How do practitioners, PhD students and postdocs in the social sciences assess topic-specific recommendations?

Mayr, Philipp

Veröffentlichungsversion / Published Version
Konferenzbeitrag / conference paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Mitglied der

Leibniz-Gemeinschaft

How do practitioners, PhD students and postdocs in the social sciences assess topic-specific recommendations?

Philipp Mayr

GESIS - Leibniz Institute for the Social Sciences,
Unter Sachsenhausen 6-8
50667 Cologne, Germany
philipp.mayr@gesis.org

Abstract. In this paper we describe a case study where researchers in the social sciences (n=19) assess topical relevance for controlled search terms, journal names and author names which have been compiled by recommender services. We call these services Search Term Recommender (STR), Journal Name Recommender (JNR) and Author Name Recommender (ANR) in this paper. The researchers in our study (practitioners, PhD students and postdocs) were asked to assess the top n preprocessed recommendations from each recommender for specific research topics which have been named by them in an interview before the experiment. Our results show clearly that the presented search term, journal name and author name recommendations are highly relevant to the researchers topic and can easily be integrated for search in Digital Libraries. The average precision for top ranked recommendations is 0.749 for author names, 0.743 for search terms and 0.728 for journal names. The relevance distribution differs largely across topics and researcher types. Practitioners seem to favor author name recommendations while postdocs have rated author name recommendations the lowest. In the experiment the small postdoc group favors journal name recommendations.

Keywords: Recommendation services, bibliometric-enhanced IR, co-word analysis, author centrality, journal productivity, relevance assessment

1 Introduction

In metadata-driven Digital Libraries (DL) typically three major information retrieval (IR) related difficulties arise: (1) the vagueness between search and indexing terms, (2) the information overload by the amount of result records obtained by the information retrieval systems, and (3) the problem that pure term frequency based rankings, such as term frequency - inverse document frequency (tf-idf), provide results that often do not meet user needs [10]. Search term suggestion or other domain-specific recommendation modules can help users -

2   Philipp Mayr

e.g. in the social sciences [7] and humanities [4] - to formulate their queries by mapping the personal vocabularies of the users onto the often highly specialized vocabulary of a digital library. A recent overview of recommender systems in DL can be found in [2]. This strongly suggests the introduction of models in IR systems that rely more on the real research process and have therefore a greater potential for closing the gap between information needs of scholarly users and IR systems than conventional system-oriented approaches.

In this paper[1] we will present an approach to utilize specific information retrieval services [10] as enhanced search stratagems [1, 17, 5] and recommendation services within a scholarly IR environment. In a typical search scenario a user first formulates his query, which can then be enriched by a Search Term Recommender that adds controlled descriptors from the corresponding document language to the query. With this new query a search in a database can be triggered. The search returns a result set which can be re-ranked. Since search is an iterative procedure this workflow can be repeated many times till the expected result set is retrieved. These iterative search steps are typically stored in a search sessions log.

The idea of this paper is to assess if topic-specific recommendation services which provide search related thesaurus term, journal name and author name suggestions are accepted by researchers. In section 2 we will shortly introduce three different recommendation services: (1) co-word analysis and the derived concept of Search Term Recommendation (STR), (2) coreness of journals and the derived Journal Name Recommender (JNR) and (3) centrality of authors and the derived Author Name Recommender (ANR). The basic concepts and an evaluation of the top-ranked recommendations are presented in the following sections. We can show that the proposed recommendation services can easily be implemented within scholarly DLs. In the conclusion we assume that a users search should improve by using the proposed recommendation services when interacting with a scientific information system, but the relevance of these recommendations in a real interactive search task is still an open question (compare the experiences with the Okapi system [13]).

2   Models for information retrieval enhancement

The standard model of IR in current DLs is the tf-idf model which proposes a text-based relevance ranking. As tf-idf is text-based, it assigns a weight to term t in document d which is influenced by different occurrences of t and d. Variations of the basis term weighing process have been proposed, like normalization of document length or by scaling the tf values but the basic assumption stays the same. We hypothesize that recommendation services which are situated in the search process can improve the search experience of users in a DL.

The recommendation services are outlined very shortly in the following section. More details on these services can be found in [10].

---

[1] This paper is a slightly revised and updated version of a talk given at the EuroHCIR 2014 workshop in London.

2.1   Search Term Recommendation

Search Term Recommenders (STRs) are an approach to compensate the long known language problem in IR [3, 12, 7]: When searching an information system, a user has to come up with the "appropriate" query terms so that they best match the document language to get qualitative results. STRs in this paper are based on statistical co-word analysis and build associations between free terms (i.e. from title or abstract) and controlled terms (i.e. from a thesaurus) which are used during a professional indexation of the documents (see "Alternative Keywords" in Figure 1). The co-word analysis implies a semantic association between the free and the controlled terms. The more often terms co-occur in the text the more likely it is that they share a semantic relation. In our setup we use STR for search term recommendation where the original topical query of the researcher is expanded with semantically "near" terms[2] from the controlled vocabulary Thesaurus for the Social Sciences (TheSoz).
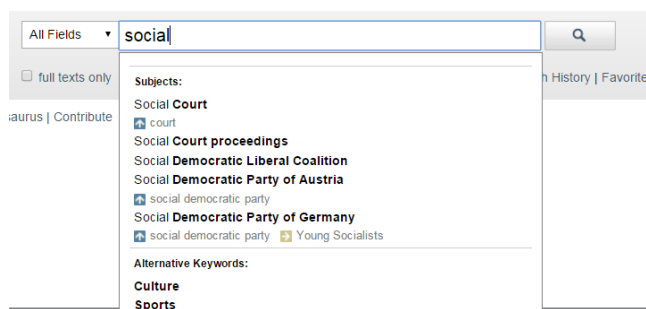

Fig. 1. Example search term "social" and search term recommendations

2.2   Recommending Journal Names

Journals play an important role in the scientific communication process. They appear periodically, they are topically focused, they have established standards of quality control and often they are involved in the academic gratification system. Metrics like the famous impact factor are aggregated on the journal level. In some disciplines journals are the main place for a scientific community to communicate and discuss new research results [11, 15]. In addition, journals or better journal names play an important role in the search process (see e.g. the famous search stratagem "journal run") [1, 8, 5]. The underlying mechanism for recommending journal names (JNR) in this paper is called Bradfordizing [16]. Bradfordizing is an alternative mechanism to re-rank journal articles according to core journals

---

[2] In the assessment we evaluated the "Alternative Keywords" generated from our STR implementation.

4   Philipp Mayr

to bypass the problem of very large and unstructured result sets. The approach
of Bradfordizing is to use characteristic concentration effects (Bradfords law
of scattering) that appear typically in journal literature. Bradfordizing defines
different zones of documents which are based on the frequency counts in a given
document set. Documents in core journals - journals which publish frequently
on a topic - are ranked higher than documents which were published in journals
from the following Bradford zones[3]. In IR a positive effect on the search result
can be assumed in favor of documents from core journals [8, 9]. Bradfordizing
is implemented as one re-ranking feature called "Journal productivity" in the
digital library sowiport[4] [6]. In our setup of the assessment we evaluated the
journal name recommendations, namely the top-ranked 5 core journals after
Bradfordizing.

Fig. 2. Recommending author names in our retrieval prototype. Example search term
"luhmann" and highly associated author names "central authors" in the right panel

2.3   Recommending Author Names

Collaboration in science is mainly represented by co-authorships between two or
more authors who write a publication together. Transferred to a whole commu-
nity, co-authorships form a co-authorship network reflecting the overall collabo-
ration structure of a community. The underlying mechanism for recommending
author names (ANR) in this paper is the author centrality measure betweenness.
Author centrality is another way of re-ranking result sets (see Figure 2). Here the
concept of centrality in a network of authors is an additional approach for the
problem of large and unstructured result sets. The intention behind this rank-
ing model is to make use of knowledge about the interaction and cooperation
behavior in special fields of research. The (social) status and strategic position

---

[3] An explanation and visualization of Bradford Zones is given in [8].
[4] http://sowiport.gesis.org/

of a person in a scientific community is used too. The model is based on a network analytical view on a field of research and differs greatly from conventional text-oriented ranking methods like tf-idf. A concrete criterion of relevance in this model is the centrality of authors from retrieved publications in a co-authorship network. The model calculates a co-authorship network based on the result set to a specific query. Centrality of each single author in this network is calculated by applying the betweenness measure and the documents in the result set are ranked according to the betweenness of their authors so that publications with very central authors are ranked higher in the result list [10, 9]. From a recent study we know that many users are searching DL with author names [5]. In addition, author name recommendations basing on author centrality can be successfully be used as query expansion mechanism [14].

2.4 Implementation

All proposed services are implemented in a live information system using (1) the Solr search engine, (2) Grails Web framework to demonstrate the general feasibility of the approaches. Both Bradfordizing and author centrality as re-rank mechanism are implemented as plugins to the open source web framework Grails. Grails is the glue to combine the different modules and to offer an interactive web-based prototype. In general these retrieval services can be applied in different query phases.

In the following section we will describe a small case study with researchers using the recommendation services STR, JNR and ANR to find search terms, journal names and author names relevant to their research topics.

3 Assessment Study

The assessment study involved 19 researchers in the social sciences who agreed to name one or two of their research topics and take part in a short online assessment exercise. We have recruited the researchers (practitioners[5], PhD students and PostDocs) via email and telephone and they were asked to qualify their primary research topic in the form of 1-3 typical search terms they would enter in a search box. These search terms have been operationalized into a valid query for our prototype by us together with an individualized login for the single researcher. Individualized assessment accounts were sent to the researchers via email for each topic and contained a link to the online assessment tool and a short description how to evaluate the recommendations. All researchers were asked to assess the topical relevance of each recommendation in relationship to their research topic into relevant or not relevant (binary assessments). All researchers got three different assessment screens, always in the same order with a maximum of 5 recommendations for each recommender on one screen: first

---

[5] With practitioners we mean research staff/research associates who are working research related e.g. as information professionals, consultants or service staff, but are not working on a PhD thesis.

6   Philipp Mayr

all search term recommendations, second all author name recommendations and last all journal name recommendations. For each query, researchers got a set of max. 15 recommendations.

This is the list of all 23 evaluated researcher topics: [east Europe; urban sociology; equal treatment; data quality; interviewer error; higher education research; evaluation research; information science; political sociology; party democracy; data quality (2)[6], party system; factor structure; nonresponse; ecology; industrial sociology; sociology of culture; theory of action; atypical employment; lifestyle; Europeanization; survey design; societal change in the newly-formed German states].

4   Evaluation

In the following section we describe the evaluation of the recorded assessments. We calculated average precision AP for each recommender service. The precision P of each service was calculated by

$$P = \frac{r}{r + nr} \tag{1}$$

for each topic, where r is the number of all relevant assessed recommendations and $r+nr$ is the number of all assessed recommendations (relevant and not relevant).

We wanted to keep the assessment exercise for the researchers very short and hence we limited the list of recommendations of each service to a maximum of 5 controlled terms, journal names and author names. According to this restriction we decided to calculate $AP@1$, $AP@2$, $AP@4$ for each service. In very rare case one recommendation service generated just one or two recommendations.

5   Results

In sum 19 researchers assessed 23 topics in the online assessment study. This resulted in total 95 STR, 111 JNR and 107 ANR assessments (see Table 1). In average the researchers assessed 4.1 search term, 4.8 journal name and 4.6 author name recommendations per topic.

Table 1. Statistics of the assessment study

Researchers	Topics	STR Assessm.	JNR Assessm.	ANR Assessm.
19	23	95	111	107

Table 2 shows the evaluation results of all STR, JNR and ANR assessments. For this case study we did no statistically testing like t-test or Wilcoxon because

---

[6] Data quality is mentioned by two researchers and assessed two times.

of our very small sample. The following results should be read as plausibility tests without any statistical significance. We just want to demonstrate here the indicative relevance of this kind of recommender systems for scholarly search systems.

Table 2. Evaluation of the assessments. AP, AP@1, AP@2 and AP@4 for recommendation from STR, JNR and ANR

	STR	JNR	ANR
AP	0.743	0.728	0.749
AP@1	0.957	0.826	0.957
AP@2	0.826	0.848	0.864
AP@4	0.750	0.726	0.750

We can see that the average precision AP of ANR (0.749) and STR (0.743) is slightly better than JNR (0.728). Consulting the $AP@1$ measures ANR and STR are clearly better the JNR. That means that the first recommended author name or search term is rated more often relevant than the first journal name in a list of 4 or 5 recommendations. Surprisingly JNR (0.848) is slightly better than STR (0.826) in $AP@2$. If we look at the last row ($AP@4$) in Table 2 we can see that all three recommendation services move closer together when more recommendation are assessed.

Table 3 shows the average precision AP of STR, JNR and ANR for our three different researcher types (practitioners, PhD students and postdocs). From the 19 researchers in our user study we group 8 researchers into the practitioners group (mostly information professionals without PhD), 8 PhD students which had 1-4 years research experience and a small group of 3 postdocs with 4 and more years research experience. We can see clearly that the author name recommendations are rated highest by the practitioners (see AP of ANR = 0.836). Surprisingly the postdocs have evaluated ANR much lower than the other two groups (see AP of ANR = 0.467). In the experiment postdocs favor journal name recommendations. PhD students rate all three recommenders more or less the same.

Table 3. Evaluation of different researcher types. Average precision for recommendation from STR, JNR and ANR

	STR	JNR	ANR
AP Practitioners (N=8)	0.727	0.709	0.836
AP PhD students (N=8)	0.742	0.719	0.737
AP PostDocs (N=3)	0.750	0.800	0.467

8   Philipp Mayr

6   Conclusion

In this small case study typical researchers in the social sciences are confronted with specific recommendations which were calculated on the basis of researchers research topics. Looking at the precision values two important insights can be noted: (1) precision values of recommendations from STR, JNR and ANR are close together on a very high level - AP is close to 0.75 - and (2) each service retrieved a disjoint set of relevant recommendations. The different services each favor quite other - but still relevant - recommendations and relevance distribution differs largely across topics and researchers. A distinction between researcher types shows that practitioners are favoring author name recommendations (ANR) while postdocs are favoring journal name recommendations precompiled by our recommender services. This can be an artifact due to the small size of the postdoc group but this is also plausible. In terms of research topics author names typically are more distinctive than journal names. An experienced researcher (e.g. postdoc) who is familiar with an authors work can quickly rate an authors name relevance for a specific topic. In this context journal names are not that problematic because they published widely on different topics. This seems to be the case in our small sample (see third row in Table 3). PhD students who typically are unexperienced find all recommendations (terms, author names, journal names) helpful (see second row in Table 3).

The proposed models and derived recommendation services open up new viewpoints on the scientific knowledge space and also provide an alternative framework to structure and search domain-specific retrieval systems [10]. In sum, this case study presents limited results but show clearly that bibliometric-enhanced recommender services can support the retrieval process.

In a next step we plan to evaluate the proposed recommendation services in a larger document assessment task where the services are utilized as query expansion mechanisms [14] and interactive services [17]. However, a lot of research effort needs to be done to make more progress in coupling bibliometric-enhanced recommendation services with IR. The major challenge that we see here is to consider also the dynamic mechanisms which form the structures and activities in question and their relationships to dynamic features in scholarly information retrieval.

7   Acknowledgment

References

1. Bates, M.J.: Where should the person stop and the information search interface start? Information Processing & Management 26(5), 575–591 (jan 1990)
2. Beel, J., Gipp, B., Langer, S., Breitinger, C.: Research-paper recommender systems: a literature survey. International Journal on Digital Libraries (2015)
3. Blair, D.C.: Information retrieval and the philosophy of language. Annual review of information science and technology 37(1), 3–50 (2003)
4. Buchanan, G., Cunningham, S.J., Blandford, A., Rimmer, J., Warwick, C.: Information seeking by humanities scholars. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 3652 LNCS, pp. 218–229 (2005)
5. Carevic, Z., Mayr, P.: Survey on High-level Search Activities based on the Stratagem Level in Digital Libraries. In: 20th International Conference on Theory and Practice of Digital Libraries (TPDL 2016) (2016)
6. Hienert, D., Sawitzki, F., Mayr, P.: Digital Library Research in Action Supporting Information Retrieval in Sowiport. D-Lib Magazine 21(3/4) (2015), `http://www.dlib.org/dlib/march15/hienert/03hienert.html`
7. Hienert, D., Schaer, P., Schaible, J., Mayr, P.: A novel combined term suggestion service for domain-specific digital libraries. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 6966 LNCS, pp. 192–203 (2011)
8. Mayr, P.: Relevance distributions across Bradford Zones: Can Bradfordizing improve search? In: Gorraiz, J., Schiebel, E., Gumpenberger, C., Hörlesberger, M., Moed, H. (eds.) 14th International Society of Scientometrics and Informetrics Conference. pp. 1493–1505. Vienna, Austria (2013), `http://arxiv.org/abs/1305.0357`
9. Mutschke, P., Mayr, P.: Science Models for Search. A Study on Combining Scholarly Information Retrieval and Scientometrics. Scientometrics 102(3), 2323–2345 (2015)
10. Mutschke, P., Mayr, P., Schaer, P., Sure, Y.: Science models as value-added services for scholarly information systems. Scientometrics 89(1), 349–364 (jun 2011), `http://arxiv.org/abs/1105.2441`
11. Nicholas, D., Williams, P., Rowlands, I., Jamali, H.R.: Researchers' e-journal use and information seeking behaviour. Journal of Information Science 36(4), 494–516 (2010)
12. Petras, V.: Translating Dialects in Search: Mapping between Specialized Languages of Discourse and Documentary Languages. Ph.D. thesis, Berkeley, USA (2006)
13. Robertson, S.: Overview of the Okapi projects. Journal of Documentation 53(1), 3–7 (1997)
14. Schaer, P., Mayr, P., Lüke, T.: Extending Term Suggestion with Author Names. In: International Conference on Theory and Practice of Digital Libraries (TPDL 2012). pp. 317–322. Springer Berlin Heidelberg, Paphos, Cyprus (2012)
15. Talja, S., Maula, H.: Reasons for the use and non-use of electronic journals and databases: A domain analytic study in four scholarly disciplines. Journal of Documentation 59(6), 673–691 (2003)
16. White, H.D.: 'Bradfordizing' search output: how it would help online users. Online Review 5(1), 47–54 (1981)
17. Wilson, M.L., Kules, B., Schraefel, m.c., Ben Shneiderman: From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. Foundations and Trends in Web Science 2(1), 1–97 (2010)