Bibliometric-Enhanced Information Retrieval: 8th International BIR Workshop

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

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

Abstract. The Bibliometric-enhanced Information Retrieval workshop series (BIR) at ECIR tackles issues related to academic search, at the crossroads between Information Retrieval and Bibliometrics. BIR is a hot topic investigated by both academia (e.g., ArnetMiner, CiteSeer^x, Doc-Ear) and the industry (e.g., Google Scholar, Microsoft Academic Search, Semantic Scholar). An 8th iteration of the one-day BIR workshop was held at ECIR 2019.

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

1 Motivation and Relevance to ECIR

Searching for scientific information is a long-lived information need. In the early 1960s, Salton was already striving to enhance information retrieval by including clues inferred from bibliographic citations [21]. The development of citation indexes pioneered by Garfield [6] proved determinant for such a research endeavour at the crossroads between the nascent fields of Information Retrieval (IR) and Bibliometrics[1]. The pioneers who established these fields in Information Science—such as Salton and Garfield—were followed by scientists who specialised in one of these [26], leading to the two loosely connected fields we know of today.

The purpose of the BIR workshop series founded in 2014 is to tighten up the link between IR and Bibliometrics. We strive to get the 'retrievalists' and 'citationists' [26] active in both academia and the industry together, who are

---

[1] Bibliometrics refers to the statistical analysis of the academic literature [20] and plays a key role in scientometrics: the quantitative study of science and innovation [9].

developing search engines and recommender systems such as ArnetMiner [24], CiteSeer$^{\chi}$ [27], DocEar [1], Google Scholar [25], Microsoft Academic Search [23], and Semantic Scholar [2], just to name a few.

Bibliometric-enhanced IR systems must deal with the multifaceted nature of scientific information by searching for or recommending academic papers, patents [7], venues (i.e., conferences or journals), authors, experts (e.g., peer reviewers), references (to be cited to support an argument), and datasets. The underlying models harness relevance signals from keywords provided by authors, topics extracted from the full-texts, coauthorship networks, citation networks, and various classifications schemes of science.

Bibliometric-enhanced IR is a hot topic whose recent developments made the news—see for instance the Initiative for Open Citations [22] and the Google Dataset Search [5] launched on September 4, 2018. We believe that BIR@ECIR is a much needed scientific event for the 'retrievalists' and 'citationists' to meet and join forces pushing the knowledge boundaries of IR applied to literature search and recommendation.

2   Past Related Activities

The BIR workshop series was launched at ECIR in 2014 [18] and it was held at ECIR each year since then [12–14,17]. As our workshop has been lying at the crossroads between IR and NLP, we also ran it as a joint workshop called BIRNDL (for Bibliometric-enhanced IR and NLP for Digital Libraries) at the JCDL [3] and SIGIR [10,11] conferences. All workshops had a large number of participants, demonstrating the relevance of the workshop's topics. The BIR and BIRNDL workshop series gave the community the opportunity to discuss latest developments and shared tasks such as the CL-SciSumm [8], which was introduced at the BIRNDL joint workshop.

The authors of the most promising workshop papers were offered the opportunity to submit an extended version for a Special Issue for the *Scientometrics* journal [4,19] and of the *International Journal on Digital Libraries* [16].

The target audience of our workshop are researchers and practitioners, junior and senior, from Scientometrics as well as Information Retrieval. These could be IR researchers interested in potential new application areas for their work as well as researchers and practitioners working with, for instance, bibliometric data and interested in how IR methods can make use of such data.

3   Objectives and Topics for BIR@ECIR 2019

We called for original research at the crossroads of IR and Bibliometrics. The accepted papers report on new approaches using bibliometric clues to enhance the search or recommendation of scientific information or significant improvements of existing techniques. Thorough quantitative studies of the various corpora to be indexed (papers, patents, networks or else) were also welcome.

Fig. 1. Main topics of the BIR and BIRNDL workshop series (2014–2018) as extracted from the titles of the papers published in the proceedings, see https://dblp.org/search?q=BIR.ECIR

The topics of the workshop are in line with those of the past BIR and BIRNDL workshops (Fig. 1): a mixture of IR and Bibliometric concepts and techniques. More specifically, the call for papers featured current research issues regarding three 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.
2. The characteristics of scientific information:
   – Measuring the reliability of bibliographic libraries;
   – Spotting research trends and research fronts.
3. Academic search/recommendation systems:
   – Modelling the multifaceted nature of scientific information;
   – Building test collections for reproducible BIR.

4   Peer Review Process and Organization

The 8th BIR edition ran as a one-day workshop, as it was the case for the previous editions. Keynote talks by leading scientists working at the crossroads between IR and Scientometrics kicked off the day.

Two types of papers were presented: long papers (15-min talks) and short papers (5-min talks). Two interactive sessions closed the morning and evening sessions with posters and demos. These sessions allowed us to discuss the latest developments in the field and opportunities. The interactive sessions were announced with the workshop program. We invited anyone attending to demonstrate their prototypes during flash presentations (5 min). These interactive sessions served as ice-breakers, sparking interesting discussions that usually continued during lunch and the cocktail party. The sessions were also an opportunity for our speakers to further discuss their work.

We ran the workshop with peer review supported by EasyChair[2]. Each submission was assigned to 2 to 3 reviewers, preferably at least one expert in IR and one expert in Bibliometrics. The stronger submissions were accepted as long papers while weaker ones were accepted as short papers, posters, or demos. All authors were instructed to revise their submission according to the reviewers' reports. All accepted papers are planned to be in the workshop proceedings hosted at ceur-ws.org, an established open access repository with no author-processing charges.

As a follow-up of the workshop, the co-chairs will write a report summing up the main themes and discussions to *SIGIR Forum* [15, for instance] and BCS Informer[3], as a way to advertise our research topics as widely as possible among the IR community. All authors are encouraged to submit an extended version of their papers to the Special Issue of the *Scientometrics* journal that will be announced in Spring 2019.

References

1. Beel, J., Langer, S., Gipp, B., Nürnberger, A.: The architecture and datasets of Docear's research paper recommender system. D-Lib Mag. **20**(11/12) (2014). https://doi.org/10.1045/november14-beel
2. Bohannon, J.: A computer program just ranked the most influential brain scientists of the modern era. Science (2016). https://doi.org/10.1126/science.aal0371
3. Cabanac, G., et al. (eds.): BIRNDL 2016: Proceedings of the Joint Workshop on Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries Co-located with the Joint Conference on Digital Libraries, vol. 1610. CEUR-WS, Aachen (2016)
4. Cabanac, G., Frommholz, I., Mayr, P.: Bibliometric-enhanced information retrieval: preface. Scientometrics **116**, 1225–1227 (2018). https://doi.org/10.1007/s11192-018-2861-0
5. Castelvecchi, D.: Google unveils search engine for open data. Nature (2018). https://doi.org/10.1038/d41586-018-06201-x. (News & Comment)
6. Garfield, E.: Citation indexes for science: a new dimension in documentation through association of ideas. Science **122**(3159), 108–111 (1955). https://doi.org/10.1126/science.122.3159.108
7. Garfield, E.: Patent citation indexing and the notions of novelty, similarity, and relevance. J. Chem. Doc. **6**(2), 63–65 (1966). https://doi.org/10.1021/c160021a001

---

8. Jaidka, K., Chandrasekaran, M.K., Rustagi, S., Kan, M.Y.: Insights from CL-SciSumm 2016: the faceted scientific document summarization shared task. Int. J. Digit. Libr. **19**(2–3), 163–171 (2018). https://doi.org/10.1007/s00799-017-0221-y

9. Leydesdorff, L., Milojević, S.: Scientometrics. In: Wright, J.D. (ed.) International Encyclopedia of the Social & Behavioral Sciences, vol. 21, 2nd edn, pp. 322–327. Elsevier, Amsterdam (2015). https://doi.org/10.1016/b978-0-08-097086-8.85030-8

10. Mayr, P., Chandrasekaran, M.K., Jaidka, K. (eds.): BIRNDL 2017: Proceedings of the 2nd Joint Workshop on Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries Co-located with the Joint Conference on Digital Libraries, vol. 1888. CEUR-WS, Aachen (2017)

11. Mayr, P., Chandrasekaran, M.K., Jaidka, K. (eds.): BIRNDL 2017: Proceedings of the 3rd Joint Workshop on Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries Co-located with the Joint Conference on Digital Libraries, vol. 2132. CEUR-WS, Aachen (2018)

12. Mayr, P., Frommholz, I., Cabanac, G. (eds.): BIR 2016 Proceedings of the 3rd Workshop on Bibliometric-Enhanced Information Retrieval Co-located with the 38th European Conference on Information Retrieval, vol. 1567. CEUR-WS, Aachen (2016)

13. Mayr, P., Frommholz, I., Cabanac, G. (eds.): BIR 2017 Proceedings of the 5th Workshop on Bibliometric-Enhanced Information Retrieval Co-located with the 39th European Conference on Information Retrieval, vol. 1823. CEUR-WS, Aachen (2017)

14. Mayr, P., Frommholz, I., Cabanac, G. (eds.): BIR 2018 Proceedings of the 7th Workshop on Bibliometric-Enhanced Information Retrieval Co-located with the 40th European Conference on Information Retrieval, vol. 2080. CEUR-WS (2018)

15. Mayr, P., Frommholz, I., Cabanac, G.: Report on the 7th International Workshop on Bibliometric-Enhanced Information Retrieval (BIR 2018). SIGIR Forum **52**(1), 135–139 (2018). https://doi.org/10.1145/3274784.3274798

16. Mayr, P., et al.: Special issue on bibliometric-enhanced information retrieval and natural language processing for digital libraries. Int. J. Digit. Libr. **19**(2–3), 107–111 (2018). https://doi.org/10.1007/s00799-017-0230-x

17. Mayr, P., Frommholz, I., Mutschke, P. (eds.): BIR 2015 Proceedings of the 2nd Workshop on Bibliometric-Enhanced Information Retrieval Co-located with the 37th European Conference on Information Retrieval, vol. 1344. CEUR-WS, Aachen (2015)

18. Mayr, P., Schaer, P., Scharnhorst, A., Larsen, B., Mutschke, P. (eds.): BIR 2016 Proceedings of the 1st Workshop on Bibliometric-Enhanced Information Retrieval Co-located with the 36th European Conference on Information Retrieval, vol. 1143. CEUR-WS, Aachen (2014)

19. Mayr, P., Scharnhorst, A.: Scientometrics and information retrieval: weak-links revitalized. Scientometrics **102**(3), 2193–2199 (2015). https://doi.org/10.1007/s11192-014-1484-3

20. Pritchard, A.: Statistical bibliography or bibliometrics? J. Doc. **25**(4), 348–349 (1969). https://doi.org/10.1108/eb026482. (Documentation notes)

21. Salton, G.: Associative document retrieval techniques using bibliographic information. J. ACM **10**(4), 440–457 (1963). https://doi.org/10.1145/321186.321188

22. Shotton, D.: Funders should mandate open citations. Nature **553**(7687), 129 (2018). https://doi.org/10.1038/d41586-018-00104-7

23. Sinha, A., et al.: An overview of Microsoft Academic Service (MAS) and applications. In: Gangemi, A., Leonardi, S., Panconesi, A. (eds.) WWW 2015: Proceedings of the 24th International Conference on World Wide Web, pp. 243–246. ACM, New York (2015). https://doi.org/10.1145/2740908.2742839
24. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: ArnetMiner: extraction and mining of academic social networks. In: KDD 2008: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 990–998. ACM, New York (2008). https://doi.org/10.1145/1401890.1402008
25. Van Noorden, R.: Google scholar pioneer on search engine's future. Nature (2014). https://doi.org/10.1038/nature.2014.16269
26. 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
27. Williams, K., Wu, J., Choudhury, S.R., Khabsa, M., Giles, C.L.: Scholarly big data information extraction and integration in the CiteSeer^x digital library. In: ICDE 2014: Proceedings of the 30th IEEE International Conference on Data Engineering Workshops, pp. 68–73. IEEE (2014). https://doi.org/10.1109/icdew.2014.6818305