

Jurisprudential relevance ranking

Staffan Malmgren

October 1, 2008

Contents

1	Introduction	2
1.1	Hypotheses	2
1.2	Terminology	3
1.3	Method description	3
2	Jurisprudence and legal method	4
2.1	Legal relevance in jurisprudence	4
2.2	Legal method	4
2.3	Jurisprudential relevance	4
2.4	Importance, influence and relevance	4
2.5	A method for measuring relevance	4
3	Information retrieval and relevance functions	4
3.1	The boolean retrieval model	4
3.2	The probability ranking principle	4
3.3	Domain specific relevance functions	4
3.4	Formalizing a function for jurisprudential relevance	4
4	Evaluation	4
4.1	Description of corpus and queries	4
4.2	Description of the IR system prototype	4
4.3	User study	4
5	Conclusion	4
6	References	4

1 Introduction

The concept of *relevance* is central in the field of Information Retrieval (IR). The classic *boolean retrieval model*¹ of relevance in IR assumes that a document is either fully relevant to a query, or not relevant at all, and that the information retrieval system can determine this by way of the keywords and operators used.² More sophisticated models used in modern IR instead ranks documents retrieved according to some principle. A commonly used principle is *probability ranking*,³ in which documents are ordered according to the probability that the document is relevant in respect to the user's query.

Most relevance ranking functions operate on statistical and linguistic levels, e.g. by counting the frequency and placement of keywords, or by recognizing different forms of the same word (stemming). Such *general* functions are by definition not particularly tailored to a specific field of information such as legal information or biomedical texts, i.e. the functions does not contain any *domain knowledge*.

This thesis presents a relevance ranking system that is *specific* for legal documents, in that it incorporates jurisprudential knowledge. The purpose of designing a domain specific function is to achieve a better relevance ranking than general ranking functions such as Okapi BM25⁴ or PageRank.⁵ "Better" is here defined as "more similar to the order a legal expert, familiar with the area that the query concerns, would place the documents". This will show the need for designers of legal information systems to take jurisprudence into account when designing such systems.⁶

1.1 Hypotheses

Three hypotheses are proposed:

1. The legal method used by legal professionals contains rules for determining the relevance of a legal document with respect to a particular legal problem. These rules are for most part not formalized and explicit but rather exists as tacit knowledge.
2. These relevance-determining rules can be *approximately* described in a formal fashion as a ranking function.
3. A legal IR system will have better relevance ranking when using such an specific function than when using standard relevance functions.

¹Querying a document collection using keywords combined with AND, OR and NOT operators - see section 3.1 on page 3.1 for a more detailed description

²Manning et al, Introduction to information retrieval, p 4

³ibid, p. 221

⁴ibid, p. 232

⁵Page, Lawrence et al, "The PageRank citation ranking: Bringing order to the Web"

⁶Susskind, Richard, "The Future of Law", Oxford, 1996

1.2 Terminology

In this thesis, the following terms are used with the meaning indicated:

Information retrieval system: A computerized system for storing and retrieving texts according to queries. A query can be explicitly specified by the user, or it can be implicit, eg in relation to a particular context.

Jurisprudence: The theory and philosophy of law. The central question that jurisprudence seeks to answer is “What is law?”⁷

Jurisprudential relevance: a measure of how “important” a given document is for a given query for purposes of answering a legal problem.⁸

1.3 Method description

This work uses an interdisciplinary method, analysing methods used in jurisprudence and using them in an information science context.

In order to design a jurisprudential relevance ranking system, one must attempt to answer the question “What makes something legally relevant?”. The first part of this thesis concerns this question, and answers it in the form of a set of rules for assigning jurisprudential relevance (answering hypothesis 1 above). It is assumed that importance and influence are closely related to relevance.

An untested theory might be interesting and thought-provoking, but ultimately not very useful. Therefore, the second part of the thesis describes an implementation of these relevance-determining rules in the form of a ranking function (answering hypothesis 2).

In order to measure the functions effectiveness and correctness, the output of the system is then compared to what an expert in the area (EC law) would expect (answering hypothesis 3).

The corpus for the system will be all verdicts from the European Court of Justice (ECJ), augmented with various third-party sources of information, such as the journal “Europarättslig Tidskrift”. The query will not be a traditional free-text search string such as “abuse of dominant market position”, but rather the question “For a given piece of legislation, such as Article 82 of the treaty of Rome, which are the central ECD cases that interpret that article?”.⁹

⁷Wacks, *Understanding Jurisprudence*, p 1

⁸Summers, Robert S. “*Essays in Legal Theory*”, p 155

⁹For that particular query, one would expect that C-27/76, *United Brands*, would be near the top of the list.

- 2 Jurisprudence and legal method**
 - 2.1 Legal relevance in jurisprudence
 - 2.2 Legal method
 - 2.3 Jurisprudential relevance
 - 2.4 Importance, influence and relevance
 - 2.5 A method for measuring relevance
- 3 Information retrieval and relevance functions**
 - 3.1 The boolean retrieval model
 - 3.2 The probability ranking principle
 - 3.3 Domain specific relevance functions
 - 3.4 Formalizing a function for jurisprudential relevance
- 4 Evaluation**
 - 4.1 Description of corpus and queries
 - 4.2 Description of the IR system prototype
 - 4.3 User study
- 5 Conclusion**
- 6 References**