



B!SON: A Tool for Open Access Journal Recommendation

Elias Entrup¹, Anita Eppelin¹, Ralph Ewerth^{1,3}, Josephine Hartwig²,
Marco Tullney¹, Michael Wohlgemuth², and Anett Hoppe^{1,3}✉

¹ TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany
{bison,anett.hoppe}@tib.eu

² Saxon State and University Library, Dresden, Germany

³ L3S Research Center, Leibniz University Hannover, Hannover, Germany

Abstract. Finding a suitable open access journal to publish scientific work is a complex task: Researchers have to navigate a constantly growing number of journals, institutional agreements with publishers, funders' conditions and the risk of Predatory Publishers. To help with these challenges, we introduce a web-based journal recommendation system called B!SON. It is developed based on a systematic requirements analysis, built on open data, gives publisher-independent recommendations and works across domains. It suggests open access journals based on title, abstract and references provided by the user. The recommendation quality has been evaluated using a large test set of 10,000 articles. Development by two German scientific libraries ensures the longevity of the project.

Keywords: Recommendation system · Open access · Paper submission · Publishing

1 Introduction

The open access landscape keeps growing, making it harder and harder to choose a suitable journal to publish research findings. There is an increasing number of journals: The Directory of Open Access Journals (DOAJ), an online directory of peer-reviewed open access journals, added over 5,000 journals in the last three years. All of these journals offer a variety of publishing conditions, including different peer-review schemes, publication costs and waivers, copyright and rights retention clauses. But there is also growing support for researchers to facilitate open access publishing: The recent years brought on various agreements formed between academic institutions and publishers, which determine publication costs and conditions; scientific libraries increasingly offer support services. Also, more funding agencies expect scientists to use open access options and specify clear conditions with respect to how funded work is to be published. This adds to the overall workload of researchers: They have to assess the open access landscape while taking into account factors such as newly established journals, predatory publishing schemes, quality measures, and, finally, individual publication costs.

© The Author(s) 2022

G. Silvello et al. (Eds.): TPD 2022, LNCS 13541, pp. 357–364, 2022.

https://doi.org/10.1007/978-3-031-16802-4_33

In this paper, we present B!SON, a web-based recommendation system which aims to alleviate these problems. Open access journals are recommended based on content similarity (using title, abstract and keywords) and the cited references. User requirements were systematically collected in a survey [9], focusing primarily on researchers, but also addressing libraries, publishers and editors of scholar-led journals. Findings from these surveys have been discussed in depth with a user community and were incorporated into the system specification accordingly. The quality of B!SON's recommendation is evaluated on a large test set of 10,000 articles. The rest of the paper describes the B!SON prototype by first reviewing existing work on scientific recommendation (Sect. 2), then describing the B!SON service, its data sources, algorithm and functionality (Sect. 3), and concluding in Sect. 4.

2 Related Work

Scientific recommendation tasks span the search for potential collaborators [17] and reviewers [18], of papers to read [1] and to cite [2, 5]. With more and more ways of publishing scientific articles, the recommendation of scientific publication outlets (journals/conferences) is a task which is on the rise (e.g. [22, 25]).

Prototypical approaches explore diverse data sources to provide recommendations. A major source of information is the article to be published: The text's title, abstract or keywords are used to compare against papers that previously appeared in an outlet [14, 25]. Other systems exploit the literature cited by the article, and try to determine the best publication venue using bibliometric metrics [7, 20]. An alternative stream of research focuses more on the article authors, exploring their publication history [24] and co-publication networks [16, 23].

While there is a number of active journal recommender sites, they all come with limitations. Several publishers offer services limited to their own journals like Elsevier's Journalfinder¹ [10] or Springer's Journal suggester². Others, like Journal Guide³ are closed-source and do not provide transparent information on their recommendation approach. Several collect user and usage data, e.g. Web of Science's manuscript matcher⁴. Notably, some open recommenders exist, e.g. Open Journal Matcher⁵, Pubmender⁶ [4] and Jane⁷ [21], the latter two being limited to medical journals. All of these open services provide little information on journals, are limited to abstract input and do not offer advanced filter options.

¹ <https://journalfinder.elsevier.com/>.

² <https://journalsuggester.springer.com/>.

³ <https://www.journalguide.com/>.

⁴ <https://mjl.clarivate.com/manuscript-matcher>.

⁵ <https://ojm.ocert.at/>.

⁶ <https://www.keaml.cn/abc/>.

⁷ <https://jane.biosemantics.org/>.

3 B!SON – The Open Access Journal Recommender

B!SON is the abbreviation for *Bibliometric and Semantic Open Access Recommender Network*. It combines several available data sources (see Sect. 3.1 for details) to provide authors and publication support services at libraries with recommendations of suitable open access journals, based on the entered title, abstract and reference list of the paper to be published.

3.1 Data Sources

The B!SON service is built on top of several open data sources with strong reputation in the open access community:

DOAJ: The Directory of Open Access Journals (DOAJ)⁸ collects information on open access journals which fulfill a set of quality criteria (available full text, dedicated article URLs, at least one ISSN, etc.). The dataset includes basic information on the journal itself, but also metadata of the published articles (title, abstract, year, DOI, ISSN of journal, etc.). The DOAJ currently contains 17,669 journals and 7,489,975 articles. The data is available for download in JSON format under CC0 for articles and CC BY-SA for journal data [3].

OpenCitations: The OpenCitations initiative⁹ collects (amongst other) the CC0-licensed COCI data set for citation data. It is based on Crossref data and contains over 72,268,850 publications and 1,294,283,603 citations [15]. The information is available in the form of DOI-to-DOI relations and it covers 44% of citations in Scopus and 51% of the citations in Dimensions [13]. COCI lacks citations in comparison to commercial products, but can be used to check which articles published in DOAJ journals cite the references given by the user (details in Sect. 3.2). As open access journals are incentivized to submit their articles' metadata, we can assume that the coverage of COCI in this regard is better.

Journal Checker Tool: The cOAlition S initiative (a group of funding agencies that agreed on a set of principles for the transition to open access) provides the Journal Checker Tool¹⁰. A user can enter journal ISSN, funder and institution to check whether a journal is open access according to Plan S, if the journal is a transformative journal or has a transformative agreement with the user's institution, or whether there is a self-archiving option [8]. An API allows to fetch this information automatically. Since B!SON does not retrieve data on funder or institution, we use the funder information of the European Commission to check if a journal is Plan-S compliant.

Other data sources: There are many other projects whose data might be used in B!SON's future to supplement the currently used data sets. Crossref would allow us to extend the article data of the DOAJ which are occasionally incomplete. OpenAlex (by OurResearch) could add e.g. author information.

⁸ <https://doaj.org>.

⁹ <https://opencitations.net>.

¹⁰ <https://journalcheckertool.org/>.

Table 1. Top@N accuracy for the different search methods.

Method	top@1	top@5	top@10	top@15
Bibliometric search	0.2048	0.3789	0.4185	0.4342
Title search	0.2102	0.4481	0.5489	0.5989
Abstract search	0.2501	0.5148	0.6160	0.6621
Combined search	0.2770	0.5644	0.6716	0.7229

3.2 Technology

B!SON consists of a Django backend¹¹ and a Vue.js¹² frontend.

Data Integration: PostgreSQL and Elasticsearch are used as databases. Data from DOAJ and OpenCitations’ COCI index are bulk downloaded and inserted into PostgreSQL and Elasticsearch. The information on Plan-S compliance stems from Journal Checker Tool and is fetched from their API using “European Commission Horizon Europe Framework Programme” as the funder.

All developed software will be published open source in the upcoming weeks¹³.

Recommendation: The recommendation is based on similarity measures with regard to the entered text data (title, abstract and keywords) and reference list.

Text similarity: Elasticsearch has a built-in functionality for text similarity search based on the Okapi BM 25 algorithm [19]. This is used to determine those articles already indexed in the DOAJ which are similar to the entered information. Stop word removal is performed as a pre-processing step. As the DOAJ contains articles in several languages, we combine the stop word lists from Apache Lucene for this purpose.

Bibliographic coupling: Additionally to textual data, the user can enter the list of cited articles, allowing to match journals based on bibliometric coupling [11]. For this, we extract the DOIs from the input list using regular expressions, then rely on the OpenCitations’ COCI index to find existing articles citing the same sources. The current solution is in a prototypical state: The number of matching citations is divided by the highest number of matching citations of the compared articles; if the normalized value is higher than a threshold (which is currently manually defined), the article is considered similar. We are currently working on integrating more sophisticated normalisation methods (e.g. [12]) and exploring options on how to dynamically define the threshold value.

Combination of text-based and bibliographic similarity: Similar articles are matched with their journal and the total score is calculated as a sum. Refined aggregation methods are currently explored and will be available soon.

¹¹ <https://www.djangoproject.com/>.

¹² <https://vuejs.org/>.

¹³ <https://gitlab.com/TIBHannover/bison/>.

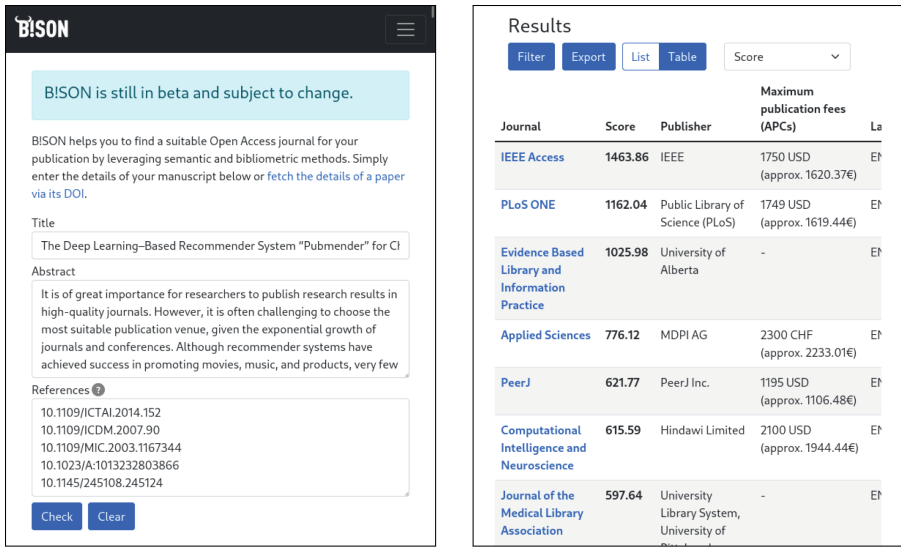


Fig. 1. Screenshot of B!SON prototype with an exemplary query.

Recommendation Evaluation: The algorithm is evaluated on a separate test data set of 10,000 DOAJ articles. To ensure realistic input data, all articles in the test set have a minimal abstract length of 100 characters and a minimal title length of 30 characters. As the references are not part of the DOAJ data, the COCI index was used to fetch references via the article DOI. Only articles with at least one reference were included. We assume that the articles were published in a suitable journal to begin with, counting a positive result if the originating journal appears in the top-n results of the recommendation. While this may not be correct for each individual article, we rely on the assumption that the overall journal scope is defined by the articles published in a journal. This current recommendation algorithm reaches the top@N accuracy shown in Table 1 when tested on a test set of 10,000 DOAJ articles.

3.3 User Interface and Functionality

The current state of the B!SON prototype is available for testing¹⁴. The user interface has been designed deliberately simple, a screenshot is shown in Fig. 1.

Data entry: The start page directly allows the user to enter title, abstract and references or fill them out automatically by fetching the information from Crossref with a DOI so that open access publication venues can be found based on previously published research.

Result page: To inspect the search results, the user has the choice of representing them as a simple list or a table which offers a structured account of

¹⁴ <https://service.tib.eu/bison/>.

additional details, enabling easy comparison of the journals. Author publishing costs (APCs) are displayed based on the information available in DOAJ, and automatically converted to Euro if necessary.

Currently, the displayed similarity score is calculated based on simple addition of the Elasticsearch similarity score and the bibliometric similarity score. By clicking on the score field, the user has the option to display a pop-over with explanatory information: B!SON will then display the list of articles which previously appeared in said journal which were determined to be similar by the recommendation engine. Clicking on a journal name leads the users to a separate detail page which offers even more information including keywords, publishing charges, license, Plan-S compliance, and more.

Data export and transparency: Search results can be exported as CSV for further use and analysis. A public API is available for programmatic access. It is also planned to provide the recommendation functionality in a form that is easily integrated and adapted to local library systems. For transparency on data sources, the date of the last update of the data is shown on the “About Page”.

4 Conclusion

In this paper, we have presented a novel prototypical recommendation system for open access journals. The system combines semantic and bibliometric information to calculate a similarity score to the journals’ existing contents, and provides the user with a ranked list of candidate venues.

The B!SON prototype is available online for beta-testing. Based on the community feedback received so far, we are currently working on further optimisations. This concerns, for instance, the computation of the similarity score which is, to date, a simple addition of semantic and bibliometric similarity results. More sophisticated aggregation methods will allow optimised weighting of both components and, perspective, better interpretability of the resulting score. Furthermore, we are currently exploring embedding-based text (used by e.g. Pubmender [4]) and citation graph representations [6] to further improve the recommendation results. Several remarks also concern the user interface. The community’s wishlist includes more sophisticated methods for the exploration of the result list, e.g. graph-based visualisations, an extension of the filtering options and an improved representation of the similarity score.

Going beyond the scope of the B!SON project, it could be interesting to extend recommendations to other venues which offer open access publication such as conferences. Moreover, the integration of person-centred information, such as prior publication history and frequent co-authors seems promising.

Acknowledgements. B!SON is funded by the German Federal Ministry of Education and Research (BMBF) for a period of two years (funding reference: 16TOA034A) and partners with DOAJ and OpenCitations. The publication was funded by the Open Access Fund of the TIB - Leibniz Information Centre for Science and Technology.

References

1. Beel, J., Gipp, B., Langer, S., Breitingner, C.: Research-paper recommender systems: a literature survey. *Int. J. Digit. Libr.* **17**(4), 305–338 (2015). <https://doi.org/10.1007/s00799-015-0156-0>
2. Brack, A., Hoppe, A., Ewerth, R.: Citation recommendation for research papers via knowledge graphs. In: Berget, G., Hall, M.M., Brenn, D., Kumpulainen, S. (eds.) *TPDL 2021*. LNCS, vol. 12866, pp. 165–174. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86324-1_20
3. Directory of Open Access Journals. <https://doaj.org/>. Accessed 25 May 2022
4. Feng, X., et al.: The deep learning-based recommender system “Pubmender” for choosing a biomedical publication venue: development and validation study. *J. Med. Internet Res.* **21**(5), e12957 (2019). <https://doi.org/10.2196/12957>
5. Färber, M., Sampath, A.: HybridCite: a hybrid model for context-aware citation recommendation. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, pp. 117–126. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3383583.3398534>
6. Ganguly, S., Pudi, V.: Paper2vec: combining graph and text information for scientific paper representation. In: Jose, J.M., et al. (eds.) *ECIR 2017*. LNCS, vol. 10193, pp. 383–395. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56608-5_30
7. Ghosal, T., Chakraborty, A., Sonam, R., Ekbal, A., Saha, S., Bhattacharyya, P.: Incorporating full text and bibliographic features to improve scholarly journal recommendation. In: 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL), pp. 374–375. IEEE, Champaign, IL, USA (2019). <https://doi.org/10.1109/JCDL.2019.00077>
8. Hannah Hope: Unboxing the Journal Checker Tool | Plan S. <https://www.coalition-s.org/blog/unboxing-the-journal-checker-tool/>. Accessed 01 June 2022
9. Hartwig, J., Eppelin, A.: Which journal characteristics are crucial for scientists when selecting journals for their publications? Results tables of an online survey (2021). <https://doi.org/10.5281/zenodo.5728148>. Type: dataset
10. Kang, N., Doornenbal, M.A., Schijvenaars, R.J.: Elsevier journal finder: recommending journals for your paper. In: Proceedings of the 9th ACM Conference on Recommender Systems, pp. 261–264. RecSys 2015, Association for Computing Machinery, New York, NY, USA (2015). <https://doi.org/10.1145/2792838.2799663>
11. Kessler, M.M.: Bibliographic coupling between scientific papers. *Am. Doc.* **14**(1), 10–25 (1963). <https://doi.org/10.1002/asi.5090140103>
12. Leydesdorff, L.: On the normalization and visualization of author co-citation data: Salton’s cosine versus the Jaccard index. *J. Assoc. Inf. Sci. Technol.* **59**(1), 77–85 (2008). <https://doi.org/10.1002/asi.20732>
13. Martín-Martín, A., Thelwall, M., Orduna-Malea, E., Delgado López-Cózar, E.: Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics* **126**(1), 871–906 (2021). <https://doi.org/10.1007/s11192-020-03690-4>
14. Nguyen, D., Huynh, S., Huynh, P., Dinh, C.V., Nguyen, B.T.: S2CFT: a new approach for paper submission recommendation. In: Bureš, T., et al. (eds.) *SOFSEM 2021*. LNCS, vol. 12607, pp. 563–573. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67731-2_41

15. OpenCitations: COCI CSV dataset of all the citation data (2022). <https://doi.org/10.6084/m9.figshare.6741422.v14>
16. Pradhan, T., Pal, S.: A hybrid personalized scholarly venue recommender system integrating social network analysis and contextual similarity. *Futur. Gener. Comput. Syst.* **110**, 1139–1166 (2020). <https://doi.org/10.1016/j.future.2019.11.017>
17. Pradhan, T., Pal, S.: A multi-level fusion based decision support system for academic collaborator recommendation. *Knowl.-Based Syst.* **197**, 105784 (2020). <https://doi.org/10.1016/j.knosys.2020.105784>
18. Pradhan, T., Sahoo, S., Singh, U., Pal, S.: A proactive decision support system for reviewer recommendation in academia. *Expert Syst. Appl.* **169**, 114331 (2021). <https://doi.org/10.1016/j.eswa.2020.114331>
19. Robertson, S.E., Zaragoza, H.: The probabilistic relevance framework: BM25 and beyond. *Found. Trends Inf. Retr.* **3**(4), 333–389 (2009). <https://doi.org/10.1561/1500000019>
20. Rollins, J., McCusker, M., Carlson, J., Stroll, J.: Manuscript matcher: a content and bibliometrics-based scholarly journal recommendation system. In: Proceedings of the Fifth Workshop on Bibliometric-enhanced Information Retrieval (BIR) co-located with the 39th European Conference on Information Retrieval (ECIR 2017), Aberdeen, UK, 9th April 2017. <http://ceur-ws.org/Vol-1823/paper2.pdf>
21. Schuemie, M.J., Kors, J.A.: Jane: suggesting journals, finding experts. *Bioinformatics (Oxford, England)* **24**(5), 727–728 (2008). <https://doi.org/10.1093/bioinformatics/btn006>
22. Schäfermeier, B., Stumme, G., Hanika, T.: Towards Explainable Scientific Venue Recommendations. [arXiv:2109.11343](https://arxiv.org/abs/2109.11343) (2021). <http://arxiv.org/abs/2109.11343>
23. Yu, S., et al.: PAVE: personalized academic venue recommendation exploiting co-publication networks. *J. Netw. Comput. Appl.* **104**, 38–47 (2018). <https://doi.org/10.1016/j.jnca.2017.12.004>
24. Zawali, A., Boukhris, I.: Academic venue recommendation based on refined cross domain. In: Abraham, A., Gandhi, N., Hanne, T., Hong, T.-P., Nogueira Rios, T., Ding, W. (eds.) ISDA 2021. LNNS, vol. 418, pp. 1188–1197. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-96308-8_110
25. ZhengWei, H., JinTao, M., YanNi, Y., Jin, H., Ye, T.: Recommendation method for academic journal submission based on doc2vec and XGBoost. *Scientometrics* **127**(5), 2381–2394 (2022). <https://doi.org/10.1007/s11192-022-04354-1>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

