Universiteit
Leiden
The Netherlands

The relevance of impact: bibliometric-enhanced legal information retrieval

Wiggers, G.

Citation
Wiggers, G. (2023, March 8). *The relevance of impact: bibliometric-enhanced legal information retrieval*. *SIKS Dissertation Series*. Retrieved from https://hdl.handle.net/1887/3570499

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3570499

Note: To cite this publication please use the final published version (if applicable).

Chapter 6

Discussion and conclusions

"So once you know what the question actually is, you'll know what the answer means." Douglas Adams – The Hitchhiker's Guide to the Galaxy

This research started from the top down, with the idea that it must be possible to improve the ranking of legal IR systems by adding meta-information about the documents, and the availability of a substantial amount of data. Certain that something like this had already been done, a Google search led me[1] to the discovery of altmetrics [100] and the theory behind Google's PageRank algorithm [95]. With this confirmation that a theoretical underpinning for this idea existed, and with sufficient data available, this could have led to immediate implementation in the Legal Intelligence system. Users would have been happy, or not, and the development team would move on to the next project. But the question 'what does it mean', moved it from a mere idea, ready to be implemented in two sprints, to a full PhD project.

It soon became obvious that this research would become interdisciplinary. The scholarly field of IR focuses a lot on state-of-the-art web-search, whilst domain specific (e.g. legal or archaeological) IR applications often still rely heavily on BM25[70], developed in the 1980's and 1990's [65]. That the latest academic developments focus only on web-search is unfortunate, since domain specific IR applications are often very valuable to end-users.

Because of this academic focus on web-search it can be hard to find the right context for domain specific IR research (and to find venues to publish that work). This missing context means an interdisciplinary approach is required which looks at all the steps in the process from foundational theory to application [22]. It requires a combination of domain specific (user) knowledge as well as information science and information retrieval.

Every step towards implementation of the bibliometric-enhanced ranking model led to more questions. A big challenge of interdisciplinary research is that of vocabulary. The vocabulary in bibliometrics and information science is not the same as that of information retrieval, which makes it hard to find relevant literature. In that regard the work of Van Opijnen

[1]Because of the more personal nature of this discussion, and because there are no co-authors for this chapter, this discussion is written in the first-person singular as opposed to the earlier chapters, which were written in the first-person plural.

and Santos [131] became a Rosetta Stone of sorts. Their paper applies the work of Saracevic [110] to legal IR, and thereby not only introduced the concept of bibliographic relevance as a feature of legal IR, but for me also bridged a language gap between information science and computer science. The BIR community[2] helped further bridge this gap and gave a name to the research, whilst the JURIX community[3] helped bridge the gap between the legal domain and computer science.

The question whether legal scholars and legal professionals have the same perception of relevance, and thereby whether one legal IR system can serve both user groups, was the first question answered in this research, in what has become Chapter 2. Using a survey and conducting a PERMANOVA on the answers we found no significant difference in the factors reported by these two user groups. This meant that there is no reason to treat these sub-groups differently in legal IR systems (sub-question 2). The agreement of the respondents on factors of relevance, in a survey setting without situational relevance, also shows the existence of domain relevance as described by Van Opijnen and Santos as 'relevance of information objects within the legal domain ...' [131] (sub-question 1).

The next challenge was the question 'what does a citation mean in the legal domain'? The work of Stolker [120] provided valuable information about the publication culture of Dutch legal professionals. The work of Merton [87] lead to the question what a citation in legal documents represents, a question answered by Snel [116, 117] (sub-question 3). This also introduced the second main challenge of interdisciplinary research: sidetracks. Citation metrics for scholarly evaluation have been discussed in the Dutch legal domain [130, 108], but their use for Dutch legal IR less [94]. The negative light in which they had been discussed for research evaluation may prove to be the reason why this is the case. It was tempting to become part of the debate on the merits of bibliometrics for research evaluation, but the scope of the research had to be limited.

In Chapter 3 we conducted a data analysis, which confirmed the work

[2]See `https://sites.google.com/view/bir-ws/home`
[3]`http://jurix.nl/`

of Snel and Stolker and showed many cross-citations between scholarly and non-scholarly documents. We also found cross-usage. The literature suggested, and the data showed, a disregard by both scholarly and non-scholarly users of the distinction between scholarly and practitioner-oriented publications. This confirmed to us that the affiliation of the user (legal scholar or legal practitioner) is not a suitable factor to differentiate rankings on (sub-question 4). It also provided us with the theoretical insight that citations in legal documents measure part of a broad scope of impact, or relevance, on the entire legal field. We say *part* of a broad scope of impact, because for documents that are never cited, the illusion could exist that they have had no impact on the field even though they may have had a different form of impact. We therefore suggest to combine citation metrics with usage metrics.

The third question, on the interdisciplinary sphere of academia and industry, was 'what does this mean in practice', or how to implement this. The first example I found of how to implement usage and citation counts into a live IR system was the work of Kurtz and Henneken [75]. This work led to the work of the CWTS [135] on different ways to implement citation counts, from raw counts to normalized counts, and why some methods are preferred over others.

The most prominent example of 'what does it mean in practice' has been and still is the question of evaluation. The aim of BIR is to improve IR systems. But what is 'better', and how to measure it? Järvelin [63] stated that to understand what an effective method of evaluation for a (legal) IR system is, we need to understand the theoretical background (sub-question 5). Azzopardi and colleagues [9, 11] have developed a framework to create user models to aid in this. But we ran into many practical problems trying to evaluate a live domain specific IR system, as demonstrated in Chapter 4 (sub-question 6).

In the end the works of Järvelin's [63] and Azzopardi and colleagues [9] inspired us to create a cost based model for evaluation, as discussed and implemented in Chapter 5. Chapter 5 also describes the other practical questions asked in the implementation process and how we found the answers. We discovered, through the work of Geist [50], that the completeness

ideal assumed by the IR community to exist in legal IR, is offset by the research reality of legal practice. We also found that citations appear, and are a reliable predictor of future citations, much earlier than thought [108] (sub-question 7). We also confirmed the correlation between usage and citations found by, amongst others, Brody et al. [24] (sub-question 8), and were thereby able to confirm our theory from Chapter 3 that usage data can be used along citation data to represent different flavors of impact. Applying a linear mixed model (LMM) to data from this user model before and after the introduction of our Bibliometric-enhanced ranking algorithm, we found a reduction of cost for the user of 2 to 3% for situations other than known-item retrieval (sub-question 9).

6.1 The answer (to the research question)

The research question of this thesis is **How can bibliometrics improve common ranking algorithms in legal information retrieval?** Combining the answers of the sub-questions above, we can conclude that a bibliometric-enhanced ranking feature needs to take into account both usage and citations (two flavors of impact relevance), and needs to increase in influence as the reliability of the data grows (in combination with a recency feature that gives new documents the benefit of the doubt and decreases at the same rate as the bibliometric feature increases). With such a bibliometric-enhanced ranking feature we can reduce the cost required from legal professionals (whether practitioner, scholar or legal information professional) to find enough information for their information need.

The contribution of this thesis lies not only in the answer as a whole, but in the steps taken to reach this conclusion:

1. that there is no reason at this point to differentiate the ranking for sub-groups of users of legal IR systems based on their role or affiliation;

2. that bibliometrics can be seen as a manifestation of impact relevance and that citations in legal documents represent part of a broader

form of impact on the legal domain as a whole, and should be used alongside usage data to also see the impact on non-authors;

3. how common evaluation methods are limited by the characteristics of legal IR when used for the evaluation of live domain specific search engines, and how a cost-based evaluation can be used instead;

4. a clear step-by-step description how such a bibliometric-enhanced ranking variable can be created, and that ranking algorithms in legal IR can indeed be improved using bibliometrics.

6.2 What does this answer mean for the future?

The research in this thesis has raised even more questions for me, and possibly for others. For example: to what extent the perception of relevance differs from snippets as opposed to full documents, what the optimal level of detail is to use in normalization of bibliometrics, and how it is possible that legal documents get cited so quickly in published documents.

Future work should focus on the development of domain specific, live evaluation models, so that non-academic developers can adequately evaluate their system. Aside from the benefits for their own system, this will allow them to find venues to publish their work more easily. In a similar way that companies like Google, Yahoo and Microsoft contribute a lot to the scientific community around web-search by producing research output and datasets, we need companies to further the academic debate about domain specific IR. But in order to participate in the academic debate, these companies need to use evaluation methods that the scientific community agrees on.

It is in the interest of the users, and the legal profession as a whole, that the legal IR systems implement these evaluation methods, and cooperate with researchers to improve the systems. By providing insight into the completeness ideal and research reality, users themselves can also contribute to the improvement of these systems.

Azzopardi and colleagues [9] have developed the C/W/L framework (pronounced 'cool'), which (based on user data) can be used to predict user interactions with the search results based on their position in the results list, and by extension can also be used to evaluate IR systems. A similar model, tailored to small-scale live IR systems, would be very suitable, since it can be altered to suit the characteristics of the users of the domain. With such a system changes in domain specific IR systems can be reported in a uniform manner, and easily be interpreted by the wider IR community. This will force open the door for the IR community to pay more attention to domain specific IR.

A unified and interpretable evaluation method like C/W/L will hopefully also remove roadblocks for publication of such work. Currently IR journals consider the sample sizes of domain specific IR small, and a barrier to publication. A recognised evaluation method may remedy this. Similarly, work in the BIR community is often published in a special issue of the journal of Scientometrics[4], but does not have an obvious outlet in the IR community. This makes it harder for researchers in BIR to reach the IR community. These challenges and limitations of interdisciplinary work are often not recognised by journals when making their publication decisions while they should be.

Within universities, interdisciplinary research should be further normalized. Not just through the creation of interdisciplinary research groups, which appears to be a growing development, but also through the facilitating of introductory skill-based courses for (senior) researchers and assistance with navigating publishing interdisciplinary work. Often, the skills associated with a field or discipline are taught throughout content-heavy courses. But for researchers from other disciplines, this means following courses of which the majority of information is irrelevant, or independent (online) study where they do not benefit from the knowledge and skills of their colleagues. By creating modular, skill-based courses for researchers (e.g. programming or descriptive statistics), they are able to select those (LEGO) building blocks that they need, in an environment which best suits

[4] e.g. `https://sites.google.com/view/scientometrics-si2019-bir`

their needs.[5]

Creating a learning environment where researchers can learn new research methods with peers might also encourage researchers to use a wider range of research methods (e.g. a law scholar might start using quantitative research methods next to their normative research works). Using interdisciplinary skills in their domain (a transfer learning of sorts) may also lead to the construction of novel ideas. With the added benefit that increased contact between researchers from different faculties may lead to more interdisciplinary collaboration.

[5]This research was part of the interdisciplinary data science research program. I started this research with limited knowledge of statistics, so my own PhD trajectory is a good example that it can be difficult to find such courses. The options available were several bachelor courses with the skills weaved into domain specific knowledge, or online learning.