Bibliometric-Enhanced Information Retrieval: 11th International BIR Workshop

Ingo Frommholz[1(✉)], Philipp Mayr[2,3], Guillaume Cabanac[4],
and Suzan Verberne[5]

[1] School of Mathematics and Computer Science, University of Wolverhampton,
Wolverhampton, UK
`ifrommholz@acm.org`
[2] GESIS – Leibniz-Institute for the Social Sciences, Cologne, Germany
`philipp.mayr@gesis.org`
[3] Institute of Computer Science, University of Göttingen, Göttingen, Germany
[4] Computer Science Department, University of Toulouse, IRIT UMR 5505,
Toulouse, France
`guillaume.cabanac@univ-tlse3.fr`
[5] LIACS, Leiden University, Leiden, The Netherlands
`s.verberne@liacs.leidenuniv.nl`

Abstract. The Bibliometric-enhanced Information Retrieval (BIR) workshop series at ECIR tackles issues related to academic search, at the intersection of Information Retrieval, Natural Language Processing and Bibliometrics. BIR is a hot topic investigated by both academia and the industry. In this overview paper, we summarize the 11th iteration of the workshop and present the workshop topics for 2021.

Keywords: Academic search · Information retrieval · Digital libraries · Bibliometrics · Scientometrics

1 Motivation and Relevance to ECIR

Bibliometric-enhanced IR is a hot topic with growing recognition in the information retrieval as well as the scientometrics community in recent years. This is motivated by the exploding number of scholarly publications and the need to satisfy scholars' specific information needs to find relevant research contribution for their own work. As a very recent example, the COVID-19 crisis and the large number of scientific publications triggered by it has made the effective and efficient scholarly search for and discovery of high-quality publications, for instance in pre-print repositories to ensure the quick dissemination of crucial research results, a priority. Bibliometric-enhanced information retrieval tries to provide solutions to the peculiar needs of scholars to keep on top of the research in their respective fields, utilising the wide range of suitable relevance signals

that come with academic scientific publications, such as keywords provided by authors, topics extracted from the full-texts, co-authorship networks, citation networks, and various classification schemes of science. Bibliometric-enhanced IR systems must deal with the multifaceted nature of scientific information by searching for or recommending academic papers, patents, venues (i.e., conferences or journals), authors, experts (e.g., peer reviewers), references (to be cited to support an argument), and datasets.

The purpose of the BIR workshop series founded in 2014 [3] is to tackle these challenges by tightening up the link between IR and Bibliometrics. We strive to bring the 'retrievalists' and 'citationists' [4] active in both academia and industry together. The success of past BIR events (Table 1) evidences that BIR@ECIR is a much needed scientific event for the different communities involved to meet and join forces to push the knowledge boundaries of IR applied to literature search and recommendation.

Table 1. Overview of the BIR workshop series and CEUR proceedings

Year	Conference	Venue		Papers	Proceedings
2014	ECIR	Amsterdam,	NL	6	Vol-1143
2015	ECIR	Vienna,	AT	6	Vol-1344
2016	ECIR	Padua,	IT	8	Vol-1567
2016	JCDL	Newark,	US	$10 + 10^a$	Vol-1610
2017	ECIR	Aberdeen,	UK	12	Vol-1823
2017	SIGIR	Tokyo,	JP	11	Vol-1888
2018	ECIR	Grenoble,	FR	9	Vol-2080
2019	ECIR	Cologne,	DE	14	Vol-2345
2019	SIGIR	Paris,	FR	$16 + 10^b$	Vol-2414
2020	ECIR	Lisbon (Online),	PT	9	Vol-2591
2021	ECIR	Lucca (Online),	IT	9	TBA

[a]With CL-SciSumm 2016 Shared Task; [b]With CL-SciSumm 2019 Shared Task

2   Objectives and Topics

The call for papers for the 2021 workshop (the 11th BIR edition) addressed current research issues regarding 3 aspects of the search/recommendation process:

1. User needs and behaviour regarding scientific information, such as:
   – Finding relevant papers/authors for a literature review.
   – Measuring the degree of plagiarism in a paper.
   – Identifying expert reviewers for a given submission.
   – Flagging predatory conferences and journals.

  – Information seeking behaviour and HCI in academic search.
2. Mining the scientific literature, such as:
    – Information extraction, text mining and parsing of scholarly literature.
    – Natural language processing (e.g., citation contexts).
    – Discourse modelling and argument mining.
3. Academic search/recommendation systems, such as:
    – Modelling the multifaceted nature of scientific information.
    – Building test collections for reproducible BIR.
    – System support for literature search and recommendation.

At the time of writing, three keynote speakers accepted our invitation to give an invited talk during the workshop: Ludo Waltman, Lucy Lu Wang and Jimmy Lin.

Ludo Waltman is professor and deputy director at the Centre for Science and Technology Studies (CWTS) at Leiden University. He leads the Quantitative Science Studies (QSS) research group at CWTS, which does research in the fields of bibliometrics and scientometrics. Ludo is coordinator of the CWTS Leiden Ranking, a bibliometric ranking of major universities worldwide.

Lucy Lu Wang is a postdoctoral investigator at the Allen Institute for AI (AI2) in the Semantic Scholar research group. Lucy works in the areas of knowledge representation and biomedical ontologies, natural language processing applications for biomedical and scientific text, open access, and meta-science. She is one of the authors of the COVID-19 open research dataset (CORD-19).

Jimmy Lin is a professor and the David R. Cheriton Chair in the David R. Cheriton School of Computer Science at the University of Waterloo, Canada. His main work is at the intersection of information retrieval, natural language processing, databases and data management. He has spearheaded the development of a search engine that provides access to over 45,000 scholarly articles about COVID-19 in the Allen Institute for AI's CORD-19 research dataset.

3   Target Audience

The target audience of the BIR workshops are researchers and practitioners, junior and senior, from Scientometrics as well as Information Retrieval and Natural Language Processing. These could be IR/NLP researchers interested in potential new application areas for their work as well as researchers and practitioners working with bibliometric data and interested in how IR/NLP methods can make use of such data. The 10th anniversary edition in 2020 ran online with an audience peaking at 97 online participants [1]. In December 2020, we published our third special issue emerging from the past BIR workshops [2].

4   Peer Review Process and Workshop Format

Our peer review process is supported by Easychair. Each submission is assigned to 2 to 3 reviewers, preferably at least one expert in IR and one expert in Bibliometrics or NLP. The programme committee for 2021 consists of peer reviewers from all participating communities. Accepted papers are either long papers (15-min talks) or short papers (5-min talks). Two interactive sessions close the morning and afternoon sessions with posters and demos, allowing attendees to discuss the latest developments in the field and opportunities (e.g., shared tasks such as CL-SciSumm). These interactive sessions serve as ice-breakers, sparking interesting discussions that, in non-pandemic times, usually continue during lunch and the cocktail party. The sessions are also an opportunity for our speakers to further discuss their work.

As a follow-up of the workshop, the co-chairs will write a report summing up the main themes and discussions to *SIGIR Forum* [1, for instance] and BCS Informer[1], as a way to advertise our research topics as widely as possible among the IR community.

5   Next Steps

Research on scholarly document processing has for many years been scattered across multiple venues such as ACL, SIGIR, JCDL, CIKM, LREC, NAACL, KDD, and others. Our next strategic step is the Second Workshop on Scholarly Document Processing (SDP)[2] that will be held in June 2021 in conjunction with the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). This workshop and initiative will be organized by a diverse group of researchers (organizers from BIR, BIRNDL, Workshop on Mining Scientific Publications/WOSP and Big Scholar) which have expertise in NLP, ML, Text Summarization/Mining, Computational Linguistics, Discourse Processing, IR, and others.

References

1. Cabanac, G., Frommholz, I., Mayr, P.: Report on the 10th anniversary workshop on bibliometric-enhanced information retrieval (BIR 2020). In: SIGIR Forum, vol. 54, no. 1 (2020). https://doi.org/10.1145/3451964.3451974

---

[1] BIR 2020 appeared in https://irsg.bcs.org/informer/category/spring-2020/.

[2] https://sdproc.org/2021/.

2. Cabanac, G., Frommholz, I., Mayr, P.: Scholarly literature mining with information retrieval and natural language processing: preface. Scientometrics **125**(3), 2835–2840 (2020). https://doi.org/10.1007/s11192-020-03763-4
3. Mayr, P., Scharnhorst, A., Larsen, B., Schaer, P., Mutschke, P.: Bibliometric-enhanced information retrieval. In: de Rijke, M., et al. (eds.) ECIR 2014. LNCS, vol. 8416, pp. 798–801. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06028-6_99
4. White, H.D., McCain, K.W.: Visualizing a discipline: an author co-citation analysis of Information Science, 1972–1995. J. Am. Soc. Inf. Sci. **49**(4), 327–355 (1998). https://doi.org/b57vc7