

Editorial for the 2nd Bibliometric-Enhanced Information Retrieval Workshop at ECIR 2015

Philipp Mayr, Ingo Frommholz, Peter Mutschke

1 Introduction

This workshop brought together experts of communities which often have been perceived as different ones: bibliometrics / scientometrics / informetrics on the one side 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.

The first workshop¹ at ECIR 2014 set the research agenda by introducing in each other methods, reporting about current research problems and brainstorming about common interests. See the editorial from 2014 [6] and the workshop proceedings².

This second “Bibliometric-Enhanced Information Retrieval” (BIR 2015) workshop³ continued the overall communication and contributes to create a common ground for the incorporation of bibliometric-enhanced services into retrieval at the scholarly search engine interface. The goal of BIR 2015 was to apply insights from bibliometrics, scientometrics, and informetrics to concrete, practical problems of information retrieval and browsing. To support the previously described goals the workshop topics included the following:

- IR for digital libraries and scientific information portals
- IR for scientific domains, e.g. social sciences, life sciences etc.
- Information Seeking Behaviour
- Bibliometrics, citation analysis and network analysis for IR
- Query expansion and relevance feedback approaches
- Science Modelling (both formal and empirical)
- Task based user modelling, interaction, and personalisation
- (Long-term) Evaluation methods and test collection design
- Collaborative information handling and information sharing
- Classification, categorisation and clustering approaches
- Information extraction (including topic detection, entity and relation extraction)
- Recommendations based on explicit and implicit user feedback

¹ <http://www.gesis.org/en/events/conferences/ecirworkshop2014/>

² <http://ceur-ws.org/Vol-1143/>

³ <http://www.gesis.org/en/events/conferences/ecirworkshop2015/>

We thank our reviewers, authors and keynote speaker for contributing to this workshop and sparking a lively discussion. In the following section we will shortly outline and summarize the papers presented at the workshop. All papers have been peer reviewed by at least two reviewers from the PC.

2 Overview of the papers

In his keynote paper “In Praise of Interdisciplinary Research through Scientometrics” **Guillaume Cabanac** [3] accentuates the potential of interdisciplinary research at the interface of information retrieval and bibliometrics/scientometrics. He comes up with many research questions that lie at the crossroad of scientometrics and other fields, namely information retrieval, digital libraries, psychology and sociology. In his paper he summarizes papers, data sets and methods which have been used in interdisciplinary research.

Citations offer an invaluable source of evidence that can be exploited when searching for relevant documents. However, we need to know how to best handle and extract the citation context to improve retrieval effectiveness. These questions are at the core of **Anna Dabrowska**’s and **Birger Larsen**’s submission “Exploiting Citation Contexts for Physics Retrieval” [5]. The authors examine how to best exploit and integrate the context from citing documents to cited ones in the iSearch collection, which contains roughly half a million documents about physics. Their study evaluates which window size should be used to establish the citation context; their findings suggest that a 25-word window is preferable when it comes to retrieval effectiveness. The authors also discuss the challenges that come along with extracting the citation context. Opening the iSearch collection to a range of future experiments with citation contexts is a further contribution of their work.

In their paper “Factorial Correspondence Analysis Applied to Citation Contexts” **Bertin** and **Atanassova** [2] describe a study of in-text citations focusing on the relationship between the rhetorical structure of papers and the verbs used in the contexts of citations. For this, the study performs a Factorial Correspondence Analysis (CA) using the introduction, methods, results and discussion sections of papers as categories. The paper shows that citation contexts are strongly dependent on the structure of papers.

The paper “An Experimental Platform for Scholarly Article Recommendation” by **Wesley-Smith**, **Dandrea** and **West** [8] describes an experimental platform constructed in collaboration with the open source repository Social Science Research Network (SSRN) in order to test the effectiveness of different approaches for scholarly article recommendations. The paper compares a usage-based, co-download recommendation algorithm with a citation-based Eigenfactor recommender and shows a significant advantage in favor of the co-download recommendation approach.

The position paper “Extending search facilities via bibliometric-enhanced stratagems” by **Carevic** and **Mayr** [4] introduces simple bibliometric-enhanced search facilities which are derived from the famous stratagems by Marcia Bates. The authors argue that the conceptual model by Bates and its stratagems which have been widely dis-

cussed and partly implemented in state of the art digital libraries (DL) could be extended as bibliometric-enhanced stratagems. In their definition is a bibliometric-enhanced stratagem an extended search stratagem which utilizes bibliometric information to re-rank and/or rearrange DL entities in a specific search situation. The authors elaborate on two examples of bibliometric-enhanced stratagems: an extended journal run and an extended citation stratagem.

Are linear ranked lists really the best search engines and literature databases can offer the user? Should there be more interactive means provided and more representational and contextual evidence considered? How do citations fit in here? In their paper "Polyrepresentative Clustering: A Study of Simulated User Strategies and Representations", **Abbasi** and **Frommholz** [1] combine bibliographics and information retrieval by studying how the principle of polyrepresentation can be exploited to support interactive search in literature databases. They combine different representations of a user's information need with document and bibliographic features (i.e. evidence coming from the citation context). Their study, based on aforementioned iSearch collection, suggests incorporating these features increases retrieval effectiveness if we assume even a simple user behaviour pattern. Since their approach is based on document clustering, it can be applied in literature databases which support a broader range of user interaction and seeking strategies, going beyond mere queries and ranked result lists.

3 Outlook

After the ISSI workshop "Combining Bibliometrics and Information Retrieval"⁴ in 2013 we aimed with the BIR workshop series for a dissemination strategy oriented towards core-IR which is the reason why we located this workshop at ECIR. The variety of papers we received and the subset we could accept for this workshop show the different ways of combining bibliometrics and IR and show the mutual benefits the two disciplines can offer each other.

Meanwhile a special issue on "Combining Bibliometrics and Information Retrieval" in *Scientometrics* edited by Philipp Mayr and Andrea Scharnhorst [7] has been published.

We hope to bring both disciplines closer together and start a sequence of explorations, visions, results documented in scholarly discourse, and set up new material for a sustainable bridge between bibliometrics and IR.

⁴ <http://www.gesis.org/en/events/conferences/issiwkshop2013/>

References

1. Abbasi, M. K., & Frommholz, I. (2015). Polyrepresentative Clustering: A Study of Simulated User Strategies and Representations. In Proc. of the 2nd Workshop on Bibliometric-enhanced Information Retrieval (BIR2015) (pp. 47–54). Vienna, Austria: CEUR-WS.org.
2. Bertin, M., & Atanassova, I. (2015). Factorial Correspondence Analysis Applied to Citation Contexts. In Proc. of the 2nd Workshop on Bibliometric-enhanced Information Retrieval (BIR2015) (pp. 22–29). Vienna, Austria: CEUR-WS.org.
3. Cabanac, G. (2015). In Praise of Interdisciplinary Research through Scientometrics. In Proc. of the 2nd Workshop on Bibliometric-enhanced Information Retrieval (BIR2015) (pp. 5–13). Vienna, Austria: CEUR-WS.org.
4. Carevic, Z., & Mayr, P. (2015). Extending search facilities via bibliometric-enhanced stratagems. In Proc. of the 2nd Workshop on Bibliometric-enhanced Information Retrieval (BIR2015) (pp. 40–46). Vienna, Austria: CEUR-WS.org.
5. Dabrowska, A., & Larsen, B. (2015). Exploiting Citation Contexts for Physics Retrieval. In Proc. of the 2nd Workshop on Bibliometric-enhanced Information Retrieval (BIR2015) (pp. 14–21). Vienna, Austria: CEUR-WS.org.
6. Mayr, P., Schaer, P., Scharnhorst, A., & Mutschke, P. (2014). Editorial for the Bibliometric-enhanced Information Retrieval Workshop at ECIR 2014. In P. Mayr et al. (Ed.), Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval (pp. 1–4). Amsterdam, NL. Retrieved from <http://ceur-ws.org/Vol-1143/editorial.pdf>
7. Mayr, P., & Scharnhorst, A. (2015). Scientometrics and Information Retrieval - weak-links revitalized. *Scientometrics*, 102(3), 2193–2199. doi:10.1007/s11192-014-1484-3
8. Wesley-Smith, I., Dandrea, R., & West, J. (2015). An Experimental Platform for Scholarly Article Recommendation. In Proc. of the 2nd Workshop on Bibliometric-enhanced Information Retrieval (BIR2015) (pp. 30–39). Vienna, Austria: CEUR-WS.org.