Editorial for the 5th Bibliometric-enhanced Information Retrieval Workshop at ECIR 2017

Philipp Mayr[1], Ingo Frommholz[2], and Guillaume Cabanac[3]

[1] GESIS – Leibniz-Institute for the Social Sciences, Cologne, Germany,
`philipp.mayr@gesis.org`
[2] Institute for Research in Applicable Computing, University of Bedfordshire,
Luton, UK,
`ingo.frommholz@beds.ac.uk`
[3] University of Toulouse, Computer Science Department, IRIT UMR 5505, France
`guillaume.cabanac@univ-tlse3.fr`

1   Introduction

Following the successful workshops at ECIR 2014[4], 2015[5] and 2016[6] [1], respectively, this workshop was the fifth in a series of events that brought together experts of communities which often have been perceived as different ones: bibliometrics / scientometrics / informetrics on the one hand and information retrieval on the other. Our motivation as organizers of the workshop started from the observation that main discourses in both fields are different, that communities are only partly overlapping and from the belief that a knowledge transfer would be profitable for both sides [2]. This fifth full-day Bibliometric-enhanced Information Retrieval (BIR) workshop[7] at ECIR 2017 aimed to foster a common ground for the incorporation of bibliometric-enhanced services into scholarly search engine interfaces. In particular we addressed specific communities, as well as studies on large, cross-domain collections like Web of Science, Scopus or Mendeley. This fifth BIR workshop addressed explicitly both scholarly and industrial researchers.

2   Overview of the papers

This year 16 papers were submitted to the workshop, 11 of which were finally accepted for presentation and inclusion in the proceedings: 6 regular papers and 5 posters. The workshop featured one keynote talk, three full paper sessions and one poster session. The following section briefly describes the keynote and sessions.

---

[4] `http://ceur-ws.org/Vol-1143/`

[5] `http://ceur-ws.org/Vol-1344/`

[6] `http://ceur-ws.org/Vol-1567/`

[7] `http://www.gesis.org/en/services/events/events-archive/conferences/ecir-workshops/ecir-workshop-2017/`

2.1 Keynote

The invited paper "Real-World Recommender Systems for Academia: The Pain and Gain in Building, Operating, and Researching them" [3] by Joeran Beel (Trinity College Dublin, Ireland) gives a very insightful overview of the practical experiences in building scholarly document recommender systems for Digital Libraries. The authors Beel and Dinesh report about their research with three different recommender systems which have been implemented and operated in the last six years. They present empirical results of various studies, discuss challenges like running A/B testing with real-world scholarly recommender systems and perform research against competitive benchmarks.

2.2 Session 1: Full papers

In the paper "Manuscript Matcher: A Content and Bibliometrics-based Scholarly Journal Recommendation System" [4], Jason Rollins, Meredith McCusker, Joel Carlson and Jon Stroll present a scholarly journal recommendation system called Manuscript Matcher which is developed and run by Clarivate (formerly Thomson Reuters). The use case of the tool is uploading manuscript title, abstract and references to Manuscript Matcher and getting back bibliometric-informed recommendations of journals ("best fit" publications). The authors present user feedback of the recommendation system and future directions.

In their paper "Use of Locality Sensitive Hashing (LSH) Algorithm to Match Web of Science and SCOPUS" [5], Mehmet Ali Abdulhayoglu and Bart Thijs report on an attempt to match the records of two flagship bibliographic databases. They considered various metadata (e.g., publication title, venue name, bylines) whilst disregarding identifiers such as DOIs, as these are not always available or assigned. Their efficient approach based on LSH found a 70% intersection between these in about an hour. This research contributes to the understanding of the coverage of leading bibliographic databases.

2.3 Session 2: Full papers

The paper "Academic Search in Response to Major Scientific Events" [6] by Li and de Rijke describes search behaviour of academic and web users in occurrence of major scientific events (the Nobel Prize announcements of Chemistry, Physics and Medicine in 2014). The authors compare the query patterns in the query log of the academic search engine ScienceDirect with the data provided by Google Trends. Google Trend is used as a proxy to observe users on the web. They found unique trends for the academic searchers, which are different from users of a web search engine.

The paper "Exploring Choice Overload in Related-Article Recommendations in Digital Libraries" [7] by Beierle, Aizawa and Beel studies choice overload in scholarly document recommendation in the social sciences search engine sowiport. The authors used click-through rate of different amounts of recommendations as a measure of recommendation effectiveness. Their preliminary results show

lower click-through rates for higher numbers of recommendations. According to the experiments, users in the social sciences seem to feel quickly overloaded by increasing choice.

2.4 Session 3: Full papers

The article "Computing Interdisciplinarity of Scholarly Objects using an Author-Citation-Text Model" [8] by Seo, Jung, Kim and Myaeng discusses the computation of the degree of interdisciplinarity of a scholarly object (e.g., an article). To this end, three different sources are used: the author network, the citation network and the actual text. Furthermore, an alternative to measure interdisciplinarity is discussed. Experiments show that the combination of the three aspects author, citations and text of articles can accurately predict the discipline distributions.

In their paper "Detecting Automatically Generated Sentences with Grammatical Structure Similarity" [9], Nguyen Minh Tien and Cyril Labbé tackle the issue of spotting machine generated texts at the sentence level. They introduce a grammatical structure similarity and benchmark it to detect passages stemming from known generators: 80% positive detection rate and less than 1% false detection rate. Editorial workflows could integrate this effective approach to detect questionable manuscripts that editorial staff should check before sending to peer review.

2.5 Poster session

Langer and Beel discuss the use of Lucene in the Docear research paper recommender in their article "Apache Lucene as Content-Based-Filtering Recommender System: 3 Lessons Learned" [10]. They compare Lucene's relevance score to the click-through rate of a document, finding that Lucene's scores indeed can be used to determine relevance. The authors also observed that returning ten recommendations out of the top 50 results might be sensible. Furthermore, Lucene is suitable to approximate the recommendation effectiveness.

In their paper "Extending Scientific Literature Search by Including the Author's Writing Style" [11], Andi Rexha, Mark Kröll, Hermann Ziak, and Roman Kern consider authors' writing style as a potential feature for paper retrieval and recommendation. They report the results of a pilot study questioning the extent to which individuals identify similarities in authorship. This is a challenging task, even for humans.

In his paper "Drakkar: a graph based All-Nearest Neighbour search algorithm for bibliographic coupling" [12], Bart Thijs discusses the creation of bibliographic coupling graphs based on citations. The proposed algorithm utilizes a bipartite graph constituted by the citing publications and the cited references as well as directed citations.

Siebert, Dinesh and Feyer discuss how scientific recommender systems can be improved by incorporating scientometric measures. In their paper "Extending a Research-Paper Recommendation System with Scientometric Measures" [13] the

authors evaluate different reranking approaches in the context of the Mr. DLib research paper recommender system. Readership data is used as an approximation for citation.

In their paper "Semantic embedding for information retrieval", Wang and Koopman [14] combine bibliometric measures with word embeddings. Word embedding results of well-known systems such as Word2Vec/Doc2Vec and GloVe are compared to the Ariadne approach, showing that Ariadne exhibits a competitive performance in a document embedding for information retrieval task.

3 Outlook

With this continuing workshop series we have built up a sequence of explorations, visions, results documented in scholarly discourse, and created a sustainable bridge between bibliometrics and IR.

This year, the authors of accepted papers at the 5th BIR workshop were invited to submit extended versions to a Special Issue on "Bibliometric-enhanced IR" of the *Scientometrics*[8] journal to be published in 2018.

As a next iteration we will organize a Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2017)[9] at the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017). The BIRNDL workshop will be co-organized together with the natural language processing group community and includes a shared task (the CL-SciSumm Shared Task[10]). The shared task tackles automatic paper summarization in the Computational Linguistics (CL) domain.

References

1. Mayr, P., Frommholz, I., Cabanac, G.: Report on the 3rd International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2016). SIGIR Forum **50**(1) (2016) 28–34
2. Mayr, P., Scharnhorst, A.: Scientometrics and Information Retrieval: weak-links revitalized. Scientometrics **102**(3) (2015) 2193–2199
3. Beel, J., Dinesh, S.: Real-World Recommender Systems for Academia: The Pain and Gain in Developing, Operating, and Researching them. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 6–17
4. Rollins, J., McCusker, M., Carlson, J., Stroll, J.: Manuscript Matcher: A Content and Bibliometrics-based Scholarly Journal Recommendation System. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 18–29

---

[8] http://www.springer.com/journal/11192

[9] http://wing.comp.nus.edu.sg/birndl-sigir2017/

[10] http://wing.comp.nus.edu.sg/cl-scisumm2017/

5. Abdulhayoglu, M.A., Thijs, B.: Use of Locality Sensitive Hashing (LSH) algorithm to match Web of Science and SCOPUS. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 30–40

6. Li, X., de Rijke, M.: Academic Search in Response to Major Scientific Events. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 41–50

7. Beierle, F., Aizawa, A., Beel, J.: Exploring Choice Overload in Related-Article Recommendations in Digital Libraries. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 51–61

8. Seo, M.G., Jung, S., Kim, K.m., Myaeng, S.H.: Computing Interdisciplinarity of Scholarly Objects using an Author-Citation-Text Model. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 62–72

9. Nguyen, M.T., Labbé, C.: Detecting Automatically Generated Sentences with Grammatical Structure Similarity. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 73–84

10. Langer, S., Beel, J.: Apache Lucene as Content-Based-Filtering Recommender System: 3 Lessons Learned. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 85–92

11. Rexha, A., Kröll, M., Ziak, H., Kern, R.: Extending Scientific Literature Search by Including the Author's Writing Style. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 93–100

12. Thijs, B.: Drakkar: a graph based All-Nearest Neighbour Search Algorithm for Bibliographic Coupling. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 101–111

13. Siebert, S., Dinesh, S., Feyer, S.: Extending a Research Paper Recommendation System with Bibliometric Measures. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 112–121

14. Wang, S., Koopman, R.: Semantic Embedding for Information Retrieval. In: Proc. of the 5th Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR-WS.org (2017) 122–132