Bibliometric-Enhanced Information Retrieval: 13th International BIR Workshop (BIR 2023)

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

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

Abstract. The 13th iteration of the Bibliometric-enhanced Information Retrieval (BIR) workshop series will take place at ECIR 2023 as a full-day workshop. BIR tackles issues related to, for instance, academic search and recommendation, at the intersection of Information Retrieval, Natural Language Processing, and Bibliometrics. As an interdisciplinary scientific event, BIR brings together researchers and practitioners from the Scientometrics/Bibliometrics community on the one hand, and the Information Retrieval community on the other hand. BIR is an ever-growing topic investigated by both academia and the industry.

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

1 Motivation and Relevance to ECIR

The aim of the Bibliometric-enhanced Information Retrieval (BIR)[1] workshop series and its 13th iteration at ECIR 2023[2] is to bring together researchers and practitioners from Scientometrics/Bibliometrics as well as Information Retrieval (IR). Scientometrics is a sub-field of Bibliometrics which, like IR, is in turn a sub-field of Information Science. Bibliometrics and Scientometrics are concerned with all quantitative aspects of information and academic literature [5], which naturally make them interesting for IR research, in particular when it comes to academic search, recommendation, and other domains in which citations play a

[1] https://sites.google.com/view/bir-ws/home?authuser=0.
[2] https://sites.google.com/view/bir-ws/bir-2023?authuser=0.

J. Kamps et al. (Eds.): ECIR 2023, LNCS 13982, pp. 392–397, 2023.
https://doi.org/10.1007/978-3-031-28241-6_43

central role, for example, legal and patent retrieval. In the early 1960 s,s, Salton was already striving to enhance IR by including clues inferred from bibliographic citations [6]. In the course of decades, both disciplines (Bibliometrics and IR) evolved apart from each other over time, leading to the two loosely connected fields we know today [7].

However, the exploding number of scholarly publications and the requirement to satisfy scholars' specific information needs led Bibliometric-enhanced IR to receive growing recognition in the IR as well as the Scientometrics communities. Challenges in academic search and recommendation became particularly apparent during the COVID-19 crisis as well as the ongoing trend of publishing on preprint servers first (and sometimes exclusively), for instance in the rapidly developing field of artificial intelligence and deep learning. This results, for instance, in the information overload researchers and practitioners are facing, as well as the need to ensure the timeliness and quality of published research. Tackling these challenges requires effective and efficient solutions for scholarly search, recommendation and discovery of high-quality publications and heterogeneous data. Bibliometric-enhanced IR tries to provide these 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, altmetrics, bibliometric figures, 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., conference proceedings, journals, books, manuals, grey literature), authors, experts (e.g., peer reviewers), references (to be cited to support an argument), and datasets. A further discussion of the various research directions in bibliometric-enhanced IR can be found in [3].

To this end, the BIR workshop series was founded in 2014 [4] to tackle these challenges by tightening up the link between IR and Bibliometrics. We strive to bring the 'retrievalists' and 'citationists' [7] active in both academia and industry together. The success of past BIR events, as shown in Table 1, evidences that BIR@ECIR is a much-needed interdisciplinary scientific event that attracts researchers and practitioners from IR, Bibliometrics, and Natural Language Processing alike.

2 Workshop Goals and Objectives

Our vision is to bring together researchers and practitioners from Scientometrics/Bibliometrics on the one hand and IR, on the other hand, to create better methods and systems for instance for academic search and recommendation. Our view is to expose people from one community to the work of the respective other community and to foster fruitful interaction across communities. Therefore, in the call for papers for the 2023 BIR workshop at ECIR, we address, but are not limited to, current research issues regarding 3 aspects of the academic search/recommendation process:

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	Vol-2847
2022	ECIR	Stavanger, NO	5	Vol-3230

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

1. User needs and behaviour regarding scientific information, such as:
   - Finding relevant papers/authors for a literature review.
   - Filtering high-quality research papers, e.g. in preprint servers.
   - Measuring the degree of plagiarism in a paper.
   - Identifying expert reviewers for a given submission.
   - Flagging predatory conferences and journals, or other forms of scientific misbehaviour
   - Understanding 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.

3 Target Audience and Dissemination

The target audience of the BIR workshops is researchers and practitioners, junior and senior, from Scientometrics as well as IR and Natural Language Processing (NLP). 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. BIR 2022 will be open to anyone interested in the topic.

The BIR organisers are well established in the IR and the Bibliometrics community, respectively and have a years-long experience as workshop organizers.

We will send the call for papers to major professional mailing lists in IR (ACM SIGIR, IR-List, JISC IR) and Bibliometrics (ASIS&T Sigmetrics and ISSI). We will also send it to the former BIR and BIRNDL participants (in the range of a few hundred people) and scientists who publish in both IR and Bibliometrics venues, based on the mining of the DBLP. We will further advertise the call for papers through our social media channels.

The 10th-anniversary edition in 2020 ran online with an audience peaking at 97 online participants [1]. BIR 2022, the 12th edition, was run as a hybrid event with on-site and online participants. We were surprised by how successful this model went and how satisfied the speakers, audience, and organizers were with the hybrid workshop.

In December 2020, we published our third special issue emerging from the past BIR workshops [2].

4 Workshop Format, Structure and Peer Review Process

Our peer review process will be 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 2023 will consist 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). We also invite attendees to demonstrate prototypes during flash presentations (5 min).

These interactive sessions serve as ice-breakers, sparking interesting discussions that usually continue during lunch and the evening social event. The sessions also allow our speakers to further discuss their work. BIR has a friendly and open atmosphere where there is an opportunity for participants (including students) to share their ideas and current work and to receive feedback from the community.

5 Previous Workshops

The BIR workshop series has a long tradition of taking place along major IR conferences such as ECIR, SIGIR, and JCDL, as documented on the BIR overview page. Table 1 provides an overview of past events. BIR@ECIR2023 would be the continuation of a highly successful conference series.

6 Organizers

Ingo Frommholz is Reader (Associate Professor equivalent) in Data Science at the University of Wolverhampton, UK. He has been working on formal IR systems and models taking user aspects into account. He has been co-organizing several IR-related research events, for instance, the IR and Foraging Autumn School

at Schloß Dagstuhl, Germany, the Future Directions in Information Access (FDIA) symposia, BCS Search Solutions, BIRDS 2020 and 2021 (at SIGIR and CHIIR, respectively) and the ongoing Bibliometrics and IR (BIR@ECIR) workshop series. He is managing editor of the *International Journal on Digital Libraries* (IJDL, Springer).

Philipp Mayr is a team leader at the GESIS – Leibniz-Institute for the Social Sciences department Knowledge Technologies for the Social Sciences. His research group *Information and Data Retrieval* is working on methods and techniques of interactive information and dataset retrieval and maintains and further develops information systems for the social sciences. He has been co-organizing several IR-related research events, for instance, the Workshops on Scholarly Document Processing at EMNLP/NAACL/COLING and the ongoing Bibliometrics and IR (BIR@ECIR) workshop series.

Guillaume Cabanac is a Professor of Computer Science at the University of Toulouse and holds a research chair at the Institut Universitaire de France. His interdisciplinary research on the quantitative study of science is at the crossroads between Information Retrieval, Digital Libraries, and Scientometrics. He serves on the editorial boards of the *Journal of the Association for Information Science and Technology (JASIST)* and *Scientometrics*. His current work on the Problematic Paper Screener contributes to the identification and reporting of algorithmically generated and fraudulent papers published-often sold-by academic publishers. Cabanac was nicknamed 'Deception sleuth' in the *Nature's 10* list of 'ten people who helped shape science in 2021.'

Suzan Verberne is an associate professor at the Leiden Institute of Advanced Computer Science (LIACS) at Leiden University and group leader of Text Mining and Retrieval Leiden. She obtained her PhD in 2010 on the topic of Question Answering and has since then been working on the edge between Natural Language Processing (NLP) and Information Retrieval (IR). Her recent work centers around interactive information access for specific domains, covering academic, biomedical, legal, and patent data. She is highly active in the NLP and IR communities, holding chairing positions in large worldwide conferences, and is an associate editor for *Transactions on Information Systems*.

References

1. Cabanac, G., Frommholz, I., Mayr, P.: Report on the 10th anniversary workshop on bibliometric-enhanced information retrieval (BIR 2020). SIGIR Forum **54**(1) (2020). https://doi.org/10.1145/3451964.3451974
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. Frommholz, I., Cabanac, G., Mayr, P., Verberne, S.: Report on the 11th bibliometric-enhanced information retrieval workshop (BIR 2021). SIGIR Forum **55**(1) (2021). https://doi.org/10.1145/3476415.3476426

4. 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
5. Pritchard, A.: Statistical bibliography or bibliometrics? [Documentation notes]. J. Document. **25**(4), 348–349 (1969). https://doi.org/10.1108/eb026482
6. Salton, G.: Associative document retrieval techniques using bibliographic information. J. ACM **10**(4), 440–457 (1963). https://doi.org/10.1145/321186.321188
7. 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). b57vc7