Report on the 10th Anniversary Workshop on Bibliometric-enhanced Information Retrieval (BIR 2020)

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) was launched at ECIR in 2014 [Mayr et al., 2014] and it was held at ECIR each year since then. This year we organized the 10th iteration of BIR as an all-virtual workshop with a peak of 97 participants. The workshop series at ECIR and JCDL/SIGIR tackles issues related to academic search, at the crossroads between Information Retrieval, Natural Language Processing and Bibliometrics. In this report, we summarize the past workshops, present the workshop topics for 2020 [Cabanac et al., 2020] and reflect on some future steps for this workshop series.

1 Introduction

Searching for scientific information is a long-lived user need. In the early 1960s, Salton was already striving to enhance information retrieval by including clues inferred from bibliographic citations [Salton, 1963]. The development of citation indexes pioneered by Garfield [1955] proved determinant for such a research endeavour at the crossroads between the nascent fields of Bibliometrics[1] and Information Retrieval (IR) — BIR. The pioneers who established these fields in Information Science — such as Salton and Garfield — were followed by scientists who specialised in one of these [White and McCain, 1998], 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 [Mayr and Scharnhorst, 2015]. We strive to get the 'retrievalists' and 'citationists' [White and McCain, 1998] active in both academia and the industry together, who

---

[1]Bibliometrics refers to the statistical analysis of the academic literature [Pritchard, 1969] and plays a key role in scientometrics: the quantitative analysis of science and innovation [Leydesdorff and Milojević, 2015].

are developing search engines and recommender systems such as ArnetMiner, Dimensions, Google Scholar, Microsoft Academic Search, and Semantic Scholar, just to name a few.

These 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 underlying models harness relevance signals from keywords provided by authors, topics extracted from the full-texts, co-authorship networks, citation networks, and various classifications schemes of science.

BIR is a hot topic with growing recognition in the community in recent years: see for instance the Initiative for Open Citations [Shotton, 2018], the Google Dataset Search [Brickley et al., 2019], the Indian JNU initiative for indexing the world's literature in full-text [Pulla, 2019], the increasing number of retractions [Brainard and You, 2018], and massive studies of self-citations [Ioannidis et al., 2019, Kacem et al., 2020]. We believe that BIR@ECIR is a much needed scientific event for the 'retrievalists', 'citationists' and others to meet and join forces pushing the knowledge boundaries of IR applied to literature search and recommendation.
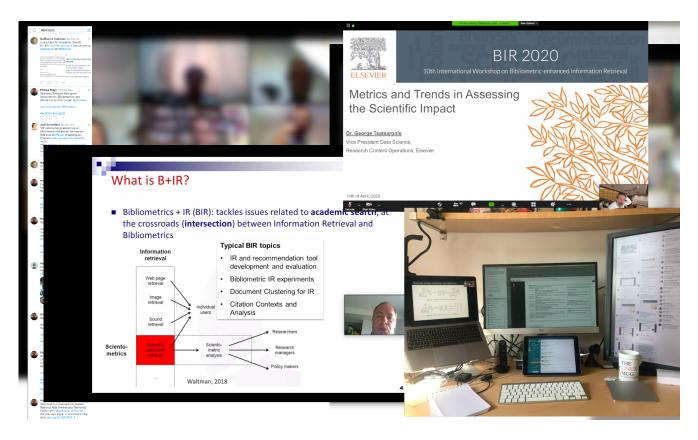

Figure 1: BIR 2020 online session – a collage as viewed from a typical conference location at home.

2 Online Experience

This year BIR faced an exceptional situation due to the COVID-19 lockdown and the fact that ECIR 2020 was an online only event (Fig. 1). The experience of the at peak times 97 participants was overall positive and provides some food for thought for the planning of similar future events and how a more inclusive online experience can be combined with the traditional face-to-face delivery.

3 Past BIR Workshop Activities

The BIR workshop series was launched at ECIR in 2014 [Mayr et al., 2014] and it was held at ECIR each year since then. As our workshop lies at the crossroads between IR and NLP, we also ran BIR as a joint workshop called BIRNDL (Bibliometric-enhanced IR and NLP for Digital Libraries) at the JCDL [Cabanac et al., 2016] and SIGIR [Chandrasekaran and Mayr, 2019] conferences. All past workshops had a large number of participants (between ∼30 and ∼60), demonstrating the relevance of the workshop's topics.

In the following, we present an overview of the past BIR workshops and keynotes at BIR (Tab. 1–2). All pointers to the workshops and proceedings are hosted on the BIR homepage[2]. Many of the presented workshop papers appeared in extended form in one of our four BIR-related special issues: 2015 [Mayr and Scharnhorst, 2015], 2018 [Cabanac et al., 2018, Mayr et al., 2018], and 2019 [Atanassova et al., 2019].

Table 1: Overview of the BIR workshop series

Year	Conference	Venue		Papers	Proceedings
		City	Country		
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[c]	PT	8	Vol-2591

[a] with CL-SciSumm 2016 Shared Task;

[b] with CL-SciSumm 2019 Shared Task;

[c] held online due to the COVID–19 pandemic

---

[2] https://sites.google.com/view/bir-ws

4 Workshop Topics

The programme of the BIR 2020 workshop is documented at the workshop's homepage[3]. The call for papers for the BIR 2020 workshop [Cabanac et al., 2020] 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:

   - Modelling the multifaceted nature of scientific information.
   - Building test collections for reproducible BIR.
   - System support for literature search and recommendation.

4.1 Keynote

This year's keynote was given by George Tsatsaronis: *Metrics and Trends in Assessing the Scientific Impact.*

4.2 Research Papers

The following research papers were presented in 3 sessions.

▷ Session: Expert Finding and Ranking Models

- Robin Brochier, Antoine Gourru, Adrien Guille and Julien Velcin:
  New Datasets and a Benchmark of Document Network Embedding Methods for Scientific Expert Finding

- Christopher Michels, Mandy Neumann, Philipp Schaer and Ralf Schenkel:
  Conference Indexing in Digital Libraries: A Ranking Model and Case Study on dblp

---

[3]https://sites.google.com/view/bir-ws/bir-2020

▷ Session: Citations, Citations, Citations

- Gineke Wiggers and Suzan Verberne:
  Usage and Citation Metrics for Ranking Algorithms in Legal Information Retrieval Systems

- Juan Pablo Bascur, Suzan Verberne, Nees Jan Van Eck and Ludo Waltman:
  Browsing Citation Clusters for Academic Literature Search: A Simulation Study with Systematic Reviews

- Michael Färber, Timo Klein and Joan Sigloch:
  Neural Citation Recommendation: A Reproducibility Study

▷ Session: Learning to Rank and Evaluation

- Daniel Kershaw, Benjamin Pettit, Maya Hristakeva and Kris Jack:
  Learning to Rank Research Articles: A Case Study of Collaborative Filtering and Learning to Rank in ScienceDirect

- Rodrigo Nogueira, Zhiying Jiang, Kyunghyun Cho and Jimmy Lin:
  Evaluating Pretrained Transformer Models for Citation Recommendation

- Timo Breuer, Philipp Schaer and Dirk Tunger:
  Relations Between Relevance Assessments, Bibliometrics and Altmetrics

4.3 Greeting Notes to the 10th Anniversary BIR Edition

As BIR 2020 was our anniversary edition, we asked renowned researchers from the Scientometrics and NLP as well as Information Retrieval community beforehand to send their thoughts about the workshop series and reflect on BIR's mission. The following videos[4] and our written greeting notes were submitted and included in the BIR proceedings.

- Andrea Scharnhorst (DANS, The Netherlands):
  Building bridges
  Companion video https://youtu.be/kNPVZZ7Mq0M

- Dietmar Wolfram (University of Wisconsin-Milwaukee, USA)
  Video https://youtu.be/BKNDYq09_-M

- Suzan Verberne (Leiden University, The Netherlands)
  Video https://youtu.be/eM1Kwevrdkc

- Mike Thelwall (University of Wolverhampton, UK):
  Why we need another ten years of Bibliometric-enhanced Information Retrieval
  Companion video https://youtu.be/Ld1s6mEpA2Q

---

[4]Please see the full playlist at https://bit.ly/BIR-greetings

Table 2: Keynotes at BIR

Year	Area[a]	Title of the keynote presentation	Presenter
2015	SCIM	In Praise of Interdisciplinary Research through Scientometrics	Cabanac, G.
2016	IR	Bibliometrics in Online Book Discussions: Lessons for Complex Search Tasks	Koolen, M.
2016	SCIM	Bibliometrics, Information Retrieval and Natural Language Processing: Natural Synergies to Support Digital Library Research	Wolfram, D.
2017	IR	Real-World Recommender Systems for Academia: The Pain and Gain in Building, Operating, and Researching them	Beel, J.
2017	NLP	Do "Future Work" sections have a purpose? Citation links and entailment for global scientometric questions	Teufel, S.
2018	NLP	Trends in Gaming Indicators: On Failed Attempts at Deception and their Computerised Detection	Labbé, C.
2018	IR	Integrating and Exploiting Public Metadata Sources in a Bibliographic Information System	Schenkel, R.
2019	NLP	Beyond Metadata: the New Challenges in Mining Scientific Papers	Atanassova, I.
2019	IR	Personalized Feed/Query-formulation, Predictive Impact, and Ranking	Wade, A.
2019	NLP	Discourse Processing for Text Analysis: Recent Successes, Current Challenges	Webber, B.
2020	NLP	Metrics and Trends in Assessing the Scientific Impact	Tsatsaronis, G.

[a] SCIM: Scientometrics; NLP: Natural Language Processing; IR: Information Retrieval

- Iana Atanassova (Université Bourgogne Franche-Comté, France) and
  Marc Bertin (University of Lyon, France):
  BIR: A time and a place to envision concepts and tools around Bibliometrics
  Companion video https://youtu.be/hLgTB0av_b0

- Henry Small (SciTech Strategies, USA):
  Some questions for information science arising from the history and philosophy of science
  Companion video https://youtu.be/xOpFB0r0WPg

- Staša Milojević (Indiana University, USA)
  Video https://youtu.be/N6WRjudG2G4

- Min-Yen Kan (National University of Singapore, Singapore)
  Video https://youtu.be/OpKXeUK3CJo

- Ludo Waltman (Leiden University, The Netherlands)
  Video https://youtu.be/YHxzppJP5Js

- Aparna Basu (South Asian University, India):
  Remembering Don Swanson: Link to Bibliometric-enhanced Information Retrieval
  Companion video https://youtu.be/tEPyL-x-R1o

- Wolfgang Glänzel (KU Leuven, Belgium):
  Bibliometrics-aided retrieval: A success story
  Companion video https://youtu.be/GkN4ngT1RIs

- Akiko Aizawa (National Institute of Informatics, Japan)
  Video https://youtu.be/Tm9sJZ8X470

- Muthu Kumar Chandrasekaran (Amazon, USA)
  BIR and BIRNDL: A marathon towards a congregation of scientific document processing community
  Companion video https://youtu.be/V1zxs_3b_CI

- Michel Zitt (INRA Nantes, France):
  Intrication between Information Retrieval and Bibliometrics: the case of scientific domain delineation

- Howard D. White (Drexel University, USA):
  Anniversary Statement

5 Further Readings and Next Steps

Since 2016 we maintain the "Bibliometric-enhanced-IR Bibliography"[5] that collects scientific papers which appear in collaboration with the BIR/BIRNDL events.

Currently the BIR organizers edit a Special issue on "Bibliometrics and Information Retrieval"[6] in the journal *Scientometrics* (Springer). All accepted and published papers are listed on the SI page accordingly.

Research on scholarly document processing has for many years been scattered across multiple venues like ACL, SIGIR, JCDL, CIKM, EMNLP, LREC, NAACL, KDD, and others. Our next strategic step is the First Workshop on Scholarly Document Processing (SDP)[7] will be held in November 2020 in conjunction with the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 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. The SDP workshop will be also a all-virtual event.

The BIR community is invited to submit research papers to a new Research Topic "Mining Scientific Papers Volume II: Knowledge Discovery and Data Exploitation" in the journal *Frontiers in Research Metrics and Analytics*[8] (Deadline for submissions is September 1, 2020).

---

[5] https://github.com/PhilippMayr/Bibliometric-enhanced-IR_Bibliography/
[6] https://sites.google.com/view/scientometrics-si2019-bir
[7] https://ornlcda.github.io/SDProc/
[8] https://bit.ly/Frontiers2020

References

Philipp Mayr, Philipp Schaer, Andrea Scharnhorst, Birger Larsen, and Peter Mutschke, editors. *BIR'14 Proceedings of the 1st Workshop on Bibliometric-enhanced Information Retrieval co-located with the 36th European Conference on Information Retrieval*, volume 1143, Aachen, 2014. CEUR-WS. URL http://ceur-ws.org/Vol-1143.

Guillaume Cabanac, Ingo Frommholz, and Philipp Mayr, editors. *Proceedings of the 10th International Workshop on Bibliometric-enhanced Information Retrieval co-located with 42nd European Conference on Information Retrieval, BIR@ECIR 2020, Lisbon, Portugal, April 14th, 2020 [online only]*, volume 2591 of *CEUR Workshop Proceedings*, 2020. CEUR-WS.org. URL http://ceur-ws.org/Vol-2591.

Gerard Salton. Associative document retrieval techniques using bibliographic information. *Journal of the ACM*, 10(4):440–457, 1963. doi:10.1145/321186.321188.

Eugene Garfield. Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 122(3159):108–111, 1955. doi:10.1126/science.122.3159.108.

Alan Pritchard. Statistical bibliography or bibliometrics? [Documentation notes]. *Journal of Documentation*, 25(4):348–349, 1969. doi:10.1108/eb026482.

Loet Leydesdorff and Staša Milojević. Scientometrics. In James D. Wright, editor, *International Encyclopedia of the Social & Behavioral Sciences*, volume 21, pages 322–327. Elsevier, 2nd edition, 2015. doi:10.1016/b978-0-08-097086-8.85030-8.

Howard D. White and Katherine W. McCain. Visualizing a discipline: An author co-citation analysis of Information Science, 1972–1995. *Journal of the American Society for Information Science*, 49(4):327–355, 1998. doi:b57vc7.

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

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

Dan Brickley, Matthew Burgess, and Natasha Noy. Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. In *The World Wide Web Conference on - WWW '19*, pages 1365–1375. ACM Press, 2019. ISBN 978-1-4503-6674-8. doi:10.1145/3308558.3313685.

Priyanka Pulla. The plan to mine the world's research papers. *Nature*, 571:316–318, 2019. doi:10.1038/d41586-019-02142-1.

Jeffrey Brainard and Jia You. What a massive database of retracted papers reveals about science publishing's "death penalty". *Science*, 2018. doi:10.1126/science.aav8384.

John P. A. Ioannidis, Jeroen Baas, Richard Klavans, and Kevin W. Boyack. A standardized citation metrics author database annotated for scientific field. *PLOS Biology*, 17(8):e3000384, August 2019. ISSN 1545-7885. doi:10.1371/journal.pbio.3000384.

Ameni Kacem, Justin W. Flatt, and Philipp Mayr. Tracking self-citations in academic publishing. *Scientometrics*, 2020. doi:10.1007/s11192-020-03413-9.

Guillaume Cabanac, Muthu Kumar Chandrasekaran, Ingo Frommholz, Kokil Jaidka, Min-Yen Kan, Philipp Mayr, and Dietmar Wolfram, editors. *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, 2016. CEUR-WS. URL http://ceur-ws.org/Vol-1610.

Muthu Kumar Chandrasekaran and Philipp Mayr, editors. *BIRNDL'19: Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries co-located with the Joint Conference on Digital Libraries*, volume 2414, Aachen, 2019. CEUR-WS. URL http://ceur-ws.org/Vol-2414.

Guillaume Cabanac, Philipp Mayr, and Ingo Frommholz. Bibliometric-enhanced information retrieval: Preface. *Scientometrics*, 116(2):1225–1227, 2018. doi:10.1007/s11192-018-2861-0.

Philipp Mayr, Ingo Frommholz, Guillaume Cabanac, Muthu Kumar Chandrasekaran, Kokil Jaidka, Min-Yen Kan, and Dietmar Wolfram. Special issue on bibliometric-enhanced information retrieval and natural language processing for digital libraries. *International Journal on Digital Libraries*, 19(2–3):107–111, 2018. doi:10.1007/s00799-017-0230-x.

Iana Atanassova, Marc Bertin, and Philipp Mayr. Editorial: Mining Scientific Papers: NLP-enhanced Bibliometrics. *Frontiers in Research Metrics and Analytics*, 2019. ISSN 2504-0537. doi:10.3389/frma.2019.00002.