Introduction

Introduction

Scientific production has experienced continuous exponential growth in the last decades (Bornmann & Mutz, 2015; Larsen & Von Ins, 2010). This is especially true for biomedical research, a trend further increased by the COVID19 pandemic, thanks to decreased journals' processing time and the diffusion of preprint databases' usage (Aviv-Reuven & Rosenfeld, 2021; Horbach, 2020; Hoy, 2020). Consequently, it gets harder for researchers and practitioners to stay up-to-date on the latest findings in their field. Secondary research is of paramount relevance in this scenario, providing valuable summaries of the state of the art, but is getting ever more demanding in terms of time and human resources (Allen & Olkin, 1999; Bastian et al., 2010; Borah et al., 2017; A. M. Cohen et al., 2010).

The article collection and screening phases of a systematic review are particularly problematic tasks [systematic review tutorial]. First, relevant published research needs to be collected from scientific databases through appropriately built search queries. The construction of search queries is a particularly delicate task (Hammerstrøm et al., 2010; Lefebvre et al., 2011), requiring both domain and some knowledge of the databases' query languages; the goal is to produce a set of results containing all relevant articles (high sensitivity) while keeping the total number low (high specificity), focusing on the first aspect at the cost of the second (Hammerstrøm et al., 2010). If an integrated search tool is not used, manual work is required to download, store and organize the publication data. Searching manually may be complicated by limits in the number of records that can be downloaded at once [] and the necessity to harmonize different formats and resolve duplication. The citation screening phase consists of classifying the collected publications as relevant or not for the topic of interest (Bannach-Brown et al., 2019). It is usually the more resource-demanding task of a systematic review: even with appropriately built search queries, the returned results easily range in the tens of thousands of which just a small fraction is actually relevant (Lefebvre et al., 2011). It was estimated that labelling 10 000 publications may take as much as 40 weeks of work, while the average clinical systematic review takes 63 weeks to be completed (Allen & Olkin, 1999; Bannach-Brown et al., 2019; Borah et al., 2017). A consequence is that often systematic reviews are already outdated once they are published (E. M. Beller et al., 2013).

The field of Data Science applied to evidence synthesis and acquisition has greatly matured in the last years (E. Beller et al., 2018; Marshall & Brereton, 2015; Tsafnat et al., 2014). Through the application of natural language processing (NLP), it is possible to transform free text into quantitative features, with various levels of abstraction and generalization (Ananiadou & McNaught, 2006; K. B. Cohen & Hunter, 2008); with machine learning (ML), such text-derived data can be used to map and reproduce human judgment and automatize the citation screening (Ikonomakis et al., 2005). One of the most successful approaches to text screening is active ML: researchers review the predictions made by the ML algorithm, focusing on the most uncertain predictions ("uncertainty" based active ML) or on those giving a document a higher probability of being relevant ("certainty" based active ML) (Miwa et al., 2014).

The automatization of systematic reviews has been ripe with improvements in the last years, and it is possible to foresee that such techniques will become a standard approach (E. Beller et al., 2018). See Ananiadou et al. (2009); O'Mara-Eves et al. (2015); Tsafnat et al. (2013); Jonnalagadda et al. (2015) for in-depth evaluations of research in this field. Many of these studies have generated a set of user-friendly commercial and free-to-use tools (see Marshall & Brereton, 2015, table 1). Despite the significant advances, each method has various shortcomings, both in usability and efficiency (O'Mara-Eves et al., 2015). We tried to address some of these limitations by creating an integrated framework that tries to simplify systematic reviews in innovative ways: our framework provides solutions for automatic citation collection and management from multiple online sources and a Bayesian, active machine learning-based tool to support humans in the screening

task, exponentially reducing the number of records to review. In particular, we exploit the characteristics of Bayesian analysis to overcome the shortcoming of both "certainty" and "uncertainty" based active learning and the difficulties in defining stopping rules (Miwa et al., 2014). Finally, we propose an experimental method to semi-automatically generate efficient search queries given an already labelled set of publications.

We apply this solution in the production of a systematic review evaluating the mathematical modelling of patient referral networks among hospitals and their impact on the diffusion of healthcare-associated pathogenic microorganisms with a focus on antimicrobial-resistant strains. The systematic review protocol is published in [].

Allen, I. E., & Olkin, I. (1999). Estimating time to conduct a meta-analysis from number of citations retrieved. *Jama*, *282*(7), 634–635.

Ananiadou, S., & McNaught, J. (2006). *Text mining for biology and biomedicine*. Citeseer.

Ananiadou, S., Rea, B., Okazaki, N., Procter, R., & Thomas, J. (2009). Supporting systematic reviews using text mining. *Social Science Computer Review*, *27*(4), 509–523.

Aviv-Reuven, S., & Rosenfeld, A. (2021). Publication patterns' changes due to the COVID-19 pandemic: A longitudinal and short-term scientometric analysis. *Scientometrics*, 1–24.

Bannach-Brown, A., Przybyła, P., Thomas, J., Rice, A. S., Ananiadou, S., Liao, J., & Macleod, M. R. (2019). Machine learning algorithms for systematic review: Reducing workload in a preclinical review of animal studies and reducing human screening error. *Systematic Reviews*, *8*(1), 1–12.

Bastian, H., Glasziou, P., & Chalmers, I. (2010). Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? *PLoS Medicine*, *7*(9), e1000326.

Beller, E., Clark, J., Tsafnat, G., Adams, C., Diehl, H., Lund, H., Ouzzani, M., Thayer, K., Thomas, J., Turner, T., & others. (2018). Making progress with the automation of systematic reviews: Principles of the international collaboration for the automation of systematic reviews (ICASR). *Systematic Reviews*, *7*(1), 1–7.

Beller, E. M., Chen, J. K.-H., Wang, U. L.-H., & Glasziou, P. P. (2013). Are systematic reviews up-to-date at the time of publication? *Systematic Reviews*, *2*(1), 1–6.

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, *7*(2), e012545.

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222.

Cohen, A. M., Adams, C. E., Davis, J. M., Yu, C., Yu, P. S., Meng, W., Duggan, L., McDonagh, M., & Smalheiser, N. R. (2010). Evidence-based medicine, the essential role of systematic reviews, and the need for automated text mining tools. *Proceedings of the 1st ACM International Health Informatics Symposium*, 376–380.

Cohen, K. B., & Hunter, L. (2008). Getting started in text mining. *PLoS Computational Biology*, *4*(1), e20.

Hammerstrøm, K., Wade, A., Jørgensen, A.-M. K., & Hammerstrøm, K. (2010). Searching for studies. *Education*, *54*(11.3).

Horbach, S. P. (2020). Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19. *Quantitative Science Studies*, *1*(3), 1056–1067.

Hoy, M. B. (2020). Rise of the rxivs: How preprint servers are changing the publishing process. *Medical Reference Services Quarterly*, *39*(1), 84–89.

Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. *WSEAS Transactions on Computers*, *4*(8), 966–974.

Jonnalagadda, S. R., Goyal, P., & Huffman, M. D. (2015). Automating data extraction in systematic reviews: A systematic review. *Systematic Reviews*, *4*(1), 1–16.

Larsen, P., & Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics*, *84*(3), 575–603.

Lefebvre, C., Manheimer, E., Glanville, J., Higgins, J., & Green, S. (2011). Searching for studies (chapter 6). *Cochrane Handbook for Systematic Reviews of Interventions Version*, *510*.

Marshall, C., & Brereton, P. (2015). Systematic review toolbox: A catalogue of tools to support systematic reviews. *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, 1–6.

Miwa, M., Thomas, J., O'Mara-Eves, A., & Ananiadou, S. (2014). Reducing systematic review workload through certainty-based screening. *Journal of Biomedical Informatics*, *51*, 242–253.

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews*, *4*(1), 1–22.

Tsafnat, G., Dunn, A., Glasziou, P., & Coiera, E. (2013). *The automation of systematic reviews.* British Medical Journal Publishing Group.

Tsafnat, G., Glasziou, P., Choong, M. K., Dunn, A., Galgani, F., & Coiera, E. (2014). Systematic review automation technologies. *Systematic Reviews*, *3*(1), 1–15.