Graefe's Archive for Clinical and Experimental Ophthalmology (2024) 262:1013–1014
https://doi.org/10.1007/s00417-023-06359-9

EDITORIAL

170 years of data-mining: history and future

Bernd Kirchhof[1]

Received: 20 December 2023 / Revised: 20 December 2023 / Accepted: 28 December 2023 / Published online: 17 January 2024

Data is not everything, but without data, everything is nothing [1]. If data were not so essential for progress and development, people would not occasionally try to falsify them (fake news). Access to original data provides us with a solid basis for discussion. However, it is still more difficult to find suitable data in the "biomedical" area. This is the reason review articles are in demand and we have been accessing journals like *Graefe's Archives* since 1854. This also explains why we search through index volumes in physical libraries and why we learn to use virtual indexed search algorithms and machines like "Pub-Med" or "Google-Scholar". Time and access to knowledge resources are identified as the most common barriers to clinical information seeking [2]. And yet, unfortunately, there are still authors whose submissions are not found suitable for publication after review because of "insufficient new information".

And yet there is the oncologist who no longer practices because he is no longer able to manage the flood of scientific publications down to the contemporary therapy of an individual patient. And he also exists, the colleague who reads several journals, subscribes to them and has them bound, and prepares excerpts for the further training of his scholars.

A solution is at hand. Instead of finding similar articles using keywords, language is vectorized using AI algorithms. In this way, literature can be searched using natural language (natural language processing (NLP)). NLP [3] is a field of artificial intelligence (AI) that enables computers to understand, interpret, and manipulate human language. In NLP-based keyword augmentation, prior knowledge and reasoning techniques are employed to remove irrelevant search results. Some of us certainly appreciate the necessary chatbots such as those from "Bard" (Google), Bing (Microsoft) or Chat GPT (Open AI) for "data mining" on general topics. These chatbots train for example on "Wikipedia". "Bard" and comrades, however, are unsuitable for the "biomedical" area: firstly, because copyright law prevents access to scientific databases such as PubMed (National Library of Medicine) and secondly, because they generate prose, mostly without a reference list. For the scientific search, suitable chatbots are for example "Elicit" [4], "Scite" [5] and "Consensus" [6]. They depend on free scientific databases (in scite's case) or have access to paywalled research articles through partnerships with publishers. Chatbots, like "Elicit", are addressed by natural language to generate reference lists. They implement NLP to remove irrelevant search results. Firms that own large proprietary databases of scientific abstracts and references are now joining [7]. Elsevier (Scopus-AI) [8], Springer Nature and Clarivate are beginning to add AI search functions to their article and book archives [9]. Clarivate integrates the Web of Science. But be careful! AI searches are prone to "hallucinations". Such childhood illnesses must be taken into account. So far, you can only go back a few years and search in abstracts. By the way, researchers in Germany only have free access to full texts through research institutions (e.g. universities) that have purchased quotas from publishers. Free access to full-text articles would certainly also be helpful for doctors in private practice, i.e. outside of universities.

In the interest of health care, would it not make more sense if the medical associations also entered into partnerships with publishers who would be open to collect infomation from full-text libraries? Then, perhaps the colleague mentioned previously would have the courage to return to patient care in this rapidly developing field of oncology.

Funding Open Access funding enabled and organized by Projekt DEAL.

✉ Bernd Kirchhof
bernd@kirchhof-koeln.de

[1] Center of Opthalmology, University of Koeln, Koeln, Germany

References

1. Obenland I (2011) Freiheit ist nicht alles, aber ohne Freiheit ist alles nichts. Weimarer Schiller Presse, Offenbach
2. Aakre CA, Maggio LA, Fiol GD, Cook DA (2019) Barriers and facilitators to clinical information seeking: a systematic review. J Am Med Inform Assoc 26:1129–1140
3. Hameed IA (2016) Using natural language processing (NLP) for designing socially intelligent robots. In: Joint IEEE International Conference on Development and Learning and Epigenetic Robotics. (ICD https://doi.org/10.1109/DEVLRN.2016.7846830. Accessed 15 Dec 2023
4. https://elicit.com/?workflow=table-of-papers. Accessed 15 Dec 2023
5. https://scite.ai/assistant?utm_source=google&utm_medium=cpc&utm_campaign=brand&utm_term=citeai&gclid=Cj0KCQiA7OqrBhD9ARIsAK3UXh0zZkpO1ouEECgclfUQldJhK0UhF7Dr2wVwIvTQ8DMTGmjQl7erio8aAnAJEALw_wcB. Accessed 15 Dec 2023
6. https://consensus.app/. Accessed 15 Dec 2023
7. Van Noorden R (2023) Chatgpt-like AIS are coming to major science searches. Nature 620:258
8. https://www.elsevier.com/products/scopus/scopus-ai. Accessed 15 Dec 2023
9. https://group.springernature.com/de/group/media/press-releases/21-titles-to-be-renamed-supporting-inclusive-research-culture/20325548. Accessed 15 Dec 2023
10. https://www.aezq.de/medien/pdf/publikationen/manual-literaturrecherche.pdf/view. Accessed 15 Dec 2023

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.