Towards a theory of jurisprudential relevance ranking
Using link analysis on EU case law

Staffan Malmgren
Graduate thesis
Written for the Master of Laws degree at Stockholm University.

Supervisor: Cecilia Magnusson Sjöberg.

Cover: The citation network of all available ECJ case law. The graph contains 14,327 nodes (decisions) with 107,473 edges (references to other cases). Case labels (CELEX numbers) are sized proportionally to their indegree (number of citing cases). The entire graph was rendered in Gephi. The nodes have been partitioned into 91 communities with different colors using the Louvin method. The graph has been laid out using the Force Atlas 2 algorithm.

This work is available under the CC BY 3.0 license, see http://creativecommons.org/licenses/by/3.0/legalcode for details. It was created using LaTeX September 4, 2011.
Abstract

The concept of relevance is central to both jurisprudence and information retrieval. But what do we mean when we say that something is relevant? Is there a difference between how relevance is understood in jurisprudence and in information science? Which aspects that are unique to legal information have effect on relevance? And can we use this to build better information retrieval systems for legal information?

This thesis discusses the concept of relevance, both as it is used in general and in legal contexts. It describes the retrieval models used in modern information systems, and what notion these models have of relevance. By examining the legal reasoning process, in particular the process of finding legal information, it attempts to find a retrieval model and a function for ranking that is adapted to legal information.

This function is implemented and evaluated against a traditional probabilistic ranking algorithm. It is shown to perform substantially better for all tested information need scenarios.
“Relevance is a human notion. How does one do anything neatly and precisely and unambiguously with human notions?”\(^1\)

“Science is what we understand well enough to explain to a computer. Art is everything else we do.”\(^2\)

---


\(^2\)Donald Knuth, in the foreword to Marko Petkovsek/Herbert S. Wilf/Doron Zeilberger: A = B, 1996
## Contents

1 Introduction 6
   1.1 Information retrieval systems 6
   1.2 Legal information, norms and meta-norms 8
   1.3 Motivation 10
   1.4 Hypotheses 13
   1.5 Method description 14
   1.6 Structure of this thesis 14

I Relevance and retrieval 16

2 The concept of relevance 17
   2.1 The relevance of relevance 17
   2.2 The many faces of relevance 18
      2.2.1 Topicality 18
      2.2.2 Implicit use orientation 19
      2.2.3 Multidimensionality 20
      2.2.4 Logical relevance 21
      2.2.5 A relevance framework 22
   2.3 Relevance, examined 23
      2.3.1 Manifestations of relevance 24
      2.3.2 Determining relevance 27
      2.3.3 Measuring relevance 31
      2.3.4 Difference in relevance judgments 32
      2.3.5 Models of relevance 33
   2.4 Relevance in jurisprudence 35
      2.4.1 Legal information as compared to scientific information 35
      2.4.2 Jurisprudence and legal method 36
      2.4.3 The notion of legal relevance 38

3 Information retrieval 45
   3.1 General information retrieval 45
      3.1.1 Retrieval models 47
II A legal relevance ranking function 70

4 Previous work 71
4.1 Automated concept indexing and classification for retrieval . 71
4.2 Knowledge engineering and case based retrieval approaches . 73
4.3 Legal citation analysis . 74

5 A prototype of a legal relevance function 76
5.1 Corpus, citation network and information needs . 76
5.1.1 Document collection . 76
5.1.2 Information needs . 77
5.1.3 Document structure . 78
5.1.4 Citation network properties . 78
5.2 Prototype construction . 80
5.3 Evaluation . 81
5.3.1 Set of information needs . 81
5.3.2 Set of baseline queries . 81
5.3.3 Gold standard relevance judgments . 81
5.3.4 Results . 82

6 Conclusion and future work 84
6.1 Analysis of results . 84
6.2 Analysis of method . 85
6.3 Suggestions for future work . 85

A Technical description of the prototype 87
A.1 Acquiring, installing and running . 87
A.1.1 Prerequisites . 87
A.1.2 Installing . 88
A.1.3 Running . 88
A.2 System and information architecture . 90
A.2.1 Processing of information . 90
A.2.2 Document representation and semantics . 90
A.2.3 Semantical analysis of the corpus . 91