building blocks, most specific query first, most general query first and pairwise facet approaches could be identified. Briefsearches and the quick approach were proven to make up for the biggest part of queries, even in the sample of mainly complex searches to which this study was restricted.

Nevertheless, more sophisticated strategies such as building blocks, pairwise facets, most specific query first, most general query first and a combination of the latter two could also be detected. Future research will have to investigate if these sophisticated strategies are also successful and effective in terms of recall and precision. If this is the case, the system may be designed to offer more support in the application of such strategies.

For deeper analysis in the future, the logged searches may be reconstructed. This way, the result lists can be analyzed in depth and the number of result documents can be determined so that strategies such as lowest postings first and most specific concept first may be identifiable.

Throughout the analyzed complex searches, the use of the advanced search mask was high. This means that users were aware of and familiar with the advanced search functionality. This is a main difference from the use of web search engines, as discussed in section 2.1. Nevertheless, the use of Boolean operators in the advanced search mode was rare as users tended to maintain the default configurations.

But they used a wide variety of term tactics throughout their searches, of which the semantic relations such as broader terms, narrower terms, synonyms and related terms made up the biggest part. While further investigations still need to verify if these tactics also prove to be effective, the use of such tactics can be supported by the system, such as by the suggestion of semantically related query expansion terms. In this context, the possibility of using an ontology for query expansion support in the German Education Index is currently examined [Carstens, 2009].

On the whole, users made frequent use of the suggestions the system already provided, which speaks for their general acceptance of search suggestions. The autonomous use of registers, by contrast, was very rare. Consequently, suggestions or possible expansion terms should be proactively offered to the user.

Compared to the users of web search engines, the users of the German Education Index made frequent reformulations in the analyzed searches. But it has to be stated that these focused on mainly complex searches. This also accounts for the comparatively high average session length of 8,8 query steps. The average query length of 2,2, by contrast, was even slightly lower than in the studies mentioned in section 2.1.

Throughout the search sessions, most users viewed less than 5 documents. These results indicate a tendency for selective searches in the analyzed sessions, while recall-oriented overview searches scarcely occurred. But it still needs to be examined if this also applies to the remaining less complex searches that were not inspected in this study. If this is the case, future enhancements of the retrieval system should focus on the support of precision-oriented processes.

While this study primarily examined subject searches, non-subject searches are planned to be analyzed in a similar way in the near future, shedding light on the way users search for documents based on already known facts.

6 Conclusion and Outlook

This study focused on the identification of tactics and strategies in complex subject searches in the German Education Index. With the help of logfile analyses, the quick approach, briefsearch, building blocks, most specific query first, most general query first and pairwise facet approaches could be identified. Briefsearches and the quick approach were proven to make up for the biggest part of queries, even in the sample of mainly complex searches to which this study was restricted.

Nevertheless, more sophisticated strategies such as building blocks, pairwise facets, most specific query first, most general query first and a combination of the latter two could also be detected. Future research will have to investigate if these sophisticated strategies are also successful and effective in terms of recall and precision. If this is the case, the system may be designed to offer more support in the application of such strategies.

For deeper analysis in the future, the logged searches may be reconstructed. This way, the result lists can be analyzed in depth and the number of result documents can be determined so that strategies such as lowest postings first and most specific concept first may be identifiable.

Throughout the analyzed complex searches, the use of the advanced search mask was high. This means that users were aware of and familiar with the advanced search functionality. This is a main difference from the use of web search engines, as discussed in section 2.1. Nevertheless, the use of Boolean operators in the advanced search mode was rare as users tended to maintain the default configurations.

But they used a wide variety of term tactics throughout their searches, of which the semantic relations such as broader terms, narrower terms, synonyms and related terms made up the biggest part. While further investigations still need to verify if these tactics also prove to be effective, the use of such tactics can be supported by the system, such as by the suggestion of semantically related query expansion terms. In this context, the possibility of using an ontology for query expansion support in the German Education Index is currently examined [Carstens, 2009].

On the whole, users made frequent use of the suggestions the system already provided, which speaks for their general acceptance of search suggestions. The autonomous use of registers, by contrast, was very rare. Consequently, suggestions or possible expansion terms should be proactively offered to the user.

Compared to the users of web search engines, the users of the German Education Index made frequent reformulations in the analyzed searches. But it has to be stated that these focused on mainly complex searches. This also accounts for the comparatively high average session length of 8,8 query steps. The average query length of 2,2, by contrast, was even slightly lower than in the studies mentioned in section 2.1.

Throughout the search sessions, most users viewed less than 5 documents. These results indicate a tendency for selective searches in the analyzed sessions, while recall-oriented overview searches scarcely occurred. But it still needs to be examined if this also applies to the remaining less complex searches that were not inspected in this study. If this is the case, future enhancements of the retrieval system should focus on the support of precision-oriented processes.

While this study primarily examined subject searches, non-subject searches are planned to be analyzed in a similar way in the near future, shedding light on the way users search for documents based on already known facts.

References


