

Report on the 8th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2019)

Guillaume Cabanac
University of Toulouse, France
guillaume.cabanac@univ-tlse3.fr

Ingo Frommholz
University of Bedfordshire, UK
ifrommholz@acm.org

Philipp Mayr
GESIS – Leibniz Institute for the Social Sciences, Germany
philipp.mayr@gesis.org

Abstract

The Bibliometric-enhanced Information Retrieval workshop series (BIR) at ECIR tackled 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, DocEar) and the industry (e.g., Google Scholar, Microsoft Academic Search, Semantic Scholar). This report presents the 8th iteration of the one-day BIR workshop held at ECIR 2019 in Cologne, Germany.

1 Introduction

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 [1]. The development of citation indexes pioneered by Garfield [2] proved determinant for such a research endeavour at the crossroads between the nascent fields of Information Retrieval (IR) and Bibliometrics¹. The pioneers who established these fields in Information Science—such as Salton and Garfield—were followed by scientists who specialised in one of these [5], 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’ [5] active in both academia and the industry together, who are developing search engines and recommender systems such as ArnetMiner [6], CiteSeer^x [7], DocEar [8], Google Scholar [9], Microsoft Academic Search [10], and Semantic Scholar [11], just to name a few.

¹Bibliometrics refers to the statistical analysis of the academic literature [3] and plays a key role in scientometrics: the quantitative study of science and innovation [4].

Bibliometric-enhanced IR systems must deal with the multifaceted nature of scientific information by searching for or recommending academic papers, patents [12], 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 [13] and the Google Dataset Search [14] launched on September 4, 2018, which give an impression of arising challenges subject to both communities. 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 [15] and it was held at ECIR each year since then [16, 17, 18, 19]. 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 [20] and SIGIR [21, 22] 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 [23], 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 [24, 25] and of the *International Journal on Digital Libraries* [26].

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. Thirteen peer-reviewed papers were accepted² [27]: 9 long papers, 3 short papers and 1 demo paper. These 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 contributed. The papers are as follows:

- Long papers:
 - An interactive visual tool for scientific literature search: Proposal and algorithmic specification [28]

²See workshop proceedings: <http://ceur-ws.org/Vol1-2345/>.

-
- A searchable space with routes for querying scientific information [29]
 - Discovering seminal works with marker papers [30]
 - How do computer scientists use Google Scholar?: A survey of user interest in elements on SERPs and author profile pages [31]
 - Feature selection and graph representation for an analysis of science fields evolution: An application to the digital library ISTEEX [32]
 - Optimal citation con-text window sizes for biomedical retrieval [33]
 - Bibliometric-enhanced arXiv: A data set for paper-based and citation-based tasks [34]
 - Mining intellectual influence associations [35]
 - Citation metrics for legal information retrieval systems [36]
- Short papers:
 - Finding temporal trends of scientific concepts [37]
 - A preliminary study to compare deep learning with rule-based approaches for citation classification [38]
 - Improving scientific article visibility by neural title simplification [39]
 - Demo:
 - Recommending multimedia educational resources on the MOVING platform [40].

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.

reports. All accepted papers were included in the workshop proceedings [27] hosted at ceur-ws.org, an established open access repository with no author-processing charges.

As a follow-up of the workshop, all authors are encouraged to submit an extended version of their papers to the Special Issue of the *Scientometrics* journal launched in Spring 2019.

References

- [1] Salton, G.: Associative document retrieval techniques using bibliographic information. *Journal of the ACM* **10**(4) (1963) 440–457
- [2] Garfield, E.: Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **122**(3159) (1955) 108–111
- [3] Pritchard, A.: Statistical bibliography or bibliometrics? [Documentation notes]. *Journal of Documentation* **25**(4) (1969) 348–349
- [4] Leydesdorff, L., Milojević, S.: Scientometrics. In Wright, J.D., ed.: *International Encyclopedia of the Social & Behavioral Sciences*. Volume 21. 2nd edn. Elsevier (2015) 322–327
- [5] White, H.D., McCain, K.W.: Visualizing a discipline: An author co-citation analysis of Information Science, 1972–1995. *Journal of the American Society for Information Science* **49**(4) (1998) 327–355
- [6] Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: ArnetMiner: Extraction and mining of academic social networks. In: *KDD'08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, ACM (2008) 990–998
- [7] 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'14: Proceedings of the 30th IEEE International Conference on Data Engineering Workshops*, IEEE (2014) 68–73
- [8] Beel, J., Langer, S., Gipp, B., Nürnberger, A.: The architecture and datasets of docear's research paper recommender system. *D-Lib Magazine* **20**(11/12) (2014)
- [9] Van Noorden, R.: Google Scholar pioneer on search engine's future. *Nature* (2014)
- [10] Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.J.P., Wang, K.: An overview of Microsoft Academic Service (MAS) and applications. In Gangemi, A., Leonardi, S., Panconesi, A., eds.: *WWW'15: Proceedings of the 24th International Conference on World Wide Web*, New York, NY, USA, ACM (2015) 243–246
- [11] Bohannon, J.: A computer program just ranked the most influential brain scientists of the modern era. *Science* (2016)
- [12] Garfield, E.: Patent citation indexing and the notions of novelty, similarity, and relevance. *Journal of Chemical Documentation* **6**(2) (1966) 63–65

-
- [13] Shotton, D.: Funders should mandate open citations. *Nature* **553**(7687) (2018) 129
- [14] Castelvechi, D.: Google unveils search engine for open data [News & Comment]. *Nature* (2018)
- [15] Mayr, P., Schaer, P., Scharnhorst, A., Larsen, B., Mutschke, P., eds.: BIR'16 Proceedings of the 1st Workshop on Bibliometric-enhanced Information Retrieval co-located with the 36th European Conference on Information Retrieval. Volume 1143., Aachen, CEUR-WS (2014)
- [16] Mayr, P., Frommholz, I., Mutschke, P., eds.: BIR'15 Proceedings of the 2nd Workshop on Bibliometric-enhanced Information Retrieval co-located with the 37th European Conference on Information Retrieval. Volume 1344., Aachen, CEUR-WS (2015)
- [17] Mayr, P., Frommholz, I., Cabanac, G., eds.: BIR'16 Proceedings of the 3rd Workshop on Bibliometric-enhanced Information Retrieval co-located with the 38th European Conference on Information Retrieval. Volume 1567., Aachen, CEUR-WS (2016)
- [18] Mayr, P., Frommholz, I., Cabanac, G., eds.: BIR'17 Proceedings of the 5th Workshop on Bibliometric-enhanced Information Retrieval co-located with the 39th European Conference on Information Retrieval. Volume 1823., Aachen, CEUR-WS (2017)
- [19] Mayr, P., Frommholz, I., Cabanac, G., eds.: BIR'18 Proceedings of the 7th Workshop on Bibliometric-enhanced Information Retrieval co-located with the 40th European Conference on Information Retrieval. Volume 2080., CEUR-WS (2018)
- [20] Cabanac, G., Chandrasekaran, M.K., Frommholz, I., Jaidka, K., Kan, M.Y., Mayr, P., Wolfram, D., eds.: BIRNDL'16: 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. Volume 1610., Aachen, CEUR-WS (2016)
- [21] Mayr, P., Chandrasekaran, M.K., Jaidka, K., eds.: BIRNDL'17: 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. Volume 1888., Aachen, CEUR-WS (2017)
- [22] Mayr, P., Chandrasekaran, M.K., Jaidka, K., eds.: BIRNDL'17: 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. Volume 2132., Aachen, CEUR-WS (2018)
- [23] Jaidka, K., Chandrasekaran, M.K., Rustagi, S., Kan, M.Y.: Insights from CL-SciSumm 2016: The faceted scientific document summarization shared task. *International Journal on Digital Libraries* **19**(2–3) (2018) 163–171
- [24] Mayr, P., Scharnhorst, A.: Scientometrics and information retrieval: weak-links revitalized. *Scientometrics* **102**(3) (2015) 2193–2199
- [25] Cabanac, G., Mayr, P., Frommholz, I.: Bibliometric-enhanced information retrieval: Preface. *Scientometrics* **116**(2) (2018) 1225–1227
-

-
- [26] Mayr, P., Frommholz, I., Cabanac, G., Chandrasekaran, M.K., Jaidka, K., Kan, M.Y., Wolfram, D.: Special issue on bibliometric-enhanced information retrieval and natural language processing for digital libraries. *International Journal on Digital Libraries* **19**(2–3) (2018) 107–111
- [27] Cabanac, G., Frommholz, I., Mayr, P., eds.: BIR’19 Proceedings of the 8th Workshop on Bibliometric-enhanced Information Retrieval co-located with the 41th European Conference on Information Retrieval. Volume 2345., Aachen, CEUR-WS (2019)
- [28] Bascur, J.P., van Eck, N.J., Waltman, L.: An interactive visual tool for scientific literature search: Proposal and algorithmic specification. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 76–87
- [29] Fabre, R.: A “searchable” space with routes for querying scientific information. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 112–124
- [30] Haunschild, R., Marx, W.: Discovering seminal works with marker papers. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 27–38
- [31] Kim, J., Trippas, J.R., Sanderson, M., Bao, Z., Croft, W.B.: How do computer scientists use Google Scholar?: A survey of user interest in elements on SERPs and author profile pages. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 64–75
- [32] Lamirel, J.C., Cuxac, P.: Feature selection and graph representation for an analysis of science fields evolution: An application to the digital library ISTEEX. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 88–99
- [33] Lykke Nielsen, B., Lavlund Skau, S., Meier, F., Larsen, B.: Optimal citation context window sizes for biomedical retrieval. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 51–63
- [34] Saier, T., Färber, M.: Bibliometric-enhanced arXiv: A data set for paper-based and citation-based tasks. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 14–26
- [35] Shah, T., Pudi, V.: Mining intellectual influence associations. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 100–111
- [36] Wiggers, G., Verberne, S.: Citation metrics for legal information retrieval systems. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 39–50
- [37] Färber, M., Jatowt, A.: Finding temporal trends of scientific concepts. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 132–139
-

-
- [38] Perier-Camby, J., Bertin, M., Atanassova, I., Armetta, F.: A preliminary study to compare deep learning with rule-based approaches for citation classification. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 125–131
- [39] Shvets, A.: Improving scientific article visibility by neural title simplification. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 140–147
- [40] Vagliano, I., Nazir, S.: Recommending multimedia educational resources on the MOVING platform. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 148–158
- [41] Atanassova, I.: Beyond metadata: the new challenges in mining scientific papers. In: Proc. of the 8th Workshop on Bibliometric-enhanced Information Retrieval, CEUR-WS.org (2019) 8–13