

# Measuring Impact of False Negatives at Citation Screening Step on the Outcome of Systematic Review

Wojciech Kusa<sup>1</sup>, Guido Zuccon<sup>2</sup>, Petr Knoth<sup>3</sup> and Allan Hanbury<sup>1</sup>

<sup>1</sup>TU Wien, Vienna, Austria

<sup>2</sup>The University of Queensland, Brisbane, Australia

<sup>3</sup>The Open University, Milton Keynes, The United Kingdom

## Abstract

The current practice for evaluating search strategies and methods for the automated citation screening of systematic literature reviews relies on counting the number of positive publications (i.e. to be included in the review) and negative publications (i.e. to be excluded). Large attention is put in promoting the retrieval of all positive publications, through a high-attention to recall metrics, and demoting the retrieval of negative publications, through precision-oriented or cost metrics. This practice, however, does not accurately reflect the real-world scenario, as not all included publications have the same level of importance to the final outcome of the review. If excluded, some publications could significantly change the conclusion of the review, while not including others may not have much impact. So, for example, failing to retrieve these publications with less impact on the review would lead to lower recall values but to no practical effects on the review.

We propose a new evaluation framework that models the effect the identification of specific publications has on the outcomes of the review. We briefly demonstrate the framework by extracting metadata from RevMan files, estimating outcomes for arbitrary rankings on a sample of CLEF TAR 2019 topics, and measuring how closely the obtained outcomes are if the arbitrary rankings were used to the outcomes of the original review. we present challenges in fully operationalising our framework with some outlook on how these challenges can be tackled in the future.

## Keywords

Systematic Literature Reviews, Citation screening, Evaluation, Study outcomes

---

*ALTARS'23: 2nd Workshop on Augmented Intelligence for Technology-Assisted Reviews Systems: Evaluation Metrics and Protocols for eDiscovery and Systematic Review Systems, April 02, 2023, Dublin, Ireland*

✉ wojciech.kusa@tuwien.ac.at (W. Kusa); g.zuccon@uq.edu.au (G. Zuccon); petr.knoth@open.ac.uk (P. Knoth); allan.hanbury@tuwien.ac.at (A. Hanbury)

🌐 <https://wojciechkusa.github.io> (W. Kusa)

🆔 0000-0003-4420-4147 (W. Kusa); 0000-0003-0271-5563 (G. Zuccon); 0000-0003-1161-7359 (P. Knoth); 0000-0002-7149-5843 (A. Hanbury)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)