Preface to the 11th Workshop on Bibliometric-enhanced Information Retrieval at ECIR 2021

Ingo Frommholz[a], Philipp Mayr[b,c], Guillaume Cabanac[d] and Suzan Verberne[e]

[a]School of Mathematics and Computer Science, University of Wolverhampton, UK
[b]GESIS – Leibniz-Institute for the Social Sciences, Cologne, Germany,
[c]University of Göttingen, Institute of Computer Science, Göttingen, Germany
[d]University of Toulouse, Computer Science Department, IRIT UMR 5505, France
[e]Leiden Institute of Advanced Computer Science, Leiden University, the Netherlands

Abstract

This preface summarizes the 11th Workshop on Bibliometric-enhanced Information Retrieval (BIR). BIR was held as virtual event at April 1st, 2021, co-located with the 43rd European Conference on Information Retrieval (ECIR 2021).

1. Introduction

The aim of the Bibliometric-enhanced Information Retrieval workshop series (BIR) is to bring together researchers from different communities, especially scientometrics/bibliometrics and information retrieval. In doing so, BIR has a long-established tradition. It was launched at ECIR in 2014 [1] and was held at ECIR each year since then. As the topic of our workshop lies at the intersection 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 and SIGIR conferences, respectively. This year marked the 11th iteration of BIR, as virtual event only for the 2nd time due to the pandemic situation.

All pointers to past and future workshops as well as to proceedings are hosted at https://sites.google.com/view/bir-ws/.

BIR 2021: 11th International Workshop on Bibliometric-enhanced Information Retrieval at ECIR 2021, April 1, 2021, online
ifrommholz@acm.org (I. Frommholz); philipp.mayr@gesis.org (P. Mayr); guillaume.cabanac@univ-tlse3.fr (G. Cabanac); s.verberne@liacs.leidenuniv.nl (S. Verberne)
https://philippmayr.github.io/ (P. Mayr)
0000-0002-5622-5132 (I. Frommholz); 0000-0002-6656-1658 (P. Mayr); 0000-0003-3060-6241 (G. Cabanac); 0000-0002-9609-9505 (S. Verberne)

2. Overview of the papers

This year five submissions were accepted as full papers and four as short papers. Both long and short papers have been scheduled for presentation during the workshop and are included in the CEUR-WS proceedings. In addition, the workshop featured three keynote talks. All workshop contributions are documented in the workshop website.[1] The following section briefly lists the various contributions.

2.1. Keynotes

We had three keynote speakers this year.

Lucy Lu Wang (Allen Institute for AI, USA) Text mining insights from the COVID-19 pandemic. Over the last year, the novel coronavirus SARS-CoV-2 generated significant upheaval in the world and in scientific publishing, yet it provided a unique sandbox in which to test and innovate upon the latest text mining and information retrieval technologies. We saw the emergence of novel information retrieval and NLP tasks with the potential to change the way information from scientific literature is communicated to healthcare providers and public health researchers. In this talk, I will discuss some of the ways the computing community came together to tackle this challenge, with the release of open data resources like CORD-19 and the introduction of various shared tasks for evaluation. I will also present our work on scientific fact checking, a novel NLP task that looks to address issues around scientific misinformation, and its practical uses in managing conflicting information arising from COVID-19 pandemic publishing.

Ludo Waltman (CWTS, the Netherlands): Openness, transparency, and inclusivity in science: What does it mean for information retrieval? The research system is moving in a direction of increased openness, transparency, and inclusivity. This offers exciting new possibilities for scholarly literature search. At the same time it also changes the expectations users have of search systems for scholarly literature. New opportunities are provided by the increased openness of the metadata of scientific outputs, and sometimes also of the outputs themselves. Calls for increased transparency and inclusivity raise complex questions about the responsibilities of those who manage search systems for scholarly literature and about the benefits as well as the risks of new AI-based approaches to scholarly literature search. While acknowledging that there are no easy answers, I will share some of my thoughts on the various issues that the BIR community may need to reflect on.

Jimmy Lin (University of Waterloo, Canada): Domain Adaptation to Scientific Texts and the Limits of Scale? A fundamental assumption behind bibliometric-enhanced information retrieval is that ranking models need to be adapted to handle scientific text, which are very different from the typical corpora (Wikipedia, books, web crawls, etc.) used to pretrain large-scale transformers. One common approach is to take a large ”general-domain” model and

[1] https://sites.google.com/view/bir-ws/bir-2021

then apply domain adaptation techniques to "customize" it for a specific (scientific) domain. While we have pursued this research direction, it appears that the far less satisfying approach of "just throwing more data at the problem" with increasingly larger pretrained transformers seems to be more effective. In fact, over the last year, my group has "won" multiple community-wide shared evaluations focused on texts related to the novel coronavirus SARS-CoV-2 using exactly this approach: document ranking (TREC-COVID, TREC Health Misinformation), question answering (EPIC-QA), and fact verification (SciFcat). We have been deeply frustrated by our own inability for "smarter" to beat "bigger". In this talk, I will share our efforts to grapple with these issues and perhaps to understand... why?

2.2. Research papers

The following research papers were presented in 3 sessions.

▷ Session 1

- Shintaro Yamamoto, Anne Lauscher, Simone Paolo Ponzetto, Goran Glavaš and Shigeo Morishima:
  Self-Supervised Learning for Visual Summary Identification in Scientific Publications (long paper)
- Pablo Accuosto, Mariana Neves and Horacio Saggion:
  Argumentation mining in scientific literature: From computational linguistics to biomedicine (long paper)
- Frederique Bordignon, Liana Ermakova and Marianne Noel:
  Preprint abstracts in times of crisis: a comparative study with the pre-pandemic period (short paper)

▷ Session 2

- Hiran H. Lathabai, Abhirup Nandy and Vivek Kumar Singh:
  Expertise based institutional recommendation in different thematic areas (short paper)
- Ahmed Abura'Ed and Horacio Saggion:
  A select and rewrite approach to the generation of related work reports (long paper)
- Jacqueline Sachse:
  Bibliometric Indicators and Relevance Criteria – An Online Experiment (short paper)

▷ Session 3

- Ken Voskuil and Suzan Verberne:
  Improving reference mining in patents with BERT (long paper)
- Manajit Chakraborty, David Zimmermann and Fabio Crestani:
  PatentQuest: A User-Oriented Tool for Integrated Patent Search (long paper)
- Daria Alexander and Arjen P. de Vries:
  "This research is funded by...": Named Entity Recognition of financial information in research papers (short paper)

3. Further reading

In 2020, the BIR organizers have edited a Special issue on "Scholarly literature mining with Information Retrieval and Natural Language Processing"[2] in the journal *Scientometrics* (Springer). In total, fourteen papers on all aspects of academic search were accepted, see an overview [2].

Since 2016 we maintain the "Bibliometric-enhanced-IR Bibliography"[3] that collects scientific papers which appeared in collaboration with the BIR/BIRNDL organizers.

Acknowledgments

The organizers wish to thank all those who contributed to this workshop series: the researchers who contributed papers, the many reviewers who generously offered their time and expertise, and the participants of the BIR and BIRNDL workshops.

References

[1] P. Mayr, A. Scharnhorst, B. Larsen, P. Schaer, P. Mutschke, Bibliometric-Enhanced Information Retrieval, in: 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings, Springer International Publishing, 2014, pp. 798–801. URL: http://arxiv.org/abs/1310.8226. doi:10.1007/978-3-319-06028-6_99.

[2] G. Cabanac, I. Frommholz, P. Mayr, Scholarly literature mining with Information Retrieval and Natural Language Processing: Preface, Scientometrics 125 (2020) 2835–2840. doi:10.1007/s11192-020-03763-4.

[2] https://sites.google.com/view/scientometrics-si2019-bir

[3] https://github.com/PhilippMayr/Bibliometric-enhanced-IR_Bibliography/