

Editorial for the 7th Bibliometric-enhanced Information Retrieval Workshop at ECIR 2018

Philipp Mayr¹, Ingo Frommholz², and Guillaume Cabanac³

¹ GESIS – Leibniz-Institute for the Social Sciences, Cologne, Germany,
philipp.mayr@gesis.org

² Institute for Research in Applicable Computing, University of Bedfordshire, UK,
ifrommholz@acm.org

³ University of Toulouse, Computer Science Department, IRIT UMR 5505, France
guillaume.cabanac@univ-tlse3.fr

Abstract. The Bibliometric-enhanced Information Retrieval (BIR) workshop series has started at ECIR in 2014 and serves as the annual gathering of IR researchers who address various information-related tasks on scientific corpora and bibliometrics. We welcome contributions elaborating on dedicated IR systems, as well as studies revealing original characteristics on how scientific knowledge is created, communicated, and used. This editorial presents all accepted papers at the 7th BIR workshop at ECIR 2018 in Grenoble, France.

Keywords: Bibliometrics, Scientometrics, Informetrics, Information Retrieval, Digital Libraries

1 Introduction

The Bibliometric-enhanced Information Retrieval (BIR) workshop series has started at ECIR in 2014 [1] and serves as the annual gathering of IR researchers who address various information-related tasks on scientific corpora and bibliometrics [2]. The workshop features original approaches to search, browse, and discover value-added knowledge from scientific documents and related information networks (e.g., terms, authors, institutions, references). We welcome contributions elaborating on dedicated IR systems, as well as studies revealing original characteristics on how scientific knowledge is created, communicated, and used.

The current incarnation is a continuation of the evolution of our workshop series. The first BIR workshops set the research agenda by introducing the workshop topics, illustrating state-of-the-art methods, reporting on current research problems, and brainstorming about common interests. For the fourth workshop, co-located with the ACM/IEEE-CS JCDL 2016, we broadened the workshop scope and interlinked the BIR workshop with the natural language processing (NLP) and computational linguistics field [3].

This 7th full-day BIR workshop at ECIR 2018⁴ aims to foster a common ground for the incorporation of bibliometric-enhanced services (including text

⁴ <http://bit.ly/bir2018>

mining functionality) into scholarly search engine interfaces. In particular we address specific communities, as well as studies on large, cross-domain collections. This workshop strives to feature contributions from core bibliometricians and core IR specialists who already operate at the interface between scientometrics and IR.

2 Overview of the papers

This year’s workshop hosts two keynotes as well as a set of regular papers and two demos. The publications are briefly introduced in the following subsections.

2.1 Keynotes

In this workshop we will have two inspirational keynotes to kick-start thinking and discussion on the workshop topic. This will be followed by paper presentations and demos in a format that we found to be successful at previous BIR workshops.

Cyril Labbé tackles a hot topic in his keynote titled “Trends in gaming indicators: On failed attempts at deception and their computerised detection” [4]. He outlines various efforts to manipulate indicators by tricking the scientific community (e.g., by submitting automatically generated papers). Other issues undermining the trust we place in peer-reviewed science are examined, such as data–results mismatch impeding the reproduction of results in cancer research. Labbé surveys his recent work in these areas while reflecting on the potential of B+IR (bibliometrics and information retrieval) to address these critical issues.

Ralf Schenkel presents in his keynote “Integrating and exploiting metadata sources in a bibliographic information system” [5] an in-depth summary of recent metadata activities in the computer science bibliography DBLP which is maintained by Schloss Dagstuhl and University of Trier. He outlines procedures for monitoring, selecting and prioritizing computer science venue for inclusion in the DBLP bibliography. A special focus is given to author disambiguation and utilization of citation data.

2.2 Regular papers

Sarol, Liu, and Schneider propose a citation and text-based publication retrieval framework [6]. After the user provides some seed articles, the system collects papers connected by citations and applies a combination of citation- and text-based filtering methods. The framework is evaluated in a systematic reviewing task.

Ollagnier, Fournier, and Bellot highlight the central references of a paper based on the mining of its fulltext, quantifying the occurrences of all in-text references [7]. They benchmarked this approach compared to a system in production at OpenEdition,⁵ and discuss the results in terms of enhanced relevance.

⁵ <https://www.openedition.org>

In their article on query expansion, Rattinger, Le Goff, and Guetl combine word embeddings and co-authorship relations [8]. The set of documents used for pseudo-relevance feedback is enriched by similar documents from co-authors, applying a locally trained Word2Vec model. Adding similar documents from co-authors significantly improved the baseline.

Bertin and Atanassova report on the construction of the InTeReC dataset [9]. Utilising different section types from PLOS articles, InTeReC consists of within text references and their surrounding sentences. Additionally, verb phrases are extracted that provide an idea of the nature of the reference.

Kacem and Mayr investigated the usage and influence of a specific search stratagem – the Journal Run – in an academic search engine log file [10]. They studied the frequency and stage of use of journal run as well as its impact on sessions. The authors found that the frequency of usage of the analyzed journals is not related to the impact factor within these sessions and that the size of the journal (Bradford Zones) has an insignificant correlation.

2.3 Demo papers

Cataldi, Di Caron, and Schifanella designed the d -index to evaluate the degree of dependence of a researcher with respect to his/her co-authors over time. They implemented this indicator and demonstrate it online⁶ with DBLP as a bibliographic datasource [11].

The demo paper by Bessagnet presents a framework which combines thematic, temporal and spatial features of Twitter tweets in the field of Human and Social Sciences [12]. The author proposes 5 W dimensions (who, when, what, where, why) for the analysis of tweets.

3 Outlook

While the past workshops laid the foundations for further work and also made the benefit of bringing information retrieval and bibliometrics together more explicit, there are still many challenges ahead. One of them is to provide infrastructures and testbeds for the evaluation of retrieval approaches that utilise bibliometrics and scientometrics. To this end, a focus of the proposed workshop and the discussion will be on real experimentations (including demos) and industrial participation. This line was started in a related workshop at JCDL (BIRNDL 2016) and continued at SIGIR (BIRNDL 2017), but with a focus on digital libraries and computational linguistics. Given the complex information needs scholars are usually facing, we will emphasize information retrieval and information seeking and searching aspects.

In 2015 we published a first special issue on “Combining Bibliometrics and Information Retrieval” in *Scientometrics* [2]. A special issue on “Bibliometrics, Information Retrieval and Natural Language Processing in Digital Libraries” will

⁶ <http://d-index.di.unito.it>

appear in 2018 in the *International Journal on Digital Libraries* [13]. Another special issue on “Bibliometric-enhanced Information retrieval and Scientometrics” is in preparation for the *Scientometrics* journal.

Since 2016 we maintain the “Bibliometric-enhanced-IR Bibliography”⁷ which collects scientific papers which appear in collaboration with the BIR/BIRNDL organizers.

4 Acknowledgement

We wish to thank all those who have contributed to the workshop proceedings: all the contributing authors and the many reviewers who generously offered their time and expertise.

References

1. Mayr, P., Scharnhorst, A., Larsen, B., Schaer, P., Mutschke, P.: Bibliometric-Enhanced Information Retrieval. In: 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands (2014) 798–801
2. Mayr, P., Scharnhorst, A.: Scientometrics and Information Retrieval: Weak-links revitalized. *Scientometrics* **102**(3) (2015) 2193–2199
3. Cabanac, G., Chandrasekaran, M.K., Frommholz, I., Jaidka, K., Kan, M.Y., Mayr, P., Wolfram, D.: Report on the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2016). *SIGIR Forum* **50**(2) (2016) 36–43
4. Labbé, C.: Trends in gaming indicators: On failed attempts at deception and their computerised detection. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 6–15
5. Schenkel, R.: Integrating and exploiting public metadata sources in a bibliographic information system. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 16–21
6. Sarol, M.J., Liu, L., Schneider, J.: Testing a citation and text-based framework for retrieving publications for literature reviews. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 22–33
7. Ollagnier, A., Fournier, S., Bellot, P.: BIBLME RecSys: Harnessing bibliometric measures for a scholarly paper recommender system. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 34–45
8. Rattinger, A., Goff, J.M.L., Guetl, C.: Local word embeddings for query expansion based on co-authorship and citations. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 46–53
9. Bertin, M., Atanassova, I.: InTeReC: An in-text reference corpus for applying natural language processing to bibliometrics. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 54–62

⁷ https://github.com/PhilippMayr/Bibliometric-enhanced-IR_Bibliography/

10. Kacem, A., Mayr, P.: Users are not influenced by high impact and core journals while searching. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 63–75
11. Cataldi, M., Caro, L.D., Schifanella, C.: All for one or one for all? Analyzing collaboration patterns in research environments. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 76–79
12. Bessagnet, M.N.: A generic framework to perform comprehensive analysis of tweets. In: Proc. of the Seventh Workshop on Bibliometric-enhanced Information Retrieval (BIR), Grenoble, France, CEUR-WS.org (2018) 80–85
13. Mayr, P., Frommholz, I., Cabanac, G., Chandrasekaran, M.K., Jaidka, K., Kan, M.Y., Wolfram, D.: Introduction to the Special Issue on Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL). International Journal on Digital Libraries (2017)