



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).

The Relevance of Impact

bibliometric-enhanced legal information retrieval

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op woensdag 8 maart 2023

klokke 11.15 uur

door

Gineke Wiggers

geboren te Hilversum

in 1988

Promotor: Prof.dr.mr. G.J. Zwenne
Co-promotor: Dr. S. Verberne

Promotiecommissie: Prof.mr.dr.ir. B.H.M. Custers
Dr. F. Dechesne
Prof.dr.ir. A.P. de Vries (Radboud Universiteit, Nijmegen)
Prof.dr. A. Hanbury (TU Wien, Austria)
Prof.mr. C.J.J.M. Stolker

Authornote:

Gineke Wiggers is also employed as a business analyst at Legal Intelligence, a Wolters Kluwer product. Legal Intelligence/Wolters Kluwer did not provide any financial contribution to this research, the research has been funded solely by Leiden University.

SIKS Dissertation Series No. 2023-04

The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Contents

1	Introduction	1
1.1	The legal importance of legal information retrieval	6
1.2	A brief introduction to information retrieval (IR)	8
1.3	Features of legal IR	9
1.4	An overview of common evaluation methods in IR	11
1.5	A brief introduction to bibliometrics	13
2	Relevance	17
2.1	Introduction	18
2.2	Background	21
2.3	Methods	27
2.3.1	Study design	27
2.3.2	Participants	28
2.3.3	Selection of stimuli	29
2.3.4	Extracting and mapping relevance factors	33
2.3.5	Statistical analysis	39
2.4	Results	40
2.4.1	Relevance factors	42
2.4.2	Differences between user sub-groups	46
2.5	Discussion	50
2.5.1	Implications	50
2.5.2	Limitations of the study	51
2.6	Conclusions	52

2.7	Acknowledgements	53
3	Citation Metrics	55
3.1	Introduction	57
3.2	Literature analysis	58
3.2.1	Citations as a Form of Impact Measurement	59
3.2.2	Citations in Dutch Legal documents	61
3.2.3	‘Scholarly’ Legal Documents	62
3.2.4	Citations in Information Retrieval	65
3.3	Methods	67
3.3.1	Document sampling criteria	68
3.3.2	Document classification	68
3.3.3	Readership	73
3.4	Results	73
3.4.1	Citations between documents	74
3.4.2	Usage of documents	75
3.5	Discussion	76
3.5.1	Inter-rater agreement	76
3.5.2	Citations between documents	77
3.5.3	Usage of documents	81
3.5.4	Using bibliometrics in legal IR	82
3.6	Conclusion	83
3.7	Acknowledgements	86
4	Evaluation	89
4.1	Introduction	90
4.2	Legal IR	92
4.2.1	The User	93
4.2.2	The IR Systems	95
4.2.3	The Documents	96
4.2.4	Relevance	97
4.2.5	Small Data	98
4.2.6	Legal Search and Professional Search	99
4.2.7	Summary	100

4.3	Expected limitations to common evaluation methods	101
4.3.1	Test Collections	101
4.3.2	User surveys	103
4.3.3	A/B testing	104
4.4	Empirical assessment of evaluation methods	105
4.4.1	Test collection based on expert relevance judgments	105
4.4.2	Test collection based on implicit feedback	107
4.4.3	Survey for ranking preferences	110
4.4.4	Survey based on the Net Promotor Score	114
4.5	Conclusion	114
4.6	Acknowledgements	116
5	Algorithm	117
5.1	Introduction	118
5.2	Background	120
5.2.1	Citations and usage in bibliometrics	120
5.2.2	Correlation between usage and citations	121
5.2.3	Usage in evaluation	121
5.3	Data analysis	123
5.4	Methods	129
5.4.1	Our proposed bibliometric-enhanced ranking variable	130
5.4.2	Evaluation	134
5.5	Results and analysis	139
5.5.1	Results	139
5.5.2	Statistical analysis (without clicks)	139
5.5.3	Statistical analysis (including clicks)	141
5.5.4	Practical significance	141
5.5.5	Analysis of long sessions	142
5.6	Conclusions	142
5.7	Acknowledgements	143
6	Discussion and conclusions	145
6.1	The answer (to the research question)	149
6.2	What does this answer mean for the future?	150

IV	<i>CONTENTS</i>
Bibliography	153
Summary	171
Samenvatting (Dutch Summary)	173
Appendices	177
A Composition of search results	179
A.1 Example query 1	179
A.2 Example query 2	180
B Queries and Seed Documents	183
C Seed Documents and Data	187
D Baseline and degraded ranking	189
E Example of Survey Question	193
F Curriculum Vitae	195
G Acknowledgements	197