Digital Library Research in Action –
Supporting Information Retrieval in Sowiport

Daniel Hienert, Frank Sawitzki, Philipp Mayr
GESIS – Leibniz Institute for the Social Sciences
Unter Sachsenhausen 6-8, 50667 Cologne, Germany
{Daniel.Hienert, Philipp.Mayr}@gesis.org

1   Abstract

Sowiport is the information portal for the social sciences that contains over 8 million literature references, research projects and full texts from 18 databases, including six English-language databases from ProQuest/CSA. Since its foundation seven years ago, Sowiport has been re-launched on the basis of the popular discovery framework VuFind. In this article we present how heterogeneous databases from different data providers can be integrated to provide the user one point of access to social science information. Further we describe several value-added services that assist the user at different stages of the information seeking process.

2   Introduction

Beside multidisciplinary commercial and non-commercial bibliographic search engines such as Google Scholar (Jacsó, 2009), MS Academic Search (Jacsó, 2011) or BASE (Pieper & Summann, 2006), there exist a number of search engines for domain specific bibliographic information such as DBLP (Ley, 2009) for the computer sciences, Pubmed (Lu, 2011) or gopubmed (Doms & Schroeder, 2005) for biomedical literature or the arXiv repository[1] for physics, mathematics and computer science along with a large number of smaller initiatives with digital libraries (DLs), OPACs and search engines for domain-specific content. Focusing on one domain can improve the search experience by getting more relevant results due to a limited scope, and most notably, domain knowledge such as taxonomies, thesauri, structures, relationships and information behavior can be utilized to improve search quality and search usability (compare e.g. Battelle, 2005, p.274).

Research information in the social sciences like in any other domain is widely distributed across different databases, systems, search tools, digital libraries and information portals. The challenge is to collect these information sources and to make them easily accessible for the end user. Heterogeneous metadata and data make it difficult to integrate different databases and often hinder information search with qualitative results. Providing different information types (like literature references, research projects, and full texts) in one search makes it even more difficult to integrate and interlink this information. Sowiport[2] uses open-source technology, established metadata schemes and terminology mappings between different thesauri to overcome some of the typical retrieval problems.

Supporting services, as described in the next section, can assist the user at each stage of the information seeking process (see for an overview Kriewel et al., 2004; Mutschke et al., 2011). In the query formulation phase the user is supported by a search term recommender that combines terms from a domain-specific thesaurus and highly associated terms from a co-word analysis. This way, the user is supported in the (re-)formulation of queries by controlled vocabulary and alternative concepts. At the next stage, in the result list, users can apply re-rankings for citation counts, journal or author

---

[1] http://arxiv.org

[2] http://sowiport.gesis.org/

productivity that provides different views for the result set. Next, in the document full view, we provide several possibilities to continue the search process based on exploratory search via metadata facets, references, citations and links. In the following section we will give a brief overview of tools we use to build a knowledge basis for further developments.

3 Supporting Information Search in Sowiport

3.1 Overview

The social science information portal Sowiport integrates quality research information from national and international data providers and makes them available in one place. Sowiport integrates literature references, research projects and full texts. It currently contains about 8 million literature references and research projects from 18 databases, including six English-language databases from ProQuest/CSA which are available by a national license funded by the German Research Foundation. Intelligent technologies such as the automatic mapping of search terms between different thesauri support the user in simultaneous searches across heterogeneous databases. Furthermore, Sowiport offers value-added services like advanced term recommendations, different novel ranking techniques and exploratory search facilities based on metadata attributes, references, citations and links to support the retrieval process. Sowiport began in 2007 and was re-launched in April 2014 on the basis of the VuFind framework[3]. The main target groups of Sowiport are users interested in the social sciences and adjacent fields. The portal serves about 20,000 unique users per week, mainly from German-speaking countries.
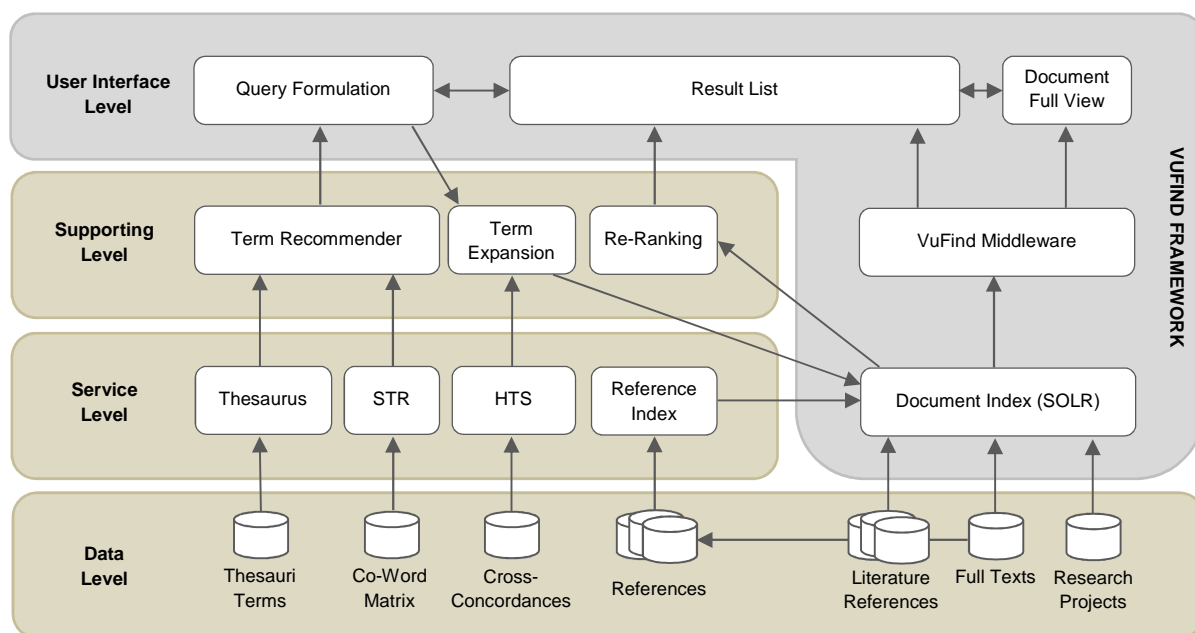
3.2 Sowiport Architecture

Figure 1 gives an overview of the Sowiport system architecture. The VuFind discovery software provides the basis for Sowiport. It offers basic search and browsing functionality including features like faceted search on the basis of Apache Solr[4], simple and extended search, search history and favorites, persistent URLs, APIs like Open Search and OAI and more features which make it an ideal basis for a modern discovery infrastructure.

The basic VuFind system was adapted on the frontend with a modern corporate design and several modifications and extensions. The backend (and respectively frontend) was extended with several services that support the user in the search process: (1) the Term Recommender supports the user in choosing appropriate terms for his information need, (2) the HTS service expands the query with synonym terms to find results from several heterogeneous databases, (3) the possibility to re-rank the result list allows different views and (4) links based on metadata attributes, references and citations allow browsing between documents and other result sets. In the following sections we explain these supporting services in more detail.

---

[3] http://vufind.org
[4] http://lucene.apache.org/solr/

Figure 1: Overview of the Sowiport architecture.

Table 1 gives a detailed overview of the databases that are currently included. Sowiport does not use a federated search approach or a harvesting via OAI-PMH, but most data providers deliver their data by uploading it to an FTP server (because e.g. they do not offer an OAI-PMH interface or the data is not publically available). Therefore, one first step for integration and integrated access is the conversion from diverse input formats into one overall metadata schema which is applied in Sowiport. The schema has to be compatible with different information types and many different metadata to avoid the loss of essential information. Even if standards such as MARC, Dublin Core, METS, MODS exist, in practice, data providers often use their own proprietary formats or variations of standards. For Sowiport we tried to automatize the process of conversion and indexation. Therefore, we used a script-based workflow, which first copies raw input files from the FTP to the conversion server. Then, for each input format we defined templates which describe the conversion process from input format to output format. In a last step, these files are indexed by Apache Solr into the document index. The overall schema for all input databases is a first step to let the user easily search in heterogeneous databases.

Table 1. Databases included in Sowiport

Information Type	Name	Description	Update Frequency	Quantity
German-language literature references	GESIS-SOLIS	Social Science Literature Information System	Daily (Diff)/ Monthly (Full)	460,077
	GESIS-BIB	GESIS Library	Quarterly	127,822
	USB Köln	University Library Cologne	Weekly	273,707
	DIPF - FIS-Bildung	FIS Education Literature	Quaterly	831,294
	FES-Katalog	Library of the Friedrich Ebert Foundation	Quaterly	636,012
	IAB - LitDokAB	Literature for Employment Research	Monthly	134,213
	DZA - Gerolit	German Centre of Gerontology Library	Monthly	162,135
	DZI - SoLit	Literature for Social Work and Social Education	Quaterly	202,257
	WZB – Library Catalogue	Berlin Social Science Center Library	Monthly	*106,640
English-language literature references	ProQuest - CSA / SA	Sociological Abstracts	Monthly	1,039,304
	ProQuest - CSA / SSA	Social Services Abstracts		172,250
	ProQuest - CSA / ASSIA	Applied Social Sciences Index and Abstracts		582,013
	ProQuest - CSA / PAIS	Public Affairs Information Service		1,943,324
	ProQuest - CSA / PEI	Physical Education Index		403,708
	ProQuest - CSA / WPSA	Worldwide Political Science Abstracts		790,972
	ProQuest - PAO	Periodicals Archive Online	Onetime	193,334
Full texts	GESIS - SSOAR	Social Science Open Access Repository	Daily	27,574
Research Projects	GESIS - SOFIS	Social Science Research Information System	Daily (Diff)/ Monthly (Full)	53,156
*Available on sowiport.gesis.org in December 2014				8,139,792

3.3 Integrating Heterogeneous Databases with Cross-Concordances

Another core challenge in integrating heterogeneous databases from different providers in one digital library is to keep precision and recall high if users perform simple keywords searches. Sowiport contains databases which use different thesauri to index their literature references. Thus, without any system side support, there is a small chance that users will get precise results from all databases. A first step was the syntactical integration in one overall metadata schema; the next step is the semantic integration with cross-concordances. In (Mayr & Petras, 2008) we defined cross-concordances as intellectually (manually) created crosswalks that determine equivalence, hierarchy, and association relations between terms from two controlled vocabularies. For example, the term 'Computer' in system A is mapped to the term 'Information System' in system B.
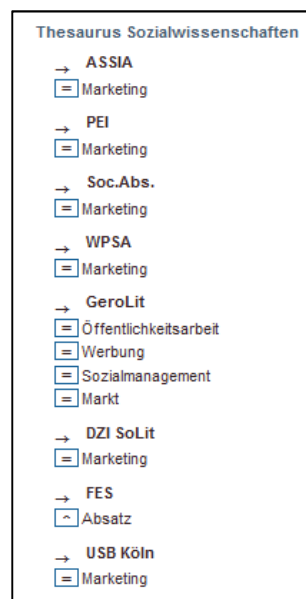
Semantic integration in our definition seeks to connect different information systems through their indexing languages (thesauri, classifications, etc.) – insuring that search over several heterogeneous collections in Sowiport can still use the advanced subject access tools provided with the individual databases. Cross-concordances can support search in several ways. First and foremost, they enable seamless search in databases with different indexing languages. Additionally, cross-concordances serve as tools for vocabulary expansion in general since they present a vocabulary network of equivalent, broader, narrower and related term relationships. As we have shown in Sowiport, this

vocabulary network of semantic mappings can also be used for automatic query expansion and reformulation.

To search and retrieve terminology data from the database, a web service (called heterogeneity service, HTS) was built to support cross-concordance searches for individual start terms, mapped terms, start and destination vocabularies as well as different types of relations. The cross-concordances database contains controlled terms from 25 different thesauri with about 513,000 entries. We use the HTS service in Sowiport to expand the user's search query with equivalence relations from all cross-concordances where the Thesaurus for the Social Sciences (TheSoz)[5] is source or target database. Figure 2 shows the user interface to the HTS integrated in the TheSoz browser. The controlled term "Marketing" in the example is expanded with terms such as (translated) "public relations", "advertising", "social management" or "market".

(Mayr & Petras, 2008) have evaluated the effect of using these cross-concordances for intra- and interdisciplinary search questions in a controlled information retrieval scenario. The expansion with "exact match"-crosswalks shows a very positive effect in terms of retrieval precision and recall, especially for topics which are searched in two databases and are situated in different research domains like psychology and medicine.

Figure 2: Query expansion in Sowiport: the search term *Marketing*
is expanded with several equivalents from other thesauri.

3.4 Term Recommender

A qualitative indexing of documents with controlled terms and the integration of heterogeneous databases with cross-concordances improves the search if the user is aware of the correct vocabulary and uses it for formulating his search query – the so-called vocabulary problem (Furnas et al., 1987). To assist the user in the search query formulation and for finding alternative concepts we have implemented a term recommender in Sowiport. Similar to solutions from commercial platforms like Google, eBay or Amazon, it proposes possible keywords to the user while entering letters and words into the search form. In addition to simple autocomplete functionality it utilizes vocabularies from (1)
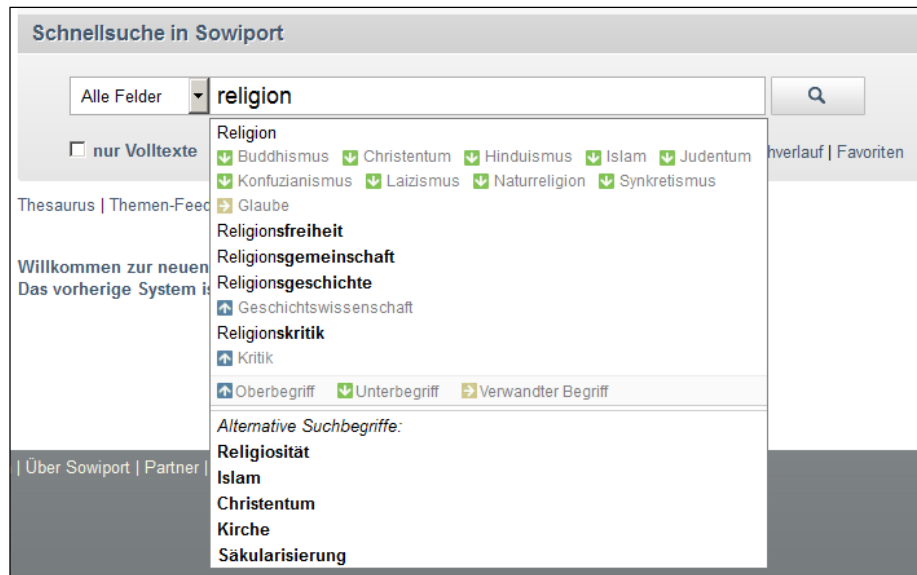
---

the Thesaurus for the Social Sciences and (2) the Search Term Recommender (STR) to provide descriptors for high-precision results.

The Thesaurus for the Social Sciences is an instrument to index and retrieve subject-specific information in Sowiport. The list of keywords contains about 11,600 entries, of which more than 7,750 are descriptors and about 3,850 are non-descriptors (see Zapilko et al., 2013). Topics in all of the social science disciplines are covered. Thesaurus terms are linked with each other with semantic relations such as "broader", "narrower" or "related". The Search Term Recommender (Lüke et al., 2012) maps arbitrary input terms to terms of a controlled vocabulary. All documents of a collection are processed by performing a co-occurrence analysis from free terms in titles and abstracts to subject-specific descriptor terms. The logarithmically modified Jaccard similarity measure is used to rank term suggestions from the controlled vocabulary. Two language-specific services have been created for Sowiport. First, for the German-language, all documents from the databases SOLIS (literature references) and SOFIS (research projects) have been processed and mapped to TheSoz terms. Second, for the English language, all documents from ProQuest/CSA have been processed and mapped to collection-specific indexing terms. To enable DL operators to create their own individual search term recommender based on OAI-harvested metadata we have built a framework (Lüke et al., 2013).

Figure 3 shows the current implementation of the term recommender. In the upper part it shows autocomplete descriptors from the TheSoz that fit to the current user input "religion" like e.g. (translated) "religious freedom", "religious association", "religious history" and "religious criticism". Additionally, underneath each descriptor it shows related terms in a lighter font color, e.g. "criticism" as a broader term for "religious criticism". This can help the user to find broader, narrower and related keywords, to identify the context of descriptors and to get suggestions for further search. In the lower section, beginning from three entered letters, it shows suggestions from the STR. Here, topically near suggestions for the input term "religion" are shown, e.g. (translated) "religiosity", "Islam", "Christianity", "church" and "secularization".

In a first evaluation (Hienert et al., 2011) we conducted a user study with over 4,000 unique visitors and four different design approaches. We used three variants with one single vocabulary: (1) user search terms, (2) terms from the heterogeneity terminology service, (3) thesaurus terms and (4) a combined recommender with thesaurus terms and terms from the STR. The different recommenders were successively activated in the live system and how often a search term suggestion was selected in relation to the number of conducted search queries was measured. The combined approach performed best with 14% usage, followed by thesaurus terms (9%), user search terms (7%) and terms from the heterogeneity service (3%). In an ongoing evaluation with Sowiport users we try to identify how users utilize the term recommender. We especially want to identify patterns of using thesaurus structures within a search session.

Figure 3: Search term recommendations for the entered term *religion*. Thesaurus terms in the upper part [translated: religious freedom, religious association, religious history, religious criticism] and statistically near terms in the lower section [translated: religiosity, Islam, Christianity, church, secularization]. Narrower terms for religion are [translated]: Buddhism, Christianity, Hinduism, Judaism and more.
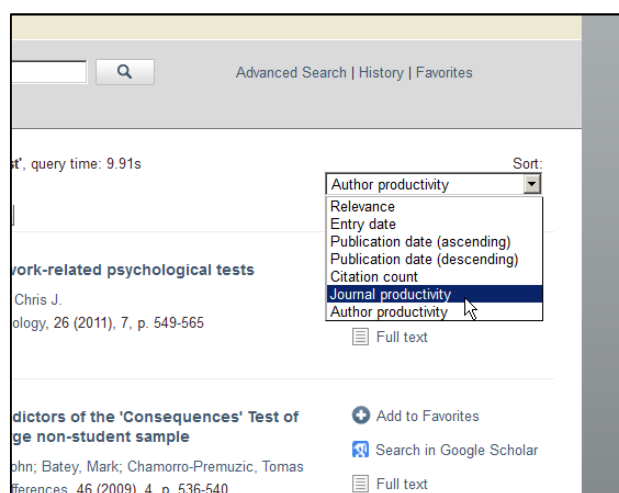
3.5   Re-Ranking Techniques

After submitting a search query, a result page is presented that lets the user explore document metadata. The VuFind framework provides standard functionality like modifying the search query, facets for filtering, saving and exporting individual documents or results sets. The standard ordering of the document list is a tf-idf measure provided by Solr. Another supporting facility that we integrated is the re-ranking of the documents in the result set to journal/author productivity or citation count. The initial ranking can be re-arranged by metrics derived from the models described below.

Traditional information retrieval has reached a high level in terms of measures like precision and recall, but scientists and scholars still face challenges present since the early days of DL: mismatches between search terms and indexing terms, overload from result sets that are too large and complex, and the drawbacks of text-based relevance rankings. Therefore, we focus on statistical modelling of the evolving science system. Such analyses have revealed not only the fundamental laws of Bradford (1948) and Lotka (1926) (see also Garfield, 1980), but also network structures and dynamic mechanisms in scientific production. In our approach these fundamental laws serve as structuring facilities (see e.g. Bradfordizing below) which have a strong empirical evidence and utility for the search process, especially for scholarly searchers. The overall aim here is to help the user to grasp the size and structure of the information space, rather than force him to precisely define the search space (see in detail Mutschke et al., 2011).

Bradfordizing is a simple re-ranking modelwhich we have called "Journal productivity" in Sowiport. Fundamentally, Bradford's law states that literature on any scientific field or subject-specific topic scatters in a typical way. A core or nucleus with the highest concentration of papers – normally situated in a set of a few so-called core journals – is followed by zones with loose concentrations of paper frequencies. The last zone covers the so-called periphery journals which are located in the model far distant from the core subject and normally contribute just one or two topically relevant papers in a defined period. Bradford's law, as a general law in informetrics, can be successfully applied to most scientific disciplines, and especially in multidisciplinary scenarios (Mayr, 2013).

Bradfordizing, originally described by (White, 1981), is a simple utilization of Bradford's law of a scattering model which sorts/re-ranks a result set according to the rank a journal gets in a Bradford distribution. The journals in a search result are ranked by the frequency of their listing in the result set, i.e. the number of articles in a certain journal. If a search result is "bradfordized", articles of core journals are ranked ahead of the journals which contain only an average number (Zone 2) or just a few articles (Zone 3) on a topic. The calculation of the individual document scores of a "bradfordized" article list is explained in (Schaer, 2011). In (Mutschke et al., 2011) we could show empirically that different bibliometric-enhanced re-ranking models can be used to improve retrieval quality. Bradfordizing as a re-ranking IR service has been evaluated in (Mayr, 2013).

"Author productivity" based on Lotka's Law is technically implemented exactly like "Journal productivity". The author name with highest productivity is ranked highest. Re-ranking subjects to "citation count" is sorting documents corresponding to their citation count in Sowiport (see more in the Section 3.6). This is a standard sorting option in all systems with citation counts (like e.g. Web of Science). In Sowiport users can re-rank results directly in the result page by choosing one of the techniques from the *Sort results by* menu (see Figure 4).

Figure 4: Alternative re-ranking techniques *Citation count*,
Journal productivity and *Author productivity*

3.6   Metadata Elements, References, Citations and Links for Exploratory Search

Typically, the last step in a basic search is choosing a document from the result list and checking its details in the document full view. At this stage we provide several possibilities to continue the search process in an exploratory way. Documents within a corpus like Sowiport (with its several databases) are not isolated elements, but form a network of interconnected entities. Connections between documents, and other result sets, are given by attributes such as authors, publishers, keywords, journals, subjects, references, citations and many more.

Metadata elements like author names, keywords or journal titles are marked as links and can easily be used to trigger new searches in Sowiport. We also provide links to full texts on the Web or to services like Google Scholar, Google Books or local copies where available. A special feature here is the linkage of research publications in Sowiport with research data in portals that contain this kind of data. Publications that cover issues such as the design, methodology, implementation and results of an

empirical study, or literature which interprets or discusses this data, are linked to the original data set with its metadata so that researchers have direct access. In the past, these links were added manually to the metadata by domain experts. In an on-going project we identified workflows and algorithms to extract these links automatically from full texts (Boland et al., 2012). At this time, about 2,706 bibliographic entries in Sowiport are linked to research data and can be followed from the full view of a record.

References and citations are harder to extract, compute and process, but if they are available, they offer the user the possibility of identifying and browsing citation networks or of re-ranking the result list by citation counts. Figure 5 shows a full view of a document in Sowiport with the capability to search for authors, journals, classifications and topics, and also to browse to referenced documents or to documents which cite the document.

Smaller data providers for Sowiport only very rarely have access to reference information and provide it for their collections. Today, only ProQuest provides approximately 8 million references for its English-language databases CSA-WPSA, CSA-SA and CSA-SSA. References are directly embedded in the raw input files and contain reference information with metadata fields such as title, author and publication year. For our own German-language full text repository SSOAR, we built an experimental workflow which automatically extracts references from PDFs based on the reference extraction tool ParsCit. In the future we will be able to extract reference information from other full text collections with very different citation styles.

To match these references with existing documents in Sowiport we apply a multi-stage process (Sawitzki et al., 2013). First, all reference information is stored in a separate reference index based on SOLR. Then, for the matching step, we search for the normalized title and publication year in the document index for every reference. If this step fails, a fuzzy search with title, publication year and ISSN is carried out. If a document is found, its ID is added to the reference entry. With this matching process about 30% of the references can be assigned to an existing record in Sowiport, which means in total about 2.5 million reference- and citation-links. The precision of the algorithm was 95% in a sample of 400 randomly chosen documents. In a final step, the reference index can be used to add reference and citation information to the metadata of each record in the document index. Metadata entries are then used to build browsable links as shown in Figure 5. In addition, computed citation counts can also be used for re-ranking the result set (compare Section 'Re-Ranking Techniques').

Figure 5: Full view of a record in Sowiport which makes it possible for the user to browse to referenced documents or to records which cited it.
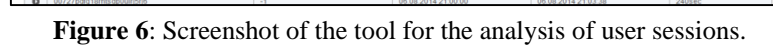
4 Analyzing User Behavior for further Developments

To build the basis for further value-added services which support a user in his information search we first have to understand why and how users search in our domain and especially in Sowiport. We try to obtain this knowledge through two measures: (1) usability studies and (2) an analysis tool which enables us to explore and examine whole retrieval sessions.

With the relaunch of Sowiport in 2014 we began a series of task-based usability tests. Initial tests made basic usability weaknesses visible which could easily be resolved with minor updates. In future tests with domain professionals we will try to understand the different search tasks that exist and how experts try to solve them (see also the initiative of an online access panel for IIR systems in Kern et al., 2014 - to appear).

To get further insight into why and how users are searching in Sowiport, we have very recently developed a tool for the analysis of whole user sessions. This framework can easily be integrated in different digital libraries with only some lines of software code. Furthermore, the examination of existing log files is possible if they are transformed into a simple schema. The purpose of this framework is to understand how users behave within and beyond sessions.

Figure 6 shows a screenshot of the user interface with user session data from August 2014. The DL operator can get an overview of how users behave in a set of search sessions. Therefore, the Sankey diagram shows which actions are performed at each search step in the session and which actions will most probably follow. In the session list, user sessions can be examined in detail, including for several sessions by the same user. The data set can be filtered to specific situations, e.g. to sessions in which users viewed a document for more than thirty seconds. The tool can help to answer questions such as "How has the search process evolved for a certain topic?", "Which documents have been finally

12

viewed?", "How has a search process evolved over several sessions?". The session based analysis can be the basis for a future set of value-added services that allow personalization, recommendation and awareness. For example, we can generate term suggestions based on the personal history of a user or can recommend documents viewed by other users that used the same search query.

Figure 6: Screenshot of the tool for the analysis of user sessions.

5   Conclusion

Sowiport currently supports the user's information seeking process at different stages. Social science information from different data providers is syntactically integrated with an overall metadata schema and semantically integrated with cross-concordances. The user is aided by a term recommender in the query formulation and reformulation phases; re-ranking mechanisms offer alternative views on the result set; links based on metadata, references and citations allow exploratory search in the document graph and related results sets. To build the basis for future developments we began with a series of usability studies and have created a tool that helps to analyze user behavior based on log data. This will help us to identify not only usability problems, but also discover at which stage in the information seeking process the user can be further supported.

References

Battelle, J. (2005) The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture. Portfolio.

Boland, K., Ritze, D., Eckert, K. & Mathiak, B. (2012) Identifying References to Datasets in Publications. In: TPDL. pp.150–161. URL: http://dx.doi.org/10.1007/978-3-642-33290-6_17.

Bradford, S.C. (1948) Documentation. London, Lockwood.

Doms, A. & Schroeder, M. (2005) GoPubMed: exploring PubMed with the gene ontology. Nucleic acids research, 33 (suppl 2), pp.W783–W786. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160231/.

Furnas, G.W., Landauer, T.K., Gomez, L.M. & Dumais, S.T. (1987) The Vocabulary Problem in Human-system Communication. Commun. ACM, 30 (11), pp.964–971. URL: http://doi.acm.org/10.1145/32206.32212.

Garfield, E. (1980) Bradford's Law and Related Statistical Patterns. Current Contents, 4 (19), pp.476–483. URL: http://www.garfield.library.upenn.edu/essays/v4p476y1979-80.pdf.

Hienert, D., Schaer, P., Schaible, J. & Mayr, P. (2011) A Novel Combined Term Suggestion Service for Domain-Specific Digital Libraries. In: Proceedings of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL). URL: http://dl.acm.org/citation.cfm?id=2042564.

Jacsó, P. (2009) Google Scholar revisited. Online Information Review, 32 (1), pp.102–114. URL: http://www.emeraldinsight.com/doi/full/10.1108/14684520810866010.

Jacsó, P. (2011) The pros and cons of Microsoft Academic Search from a bibliometric perspective. Online Information Review, 35 (6), pp.983–997. URL: http://www.emeraldinsight.com/doi/full/10.1108/14684521111210788.

Kern, D., Mutschke, P. & Mayr, P. (2014) Establishing an Online Access Panel for Interactive Information Retrieval Research. In: IEEE/ACM Joint Conference on Digital Libraries. London, UK, IEEE, pp.473–474. URL: http://arxiv.org/abs/1407.1540.

Kriewel, S., Klas, C.-P., Schaefer, A. & Fuhr, N. (2004) DAFFODIL - Strategic Support for User-Oriented Access to Heterogeneous Digital Libraries. D-Lib Magazine, 10 (6). URL: http://www.dlib.org/dlib/june04/kriewel/06kriewel.html.

Ley, M. (2009) DBLP: some lessons learned. Proceedings of the VLDB Endowment, 2 (2), pp.1493–1500. URL: http://dl.acm.org/citation.cfm?id=1687577.

Lotka, A. (1926) The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16 (12), pp.317–323.

Lüke, T., Schaer, P. & Mayr, P. (2013) A Framework for Specific Term Recommendation Systems. In: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA, ACM, pp.1093–1094. URL: http://arxiv.org/abs/1407.1539.

Lüke, T., Schaer, P. & Mayr, P. (2012) Improving Retrieval Results with Discipline-Specific Query Expansion. In: P. Zaphiris, G. Buchanan, E. Rasmussen, & F. Loizides eds. TPDL. Lecture Notes in Computer Science. Springer, pp.408–413. URL: http://arxiv.org/abs/1206.2126.

Lu, Z. (2011) PubMed and beyond: a survey of web tools for searching biomedical literature. Database, 2011, p.baq036. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3025693/.

Mayr, P. (2013) Relevance distributions across Bradford Zones: Can Bradfordizing improve search? In: J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed eds. 14th International Society of Scientometrics and Informetrics Conference. Vienna, Austria, pp.1493–1505. URL: http://arxiv.org/abs/1305.0357.

Mayr, P. & Petras, V. (2008) Cross-concordances - terminology mapping and its effectiveness for Information retrieval: Crosskonkordanzen - Terminologie Mapping und deren Effektivität für das Information Retrieval. In: World Library and Information Congress. Québec. URL: http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf.

Mutschke, P., Mayr, P., Schaer, P. & Sure, Y. (2011) Science models as value-added services for scholarly information systems. Scientometrics, 89 (1), pp.349–364. URL: http://dx.doi.org/10.1007/s11192-011-0430-x.

Pieper, D. & Summann, F. (2006) Bielefeld Academic Search Engine (BASE) An end-user oriented institutional repository search service. Library Hi Tech, 24 (4), pp.614–619. URL: http://hdl.handle.net/10760/9207.

Sawitzki, F., Zens, M. & Mayr, P. (2013) References and Citations to Support Information Retrieval in SOWIPORT. In: H. C. Hobohm ed. 13th International Symposium of Information Science (ISI 2013). Potsdam, Verlag Werner Hülsbusch, pp.267–272. URL: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-330073.

Schaer, P. (2011) Using Lotkaian Informetrics for Ranking in Digital Libraries. In: C. Hoare & A. O'Riordan eds. Proceedings of the ASIS&T European Workshop 2011 (AEW 2011). Cork, Ireland, ASIS&T. URL: http://arxiv.org/abs/1106.0217.

White, H.D. (1981) 'Bradfordizing' search output: how it would help online users. Online Review, 5 (1), pp.47–54. URL: http://dx.doi.org/10.1108/eb024050.

Zapilko, B., Schaible, J., Mayr, P. & Mathiak, B. (2013) TheSoz: A SKOS representation of the thesaurus for the social sciences. Semantic Web journal (SWJ), 4 (3), pp.257–263. URL: http://dx.doi.org/10.3233/SW-2012-0081.

Daniel Hienert

Daniel Hienert is a postdoctoral researcher at the GESIS department "Knowledge Technologies for the Social Sciences". He joined GESIS in 2007 after his graduation in computer science at the University of Koblenz and further studies of Italian and business studies at the Humboldt University Berlin. During his first year at GESIS he worked in the project vascoda-TB5, later in the projects IREON and Sowiport. In 2013 he finished his PhD on the integration of interactive visualizations in the search and linking process of heterogeneous information on the Web. From 10/2013 to 03/2014 he was an acting lead of the team GESIS Architecture. His research interests are Information Systems, Information Visualization and Information Retrieval.

Frank Sawitzki

Frank Sawitzki graduated 2007 in computer science at the University of Koblenz-Landau. His diploma thesis dealt with data visualization software under ergonomic aspects. After his studies he was employed as a software developer in the field of web development and network communication. Since August 2007 he has been with GESIS as a research assistant in the department "Knowledge Technologies for the Social Sciences". Since 2014 he works as a software developer for an insurance company.

Philipp Mayr

Philipp Mayr is a postdoctoral researcher and team lead at the GESIS department "Knowledge Technologies for the Social Sciences" (WTS). Since winter semester 2012, he teaches as a senior lecturer at Cologne University of Applied Sciences, Faculty 03 of Information Science. From October 2009 untill August 2011, he was a visiting professor for knowledge representation at the University of Applied Sciences in Darmstadt, Department of Information Science & Engineering. Philipp Mayr is a graduate of the Berlin School of Library and Information Science at the Humboldt University Berlin where he finished his doctoral research in 2009. He studied LIS, computer science and sociology. Since November 2004 he has been working in the internationally recognized projects "Competence Center Modeling and Treatment of Heterogeneity" (KoMoHe) and "Value-Added Services for Information Retrieval" (IRM) as researcher and principal investigator. Philipp Mayr published in the areas Informetrics, Information Retrieval and Digital Libraries and is member of the Networked Knowledge Organization Systems/Services (NKOS) network. He is member of the editorial board of the journals Scientometrics and Information Wissenschaft & Praxis. He serves frequently as a reviewer for various journals and international programme committees.