Why we need another ten years of Bibliometric-enhanced Information Retrieval*

Mike Thelwall

University of Wolverhampton, UK

m.thelwall@wlv.ac.uk

Congratulations to the organisers for the huge success of the Bibliometric-enhanced Information Retrieval conference series. Organising even a single workshop is difficult, and a series of ten is a major achievement. Workshops help to forge people into a community and define a field, so this series has had a lasting impact on academic research.

Scholarly information retrieval is still a major challenge that still needs more work. I teach literature searching to maths and computing students and conduct frequent literature searches. Despite this experience, I frequently must remind myself that literature searching is more difficult than it seems.

The first thing that I tell my students is that finding relevant research is difficult, even though Google Scholar (their main search tool) gives a lot of help. For a beginner, a key task is learning the terminology of a field, which is often not straightforward and is essential for good literature searches. I give the example of Google. To conduct a literature search for the technology behind a search engine like Google, the keyword Google is not particularly useful, neither is the more general phrase "search engine." Instead, after reading into the topic, the students will discover that the name given to the field producing the core research behind (the search part of) Google is information retrieval, so this is a good term to search for (although it gets too many results). I think that it is a major challenge to develop an information retrieval system that can accommodate beginners in this initial task: translating the words that they would use to describe a topic into the jargon used within the field.

Using the wrong terminology has caught me out many times when exploring unfamiliar fields. Once I submitted what I thought was the first article on a topic, but a referee pointed to a monograph and an edited volume of collected papers from a workshop, all devoted to the topic but using different terminology to me. My basic mistake is to deduce that there is no relevant research if I have used reasonable keywords to search for it and got no results.

Once a beginner has learned the jargon of a field there are still major search challenges. Article titles can be misleading, important results can be in articles with a different focus, and there may be far too many relevant articles to read individually. I tell my students to look out for productive scholars in a field and

---

* A companion video is hosted at https://youtu.be/Ld1s6mEpA2Q.

CEUR-WS.org/Vol-2591/paper-12.pdf

read their most cited work, to look for review articles, to find highly cited articles, and to citation chain from core articles to find more recent updates. These mostly exploit bibliometric information to reduce the amount of reading needed. One interesting challenge is that authors know the topic of their article so well that they may not be the best people to decide on a relevant title. Usually a title makes sense to me after a paper has been read but sometimes the content is not clear before then. In addition, whilst in some fields titles often directly describe the core outcome ("IL-1 receptor blockade restores autophagy and reduces inflammation in chronic granulomatous disease in mice and in humans"), in others titles might be funny ("Another one bites the dust: faecal silica levels in large herbivores correlate with high-crowned teeth"), speculative ("Why does unsupervised pre-training help deep learning?"), or include an illustrative quote ("'Do I really have to eat that?': A qualitative study of schoolchildren's food choices and preferences"). This makes the titles more interesting but adds to the human challenge of identifying relevant documents and adds greatly to the information retrieval challenge of finding and ranking documents matching the searcher's information need.