

Search Trails as Collaboration Artifacts – Evaluating the UX

Sebastian Franken¹, Ulrich Norbistrath², Wolfgang Prinz¹

Fraunhofer FIT, Sankt Augustin¹

University of Applied Sciences Upper Austria, Hagenberg²

Summary

Users tend to share search results such as links or excerpts from web pages when performing complex search tasks. We developed a system called SearchTrails that enables sharing of complete search logs in the form of directed graphs, resembling the users' trails through the Internet. We conducted a study with 29 participants in which we used both written reports and search trails as collaboration artifacts. We researched the value of search trails as artifacts for collaborative search compared to written reports. We did this by evaluating the user experience (UX) with both artifacts. By evaluating a user experience questionnaire (UEQ), we can confirm that search trails can support collaborative search better than written reports containing the essential information of the search trails.

1 Introduction

The sharing of search results is mostly done by sending notes or links via e-mail or verbal communication (Ringel Morris 2013). Even when some users are supposed to work on the same topic, dedicated collaborative search supporting tools are only used rarely (Kelly & Payne 2014). Such tools often work on specified datasets in which users already rated information for subsequent users of such systems. SearchTrails is an approach for supporting asynchronous collaborative search by capturing the users' trail through the Internet, which allows exchanging, recreating, and continuing search trails between users. We conducted a study with 29 participants. In this paper, we address the research question if search trails as collaboration artifacts have an advantage over written reports containing the essential information of the search trails during collaborative search with regards to user experience (UX). For our study, we divided our participants in two groups (14 and 15 participants). One group received an average report from a previous search session with our tool. The other group received the corresponding search trail from the same session. We expected the

participants to base their search on the given material to create a new report on a related complex search task. After this, we let them evaluate the UX of SearchTrails with the user experience questionnaire (UEQ). The results indicate that search trails have a major impact on the quality of the search process. Using search trails as collaboration artifacts improves the UX during asynchronous complex search and the quality of the created reports.

The work presented here is based on the SearchTrails tool for the support of complex search tasks. An early version of SearchTrails was evaluated in a qualitative study with seven participants, conducted in early 2014. This version allowed the creation of a search trail, the capturing of highlights in pages and the extraction of keywords from pages. During the first study, we found that it is possible with SearchTrails to give support for complex search tasks by creating search trails (Franken & Norbistrath 2014a, published at ‘M&C 2014’). We could also show that the generated search trails help evaluators with assessing and evaluating the outcome of complex search tasks (Franken & Norbistrath 2014b). Further implementation led to a second version of SearchTrails, which enabled exchanging search trails between users and therefore the collaboration between searchers. While the first study proved the feasibility of the selected approach, the here presented second study goes beyond the first study in the technical capability of SearchTrails and a significantly larger number of participants.

In the following sections, we will present the underlying theoretical concepts and approaches for logging search behavior as well as for collaborative search. We will then describe the architecture and implementation of the recent version of SearchTrails and present our study. The last sections will elaborate on the study results and give an outlook on future work.

2 Theoretical background and related work

Today’s search engines get better and better in helping users to find fact-based information (Jansen 2006), but lack support for gathering many pieces of information on one specific topic (e.g. organizing a trip, including comparing means of transport, hotels, and leisure activities) (Singer et al. 2012). These well-supported fact searches are often small and repeated parts of a more complex exploration process, which currently lacks support. Our approach does exactly this: it visually logs the users’ path when traversing the web, offers possibilities to capture specific information, and provides a way back to information already seen in earlier stages of the search process. The idea of creating, storing, and exchanging trails through hypertext dates back to (Bush 1945). He suggests that the navigation through the hypertext gets stored in so-called ‘trails’ that could later be recalled and extended. Bush even proposes exchanging trails between persons being interested in the same topic: “*And his trails do not fade [...] [He] photographs the whole trail out, and passes it to his friend for insertion in his own memex, there to be linked into the more general trail*“. To this day, this is not possible, and SearchTrails follows this idea. Bush’s idea influenced (Bates 1989), who suggests that the search process is comparable to ‘Berry picking’, in which highly qualitative information bits get picked during a search. She opposes this to classic information retrieval, in which a query is tried to be answered by a perfectly matching document.

Besides other approaches of classifying search tasks (e.g. Broder 2002), Marchionini classified search activities into the three categories of ‘Lookup’, ‘Learn’, and ‘Investigate’ (Marchionini 2006). He defined exploratory search tasks as consisting of the overlapping ‘Learn’ and ‘Investigate’ activities and specified sub tasks for each search activity. (Singer 2012) based his definition of ‘Complex search’ on that. He defines complex search tasks as interactive, labor-intensive, and time consuming tasks that need at least one of the three tasks aggregation, discovery, or synthesis of information. We stick to this definition, as it covers a wider range of search tasks and includes tasks that do not necessarily involve learning. (Franken & Norbistrath 2014a) holds a comparison of search characteristics.

There have been several approaches that try to capture the searchers' course through the Internet. One of the first was a logging system by (Fox et al. 2005), who developed an add-in for the Internet Explorer, that collected both implicit feedback about the searchers actions and explicit feedback by presenting small questionnaires to the searcher. A later approach was the ‘Wrapper’ system by (Jansen et al. 2006), which was installed on the user's system, and recorded the user's interaction with the browser and the browser's interaction with the system. This approach could show that users search for long times and make use of different information systems for their searches. (Singer et al. 2011) developed the SearchLogger system that consisted of a Firefox plug-in and recorded all user actions on a central server. SearchLogger allowed proving the existence of complex search tasks and their classification. Common to all presented approaches is that the users had no chance to interact with the data gathered during the search. Therefore, no support during complex search tasks was given.

The value and the rising impact of collaborative search is shown by (Ringel Morris 2013), remarking that only lightweight approaches have the chance to be accepted by users. Studies of collaborative search support tools have confirmed those findings, as they state that the actual success of such systems have been limited, which may be due to high effort of installation and during usage (Kelly & Payne 2014).

Recent publications prove emerging interest in the concept of generating trails during search and offering trail information to searchers, as it was found that *‘Individual items may be insufficient for vague or complex information needs’* (Singla et al. 2010). This approach has taken logs from web browser plug-ins which were interpreted as trails and evaluated with respect to trail length, breadth, or diversity. Later approaches talk about similarities in trails and provide suggestions to users based on the trails other users took (Awadallah et al. 2014). Several methods for measuring the UX of interactive products exist. We decided for the user experience questionnaire (UEQ), as it is a standard approach for such measurements, judging the impact of the product on the user in general (Rauschenberger et al. 2013).

3 SearchTrails

We created SearchTrails to support searchers by (1) visualizing their search trails, (2) supporting search task evaluation, and (3) the exchanging of search trails to ease collaborative search. Designing the tool, we were guided by the idea to support aggregation,

discovery, and synthesis. We support aggregation by recording the trail with all its paths and side tracks, and synthesis by enabling the user to select valuable information pieces from websites and collecting them into the highlights overview. We support discovery by a visual representation of the search process as well as search term suggestions to guide the searcher into related, but new search directions. The search trails are stored on a remote server, conserving the context of search results and allowing the exchange of search trails. User selections of valuable information (highlights) are labeled in the search trail and displayed in an overview, keeping relevant information and its sources together.

We realized our tool by building an extension for the Google Chrome browser, which unobtrusively logs the users' actions while working with the browser. Realizing SearchTrails this way poses only very minor installation effort and allows free-form, open-ended search processes. Our design also helps to avoid cluttering the visualization with commercial ads (Singer 2012), as SearchTrails stores only URLs. We have chosen a force directed graph as this is a proven concept for visualizing large information collections (Eades & Huang 2000).

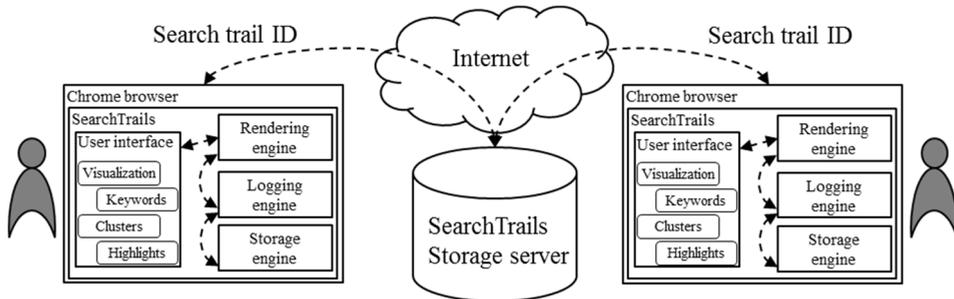


Figure 1: SearchTrails architecture supporting the exchange of search trails.

The internal structure of the now evaluated version of SearchTrails consists of three parts and the user interface (Fig. 1): The logging engine logs all events and builds a logical representation of the search process as a JSON Object. The storage engine stores the JSON trail object remotely and reports about the changes, which are translated into the visualization by the rendering engine. In addition to the version evaluated qualitatively in (Franken & Norbistrath 2014a), this version allows exchanging search trails IDs, which trigger the recreation of trails in the collaborators' instance of SearchTrails.

SearchTrails monitors the opening, closing, and switching between tabs and the change of URLs with the help of a logging engine which logs, filters, and interprets the user interactions. It also catches metadata from the web pages and generates keywords from the pages' content. The users' navigation is transformed into nodes in the graph visualization (c.f. Fig. 2) by the rendering engine and shown in a separate tab, which itself is excluded from being logged. The metadata gets attached to the nodes and is visible to the user. The visualization is based on a directed graph in a forced-directed layout (2 in Fig. 2). For each visited URL, important keywords are derived (1), stored and displayed in a table on the right, if a keyword appears on two or more pages the user has visited. This metadata gets also stored in the search logs. If more than three nodes belong to the same host, they are clustered

by a colored hull (4) and can be closed using a mouse click to reduce the complexity of the visualization, but still being sensible to user interactions. Irrelevant nodes can be deleted. The storage engine stores the search trail JSON object regularly on a remote server which allows the authors to access all data, while restricting access for the participants.

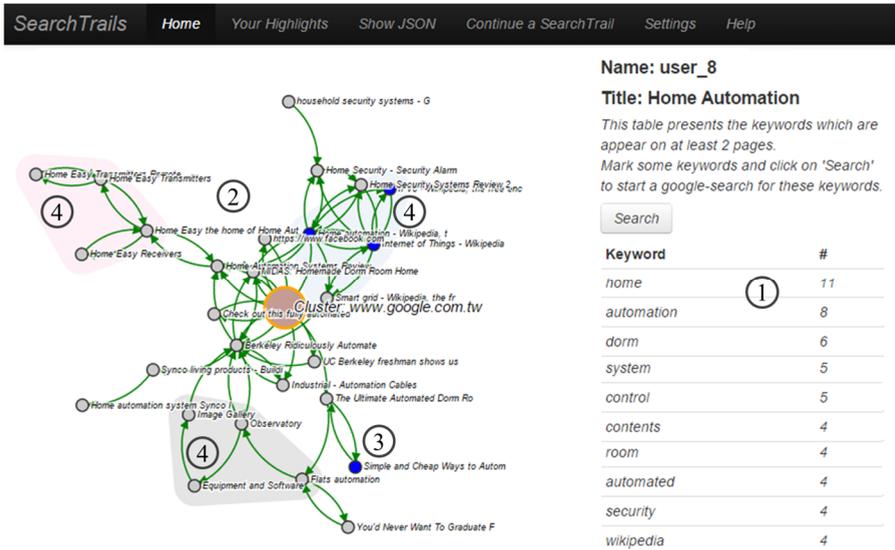


Figure 2: SearchTrails with a trail from the user study.

Besides the visualization, users can interact with the following features: To ease the synthesis of information (as one of the key points of complex search), users can select text and store it as a highlight by pressing ALT. This stores the selected text in an overview table, together with its source URL, and the corresponding node in the visualization being marked blue (3), which allows the users to immediately recognize the highlight-pages in the graph. Highlights can also be added as text from other sources, to allow manual enrichment of the information, and erroneously set highlights can be removed. A tap on ‘-’ marks a page in the search trail red, to signal that it was not helpful. The JSON object corresponding to the search can be viewed, imported, and exported, although this was mainly used for evaluation purposes. The views of the highlights and JSON can be switched on or off. The keywords derived from the pages’ contents are shown as a table and are selectable; this results in the nodes with the matching keywords being marked in the visualization. A click on ‘Search’ starts a search for the selected keywords in a new tab.

Technically, the visualization is done by making use of the ‘Data-Driven Documents’ Java Script framework d3js (<http://d3js.org/>), which allows highly dynamic and efficient visualization of large data structures. The browser plug-in inserts JavaScript code into all the pages the user visits, to be able to extract keywords from the pages. Updates as well as the keywords are shared within the application by message passing. On every loading of a site, a new node is created in the visualization. When this is finished, the node is updated by the keyword information and some page metadata, such as the description of a page. When a

searcher moves from one URL to another by using the back button or by switching browser tabs the nodes corresponding to this URL are connected.

A search session starts with opening SearchTrails and entering a username and a title for the search. After that, a unique SearchTrails ID is stored locally under which the trail is stored on the server. While searching, the search trail is stored automatically every five minutes, as well as on closing SearchTrails. For continuing a search, SearchTrails uses the stored trail ID to trigger the trail being fetched from the logging server. This way, also a trail ID received via e-mail can be used to fetch a trail from another user. This trail can then be used as a starting point for the users' own search and be extended by highlights and annotations.

4 User study

For our second user study, we selected 29 students of a university lecture on Computer Supported Collaborative Work (CSCW) in a master course as a representative example of tech-savvy users, having experience with web search. We equipped all participants with the Google Chrome Browser and the SearchTrails extension and made sure the installation went well. We developed two search tasks, to avoid biasing the study results by the selection of just one topic. The tasks required the evaluation of given artifacts and checked to fulfill the seven characteristics of complex search tasks in (Kules & Capra 2009). The search tasks were: Topic 3D '3D printing': 'Based on the given material, find applications of 3D-Printing in the car manufacturing domain. Which applications exist, which ones will come? Will 3D-Printing change the way of manufacturing cars in the future?' and Topic HA 'Home automation': 'Based on the given material, find applications of home automation dealing with home security. Which applications exist, which ones will be available? Which applications would you prefer?'

The participants were assigned two different topics per group to avoid biasing the results by the selected topic. The topics were handed out with alternating groups through the rows of the lecture hall. This prevented plagiarism and neighboring students being assigned to the same group. This resulted in four sub groups. Based on the groups the students were assigned to, the topic and the given artifact differed: Group A: 3D+Report; B: HA+Report; C: 3D+Trail; D: HA+Trail. We provided the participants either a written report or the corresponding search trail for one of the topics. For the same topic (3D or HA), these artifacts came from one average student who performed a similar search process (leaving out more specific information about cars or home security) and therefore contained basic information in each artifact. By providing a search artifact we simulated asynchronous collaboration between two searchers on the same topic and ensured a comparable level of given information for all participants. The participants were asked to use the given information for creating a new report on the extended question. This report should contain all relevant information from the search, needed to inform someone who did not do the search about the task. Including URLs was not mandatory, and we did not request a minimum amount of text, to avoid the production of filler text. The study was started as a subsidiary part of the lecture and the participants were given six days to do a search on the given task

finishing with the preparation of a report. At the sixth day, tardy participants were reminded via e-mail to hand in the reports. In the evaluation, the participants generated search trails which were stored on the server. From our participants, we received 26 search trails and UEQs and 21 reports. After the week of the study, all participants met again for the final questionnaire session. In this session, we handed out a self-developed questionnaire and the UEQ to evaluate specific opinions on SearchTrails and the general UX. We consider the UEQ an indicator for the quality of support during the search process. Only when the participants are content with the support during the search process, the UEQ will result in positive scores.

After the study, we evaluated the UEQ results and split them depending on the type of the provided artifact. All three authors also independently graded the participants' reports to achieve some level of objectivity. As the reports were requested to be able to inform someone who did not perform it about its key results, we graded the reports by their quantitative breadth and their qualitative depth of information. Based on this, the reports of group C+D were better with regards to breadth and depth. Additionally, the statistical data of the generated search trails was evaluated.

5 Evaluation results

This user study was intended to show the impact of using search trails as collaboration artifacts. First, it had to be ensured that the selected topics did not influence the study results. For this we split and compared our study results with respect to the topics. Our analyses of the UEQ, the grades of the reports, and the statistical results with respect to the two different topics show that no significant differences between the different topics can be found. This proves that our search tasks did not influence the study results.

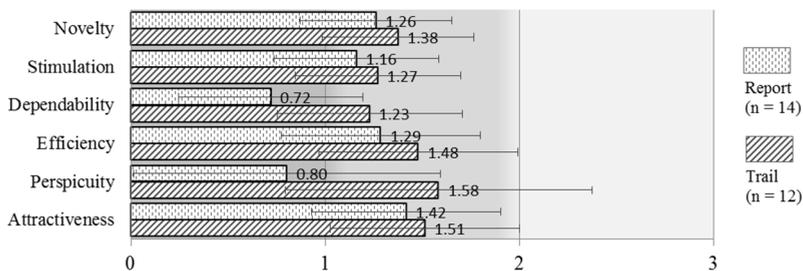


Figure 3: Search Trails UEQ scores split by the report and the trail as collaboration artifacts.

We then split our UEQ results with respect to the given artifact. The UEQ groups its 26 items pairs into the six categories of novelty, stimulation, dependability, efficiency, perspicuity, and attractiveness. Fig. 3 shows the UEQ scores for the groups that received the report and the search trail and their respective confidence intervals ($p=0.05$). Note that the scale of the UEQ reaches from -3 to +3, but for spacing reasons, we just show the positive half of the diagrams, as the negative parts are empty. The fact that all values are above 0.8 (except for the dependability, which is 0.72) show that the overall rating for each scale can

be considered explicitly positive. Cronbach's Alpha coefficient is a measure for the consistency of a scale, which is generally accepted if it is above 0.7. It was calculated for all six scales, and in case of the report group, we gain acceptable values for all scales but dependability and novelty. For the trail group, we gain acceptable values for all scales but efficiency, implying that the results for most of the scales can be considered meaningful.

When the users received a trail to start their search from, all UEQ item values are higher (Fig. 3) than for the group that received a report, and some values increase significantly. These values are even more impressive when taking into account that the participants worked with SearchTrails for the first time, and did not know its functions before. The value for perspicuity almost doubles and increases from 0.80 to 1.58. Similarly, the value for dependability increases by more than 0.5; from 0.72 to 1.23. The general increase of all scales for the trail group and the especially good values for perspicuity and dependability can be explained by the use of the search trail as an initial artifact. The trail group discovered the trail, which obviously represents the collaborators actions. This results in a high perspicuity. Similarly, the trail feels more steerable and dependable for this group, as it offers possibilities for interaction. The report group may experience the trail as a less transparent visualization of their navigation, which they interacted with much less than the trail group.

Our analyses of the reports' grades show that also their quality improved when the group received a trail to start from. Their average grade was 2.21 and as such statistically significant ($p=0.05$) roughly one full grade better than for the report group, which was graded 3.20 on average. The influence of the trail on the resulting report shows that the value added by the search trail is more than in UX, but also impacts the efficiency the participants worked with the information. This can also be seen on the value for dependability, which scored lower for the report group, as these participants conceived the force directed trail less predictable and steerable. The high values for the trail group can be explained by the fact that the participants especially liked that the trail was laid out automatically and a flexible layout offered them the chance to explore the given data much more efficiently.

Fig. 4 shows the UEQ results from all participants in a benchmark comparison with 163 studies with 4818 participants (data is provided in the UEQ evaluation sheet). Taking all user data together, Cronbach's alpha is acceptable for all scales but novelty, where the value is 0.59. This value is influenced by the weak value for the report group, which already scored low in that scale. SearchTrails scores 'Above average' or 'Good' on all scales except dependability ('Below average'). A reason for that is in the low dependability scores of the report group, influencing the overall value.

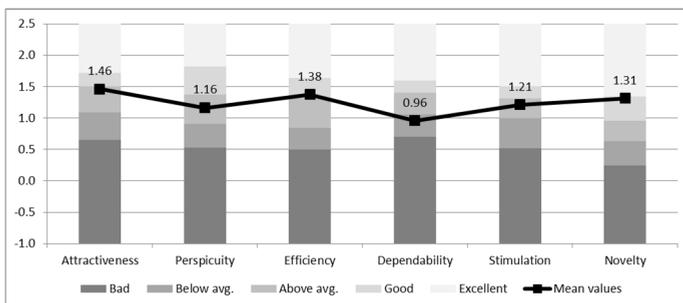


Figure 4: Search Trails UEQ benchmark comparison for all participants of the study.

6 Conclusion and outlook

The results presented here show that SearchTrails as a system is a valuable tool to support complex search tasks. Further analyses of the UEQ results with respect to the type of artifact show that the value of SearchTrails itself is more in collaboration than in individual searching: Users engaged in collaborative search much more appreciate the value of a search trail than the value of a report. Based on these results, we can answer the research question if search trails have an advantage compared to written reports as search artifacts during collaborative search with regards to user experience (UX) positively. Although the study shows that a good matching use case has a strong impact on the UX, the impact of the search trail was more than UX, as the grades for the reports were better for the trail group.

Preliminary experiments could show that the grades of the reports and the individual satisfaction with SearchTrails' features improve through practice with SearchTrails. This indicates that participants have to get acquainted with a system before they can fully appreciate its features. Our results may even have improved when we would have trained the report group more on how to efficiently use SearchTrails for their search tasks. We decided not to do too much training apart from explaining core features, to avoid any influence on the results. Apart from a certain training time that might influence the use of features, the influence of the artifact remains substantial. This is due to the type of artifact. Going back to the deconstruction of complex search into aggregation, discovery, and synthesis, a report itself is an artifact of synthesis. It consists of processed information, leaving the reader unaware of the landscape of information that lead to it. When it comes to collaborative search, less processed artifacts of collaboration are important. This is the value of a search trail, as it captures all paths, side tracks, and sources that lead to a result. Users can prevent redundant searches by knowing about the side tracks of their collaborators.

Although the study went well in technical regards, a potential flaw is that even if students are a tech savvy user group, they usually do only necessary tasks. Their behavior is less intrinsically motivated than the behavior of e.g. researchers, who can more easily spot the value of tools for complex search. Summing up, we argue in favor of the value of search trails as primary artifacts of aggregation and discovery, which are able to prevent redundant searches. We also consider SearchTrails itself a tool for enhancing synthesis of information, leading to better overview of valuable search results. The work on SearchTrails will continue to further investigate the impact on collaborative complex search.

Acknowledgments

We thank the reviewers for their work and comments that helped improving the clarity of this paper. We furthermore thank our study participants from the CSCW course of B-IT for their time invested in our study.

References

Awadallah, A., White, R. W., Pantel, P., Dumais, S., Wang, Y. (2014). Supporting Complex Search Tasks. In: *Proc. of CIKM '14*, Shanghai. ACM Press, pp. 829-838.

- Bates, M.J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Information Reviews*, Vol. 13, Issue 5, pp. 407-424.
- Broder, A. (2002). A taxonomy of web search. In: *SIGIR Forum*, Fall 2002, Vol. 36, No. 2, pp. 3-10.
- Bush, V. (1945). As we may think. In: *The Atlantic Monthly* 176, July 1945.
- Eades, P., Huang, M.L. (2000). Navigating Clustered Graphs Using Force-Directed Methods. In: *Journal of Graph Algorithms and Applications*, Vol. 4, Np. 3, pp. 157-181.
- Fox, S., Karnawat, K., Mydland, M., Dumais, S., White, T. (2005). Evaluating Implicit Measures to Improve Web Search. In: *ACM Transactions on Inf. Syst.* Vol.23, No.2, Apr. 2005, pp. 147-168.
- Franken, S., Norbistrath, U. (2014a). Trail Building During Complex Search Tasks. In: M. Koch, A. Butz & J. Schlichter (Hrsg.): *Mensch und Computer 2014*, München: Oldenbourg, pp. 135-144.
- Franken, S., Norbistrath, U. (2014b). Supporting the Evaluation of Complex Search Tasks with the SearchTrails Tool. In: *CASCON 2014*, Toronto. IBM Corp. Riverton, NJ, USA, pp. 262-274.
- Jansen, B.J. (2006). Using Temporal Patterns of Interactions to Design Effective Automated Searching Assistance. In: *Communications of the ACM*, April 2006/Vol. 49, No. 4, pp. 72-74.
- Jansen, B.J., Ramadoss, R., Zhang, M., Zang, N., (2006). Wrapper: An Application for Evaluating Exploratory Search Systems Outside of the Lab. In: *Proc. of SIGIR 2006*, Aug. 6-11, Seattle.
- Kelly, R., Payne, S. J. (2014). Collaborative Web Search in Context: A Study of Tool Use in Everyday Tasks. In: *Proceedings of CSCW'14*, Baltimore. ACM Press. pp. 807-819.
- Kules, B., Capra, R., (2009). Designing Exploratory Search Tasks for User Studies of Information Seeking Support Systems. *JCDL '09*, June 15-19, Austin, Texas, p. 419.
- Marchionini, G. (2006). Exploratory Search: From Finding to Understanding. In: *Communications of the ACM*, April 2006/Vol. 49, No. 4, pp. 41-46.
- Rauschenberger, M., Schrepp, M., Perez Cota, M., Olschner, S., Thomaschewski, J. (2013). Efficient Measurement of the User Experience of Interactive Products. In: *IJIMAI*, Vol. 2, N° 1, pp. 39-45.
- Ringel-Morris, M. (2013). Collaborative Search Revisited. In: *Proc. of CSCW '13*, San Antonio, Texas. ACM Press, pp. 1181-1192.
- Singer, G. (2012). Web search engines and complex information needs. *Dissertationes Mathematicae Universitatis Tartuensis*, Order No. 446, ISSN: 1024-4212.
- Singer, G., Danilov, D., Norbistrath, U. (2012). Complex search: aggregation, discovery, and synthesis. In: *Proceedings of the Estonian Academy of Sciences*, 2012, **61**, 2, pp. 89-106.
- Singer, G., Norbistrath, U., Vainikko, E., Kikkas, H., Lewandowski, D. (2011). Search-Logger - Analyzing Exploratory Search Tasks. In: *SAC'11*, March 21-25, Taichung, Taiwan, pp. 751-756.
- Singla, A., White, R. W., Huang, J. (2010). Studying Trailfinding Algorithms for Enhanced Web Search. In: *Proc. of SIGIR '10*, Geneva, ACM Press, pp. 443-450.

Contact information

Sebastian Franken & Prof. Wolfgang Prinz, PhD, Fraunhofer FIT, Sankt Augustin. {sebastian.franken, wolfgang.prinz}@fit.fraunhofer.de. Dr. Ulrich Norbistrath, University of Applied Sciences Upper Austria, Hagenberg. ulrich.norbistrath@fh-hagenberg.at.