CrossMark

On the concept of relevance in legal information retrieval

Marc van Opijnen[1] · Cristiana Santos[2,3]

Abstract The concept of 'relevance' is crucial to legal information retrieval, but because of its intuitive understanding it goes undefined too easily and unexplored too often. We discuss a conceptual framework on relevance within legal information retrieval, based on a typology of relevance dimensions used within general information retrieval science, but tailored to the specific features of legal information. This framework can be used for the development and improvement of legal information retrieval systems.

Keywords Legal information retrieval · Relevance · Legal information seeking behaviour

1 Introduction

Legal information retrieval (LIR) has always been a research topic within Artificial Intelligence and Law ('AI & Law'): in 'A History of AI & Law in 50 papers' (Bench-Capon et al. 2012) seven of those 50 papers have a relation to LIR. For the legal user though much research seems to be only remotely relevant for solving their daily problems in information seeking. The underrepresentation of legal practitioners within the AI & Law community might offer an explanation: "A lawyer has always the huge text body and his degree of mastery of a special topic in mind. For a computer scientist, a high-level formalisation with many ways of using and

✉ Marc van Opijnen
marc.opijnen@koop.overheid.nl

[1] Publications Office of the Netherlands (UBR|KOOP), The Hague, The Netherlands

[2] University of Barcelona (IDT-UAB), Barcelona, Spain

[3] University of Luxembourg, Luxembourg, Luxembourg

reformulating it is the aim."[1] Not surprisingly, LIR has been approached within AI & Law primarily with a focus on conceptualization of legal information, while for daily legal work that might not always be the most effective approach.

Meanwhile, due to the advancements of the information era and the Open Data movement the number of legal documents published online is growing exponentially, but accessibility and searchability have not kept pace with this growth rate. Poorly written or relatively unimportant court decisions are available at the click of the mouse, exposing the comforting myth that all results with the same juristic status are equal. An overload of information (particularly if of low-quality) carries the risk of undermining knowledge acquisition possibilities and even access to justice.

Apart from the problems with the quantities, also the qualitative complexities of legal search cannot easily be underestimated. Legal work is an intertwined combination of research, drafting, negotiation, counselling, managing and argumentation (Leckie et al. 1996). To limit the role of LIR within daily legal practice to just finding the court decisions relevant to the case at hand underestimates the complexities of the law and of legal information seeking behaviour. Any legal information retrieval system built without sufficient knowledge, not just of the actual legal information needs but also of the 'juristic mind', is apt to fail. Understanding of information needs and information-seeking behavior of legal professionals seems essential as it helps in the planning, implementation and operation of information system and services in the given work settings (Devadason and Lingam 1997). Legal information-seeking is the behavior displayed by lawyers when using a range of existing legal resources to find information required for their work.

LIR systems have been designed to support legal information-seeking, but without accommodating the characteristics of legal information-seeking behavior (Sutton 1994). If systems designers view legal information-seeking behavior, this might lead to the implementation of mechanisms and systems to support legal information-seeking at each stage of the value adding process (Cole and Kuhlthau 2000).

To aid researchers and system designers in designing or developing LIR applications it might be an interesting exercise to approach LIR more explicitly as a subtype of information retrieval (IR) instead of (merely) a topic within AI & Law. Since 'relevance' is the basic notion in IR, it could be a useful starting point for analysing the specificities of LIR. In this paper we develop a framework for the concept of relevance in legal information retrieval and come forward with suggestions for improvements in LIR systems. We do not intend to present a blueprint for a new legal search engine, nor do we assess LIR systems currently in use. We do discuss some practical examples, but only to illustrate the merits of our theoretical framework. And since we only intend to elaborate the concept of relevance, we refrain from discussing or evaluating algorithms for calculating relevance.

In Sect. 2 we define 'Legal Information Retrieval' by, on the one hand, distinguishing it from Legal Expert Systems and, on the other hand, describing the

---

[1] E. Schweighofer in Bench-Capon et al. (2012).

characteristics that justify its classification as a specific subtype of IR. In Sect. 3 we discuss the concept of relevance in LIR, guided by a topology of six 'dimensions' of relevance. In Sect. 4 we will draw some conclusions and make suggestions for future work.

2 Legal information retrieval

2.1 Inference versus querying

In a variety of ways information technology is working its way into the legal domain and even endangering the livelihood of its inhabitants (Susskind 2013). Out of all these different systems we highlight two types of information systems: legal expert systems (LES) and legal information retrieval (LIR), on the one hand with a view to articulate the particularities of LIR systems and on the other hand to underline the need—at least for many years to come—of LIR for the legal profession. The main aspects of LES and LIR are listed in Table 1.

In research interesting cross-fertilisation experiments started a long way back (Rissland and Daniels 1995) and many of the recent developments within the legal semantic web [as summarized in e.g. (Casanovas et al. 2016)] are also of importance for LIR, but it is highly unlikely that the two types will completely merge. LIR starts where LES isn't able to provide an answer. And notwithstanding the improvements AI & Law brings to LES, there will always be questions left and relevant documents to be discovered, since the lack of any final scheme is inherent to the legal domain.

2.2 Characteristics of legal information

A variety of specific features justify—and compel—the positioning of legal information retrieval as a specific subtype of information retrieval (Turtle 1995). On

Table 1 A comparison between legal expert systems (LES) and legal information retrieval (LIR)

Aspect	LES	LIR
Goal	Establish a legal position on specific case	Provide relevant legal information
Input	Facts	Request
Content	Legal rules encoding the domain expertise	Documents
Method	Inference	Querying
Output	Decision, advice, forecast	Set of documents
Preferred use	Answering 'happy flow' questions within a specific and limited domain	Finding information objects in huge repositories
Advantage	Can provide straightforward answers	Unlimited content, input and output
Drawback	What has not been modelled, cannot be answered	User always has to read, interpret and decide for himself
Basic notion	Uncertainty	Relevance

describing these features, we will briefly elucidate some shortcomings of general IR in meeting the needs arising from the legal domain.

1. *Volume* Although in the age of 'big data' the longstanding impressive volumes of legal materials have been surpassed by e.g. telecommunications and social media data, viewed upon from an information retrieval perspective the volume of legal materials is still impressive. This holds true for public repositories (like case law databases) as well as for private repositories (e.g. case files within law firms or courts).

2. *Document size* Compared to other domains, legal documents tend to be quite long. Although metadata and summaries are often added, access to (and searchability of) the full documents is of paramount importance.

3. *Structure* Legal documents have very specific (internal) structures, which often also are of substantive relevance. Although standards for structuring legal documents are emerging (Palmirani 2012), many legal documents do not have any (computer readable) structure at all.

4. *Heterogeneity of document types* In the legal sphere a variety of document types exist which are hardly seen in other domains. Apart from the obvious legislation and court decisions, one can think of parliamentary documents, contracts, loose-leaf commentaries, case-law notes a.s.o.

5. *Self-contained documents* Contrary to many other domains, documents in the legal domain are not just 'about' the domain, they actually contain the domain itself and hence they have specific authority, depending on the type of document. A statute is not merely a description of what the law is, it constitutes the law itself (Turtle 1995). Notwithstanding the notion that in a bibliographical sense a document is only a manifestation of an abstract work (IFLA 1998), for information retrieval purposes the object to be retrieved embodies the object itself.

6. *Legal hierarchy* The legal domain itself defines a hierarchical organization with regard to the type of documents and their authority. Formal hierarchies depend on the specific jurisdiction or domain, and factual hierarchies often also depend on interpretation, e.g. the general rule *lex specialis derogat legi generalis* requires a decision on its applicability in a specific situation.

7. *Temporal aspects* Within the incessant flow of legislative processes, legislative texts and amendments follow one another and may overlap. Recurrent challenges stem from tracing the history of a specific legal document by searching the temporal axis of its force and efficacy (Araszkiewicz 2014) and by retrieving the applicable law in respect to the timeframes covered by the events subject to regulation (Palmirani and Brighi 2006).

8. *Importance of citations* In most other scientific domains citation indexes exist for academic papers. In the legal domain, citations are a more integral part of text and argumentation: "Legal communication has two principal components: words and citations" (Shapiro 1991). Citations can be internal (cross-references), linking one normative provision to another normative provision in the same document or normative provisions to recitals (Humphreys et al. 2015). Citations can also be external, linking e.g. a court decision to a

normative provision, a normative document to another normative document, or an academic work to a parliamentary report. Citations can be explicit or implicit and they can express a whole variety of different relationships: they can be instrumental (or 'formal')—e.g. a court of appeal referring to the appealed first instance decision—or of a purely substantive nature, but having distinct intensions. Like the structure of legal documents in general, mentioned under (3), most citations are poorly formatted and not computer readable.

9. *Legal terminology* Legal terminology has a rich and very specific vocabulary, characterized by synonymy, ambiguity, polysemy and definitions that are very precise and vague at the same time.

10. *Audience* Legal information is queried by a wide variety of audiences. Laymen with different levels of legal knowledge and jurists with completely different professions. Scholars, judges, lawyers, notaries, library staff or legal aid workers have completely different work roles that influence their information needs (Otike 1999), where we may define 'their information needs' as the "Gap between what we know and what we want to know that motivates a search" (Dervin 1992).

11. *Personal data* Many legal documents contain personal data. Apart from the consequences for the publication of e.g. court decisions, it also weighs on LIR, since the juristic memory is often built on names of persons and places.

12. *Multilingualism and multi-jurisdictionality* In many (scientific) domains English is the pivotal language, and in the legal domain the same goes for common law jurisdictions. Civil law jurisdictions though have a variety of languages; jurisdiction and language have such a strong relationship that translated documents can only be a derivative of the original. As a result, European or international legal information retrieval poses very specific problems.

13. *Scatteredness of legal resources* Legal information is to be found in a variety of resources, scattered in a complex way, with different access regimes, technical formats and interfaces.

3 Relevance within legal search

3.1 Nature of relevance in LIR

The science of information retrieval is basically about 'Relevance': how to retrieve the most relevant documents from—in principle—an unlimited set? Before any methodology or system for retrieval can be developed or discussed, the concept of 'relevance' has to be examined. This seems to be a trivial undertaking since this concept has a tendency to be immediately understood by everybody. A thorough understanding though is of the utmost importance for the effectiveness of LIR systems, and hence it needs continuous consideration. The foundations of a conceptual framework can be adopted from general IR science.

Saracevic (1996) defined 'relevance' as: 'pertaining to the matter at hand', or, more extended: 'As a cognitive notion relevance involves an interactive, dynamic establishment of a relation by inference, with intentions toward a context.' From this definition it follows that relevance has a contextual dependency since it is measured in comparison to the 'matter at hand'. Because of its dynamic establishment relevance may change over time and it involves some kind of selection (Saracevic 2007). From the definition it also follows that relevance is a comparative concept: it is a ratio scale of measurement, although by using a specific threshold it can be turned into a binary property (relevant or not). Because of this comparative character, information objects can be ranked as to their relevance.

Because of its visibility in many end-user LIR applications, 'ranking' might appear to be a crucial concept (Geist 2016), but ranking of search results is only one of the many practical applications of relevance, next to e.g.: 'Filtering, assessing, inferring, (…) accepting, rejecting, associating, classifying… and other similar roles and processes' (Saracevic 1996). By narrowing 'relevance' to 'ranking' one not only excludes these many other applications of relevance—which are also increasingly used in modern LIR systems—but inevitably runs into theoretical problems by mistaking a derivative function for the underlying concept.

3.2 Dimensions of relevance in LIR

To understand the concept of relevance it is important to disambiguate various 'relevance dimensions' (Cosijn and Ingwersen 2000). This term compares to 'relevance manifestations' as used by Saracevic (2007). We discuss these relevance dimensions here in brief, summarizing their basic features and indicating how our typology deviates from those of Saracevic and Cosijn/Ingwersen. Along the paper we will elaborate these relevance dimensions for legal information retrieval in greater detail.

1. *Algorithmic or system relevance* The first dimension pertains to the computational relationship between a query and information objects, based on matching or a similarity between them. Traditionally, models have been described within the context of full-text search, e.g. being Boolean, probabilistic, vector-space a.s.o. Natural language processing is also perceived to be within algorithmic relevance, although in our view it covers also those processes which do not take place during the actual querying, but are intended to improve algorithmic relevance at a later stage. Examples are pre-processing of documents, automatic classification a.s.o. Unlike all other relevance dimensions that can be observed and assessed without a computer, algorithmic relevance cannot: it is system-dependent.

2. *Topical relevance* The relationship between the 'topic' (concept, subject) of a request and the information objects retrieved about that topic. A topicality relation is assumed to be an objective property, independent of any particular user. 'Aboutness' is the traditional distinctive criterion. The topics of the information objects might be hand-coded or computed, e.g. by classification algorithms.

3. *Bibliographic relevance* The relationship between a request and the bibliographic closeness of the information objects. One of the specific features of legal information, as described in Sect. 2.2 above, is its self-containment. This means that legal information systems (unlike information systems on medicine, classic cars or animals) are the final objects themselves. Hence, 'isness' is the distinctive criterion. Because of the many different versions legal information objects might have, isness is not a Boolean but a relative concept, and therefore not an issue of *data* retrieval, but of *information* retrieval. This dimension does not exist in the typologies of Saracevic and Cosijn.

4. *Cognitive relevance or pertinence* Concerns the relation between the information needs of a user and the information objects. Unlike algorithmic, bibliographic and topical relevance, cognitive relevance is user-dependent, with criteria like informativeness, preferences, correspondence and novelty as measuring elements.

5. *Situational relevance or utility* Defined as the relationship between the problem or task of the user and the information objects in the system. Also this dimension of relevance is dependent on the specific user, but unlike the cognitive relevance it does not focus on the request as formulated, but on the underlying motivation for starting the information retrieval process. Inferred criteria for situational relevance are the usefulness for decision-making, appropriateness in problem solving and reduction of uncertainty.

6. *Domain relevance* As his fifth dimension Saracevic (1996) used 'Motivational or affective relevance', but in a critical assessment Cosijn and Ingwersen (2000) replaced this dimension by 'socio-cognitive relevance', which "[I]s measured in terms of the relation between the situation, work task or problem at hand in a given socio-cultural context and the information objects, as perceived by one or several cognitive agents." Given the specific features of legal information as well as for reasons of modelling, we define this dimension as the relevance of information objects within the legal domain itself (and hence not to 'work task or problem at hand'). For convenience we label it 'domain relevance'.

The role of these dimensions in the interplay between user, information retrieval system and legal domain is depicted in Fig. 1.[2] It should be noted that both bibliographic and topical relevance relate to a relationship between the user request (as formulated in the user interface) and the information objects. They might be mutually exclusive—the user is either looking for the objects itself, or information about it—but not necessarily: one might search for a court decision and information about that decision at the same time, but even then the user wants these results separately or recognizable as 'is' and 'about' in his result list.

Already here it should be observed that relevance dimensions easily overlap and intermingle: "The effectiveness of IR depends on the effectiveness of the interplay and adaptation of various relevance manifestations, organized in a system of relevancies" (Saracevic 1996). In the design of IR systems it is hence of the utmost importance to distinguish between various dimensions and to pay specific attention
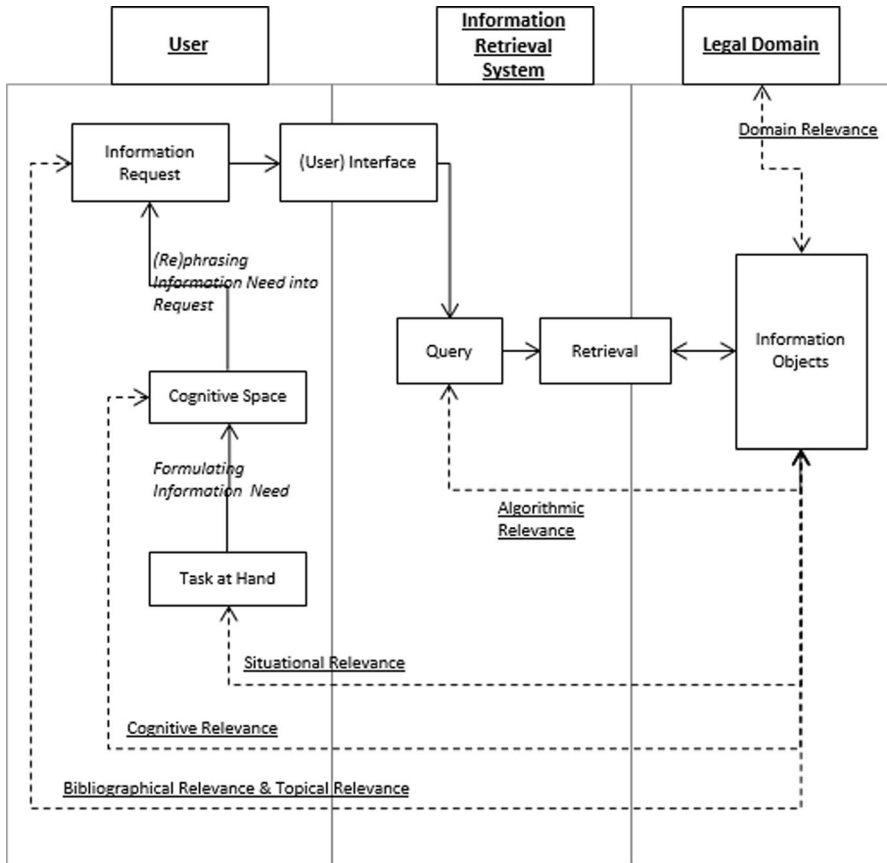
---

[2] Inspired by Cosijn and Bothma (2005).

Fig. 1 Interplay between user, information retrieval system and legal domain

to each of them, in the user interface, the retrieval engine and the document collection. It will definitely improve the user's perception of the system's performance on retrieving the most relevant information. This perception—or 'criterion for success'—depends on the relevance dimension(s) invoked. These criteria are, together with the nature of the respective dimensions, summarized in Table 2.

In the following subsections we will elaborate these six relevance dimensions of LIR and discuss how these dimensions may help to classify the past and current spectrum of approaches, how they correspond to information-seeking behaviour of legal professionals and how they might help bridging the conceptual gap between lawyers and informaticians.

3.2.1 Algorithmic relevance

Algorithmic relevance concerns the computational core of information retrieval. As expressed in Fig. 1 it is the relation between the information objects and the query;

Table 2 Dimensions of relevance compared

Dimension	Describes a relation between	Criterion for 'success'
Algorithmic relevance	Query and information objects	Comparative effectiveness in inferring relevance
Bibliographical relevance	Bibliographical object expressed in the request and information objects	Isness
Topical relevance	Topic expressed in the request and information objects	Aboutness
Cognitive relevance	Information needs of the user and information objects	Cognitive correspondence, information quality, authoritativeness, informativeness
Situational relevance	Situation/task at hand and information objects	Usefulness in decision-making and problem-solving
Domain relevance	Opinion of the legal crowd and information objects	Legal importance

this 'query' is to be understood as the computer processable translation of the request as entered in the user interface or any other intermediary component. Algorithmic relevance is about the capability of the engine to retrieve a given set of information objects (the 'gold standard') that should be retrieved with a given query (measured in 'recall') with a minimum of false positives (measured in 'precision').

From our conceptual perspective the type of query as well as the type of retrieval framework is not relevant, but given the legal information features of volume, document size and lack of structure, textual search has for long had the focus. There are various computational models for inferring similarity between query and information objects. In the early days Boolean search was the core of any legal retrieval system, and it is still an indispensable element in most LIR systems today. In a Boolean system both the user request and the documents are regarded as a set of terms, and the system will return documents containing the terms in the request. Boolean searches often result in the retrieval of a large number of documents. In addition, they provide little or no assistance to the user in formulating or refining a query and they lack domain expertise that could improve the search outcome. Relevance performance was improved by using models as the vector space model (Salton et al. 1975) and TF-IDF (term frequency—inverse document frequency). Nevertheless, recall is often below acceptable levels because the design of full-text retrieval systems: "(I)s based on the assumption that it is a simple matter for users to foresee the exact words and phrases that will be used in the documents they will find useful, and only in those documents" (Blair and Maron 1985). Ambiguity, synonymy and complexity of legal expressions contribute substantially to this problem (Dabney 1986). Natural language processing (NLP) is gaining popularity as an addition to or alternative to pure text-based search (Maxwell and Schafer 2008).

Apart from text-based search also other types of algorithmic relevance can be considered, like the use of ontologies as higher level knowledge models (Casanovas et al. 2016; Saravanan et al. 2009), network statistics, especially when used for citation analysis (Fowler and Jeon 2008; van Opijnen 2013) as well as methods that combine different approaches (Koniaris et al. 2016).

3.2.2 Topical relevance

Topical relevance is about the relevancy relation between the topic as (explicitly or implicitly) formulated in the user request and the topics of the information objects. Different strategies have been explored to improve this relevance dimension.

1. *Mapping and indexing terms* Using free text search and mapping the terms searched to the terms indexed from the information objects, too often renders poor results since legal concepts can be expressed in a variety of ways, while completely different concepts can textually be quite similar.

2. *Manual indexing* Adding head notes and keywords from taxonomies or thesauri has been a long tradition within the legal information industry. Kuhlthau and Tama (2001) pointed to the lack of flexibility within such keyword search, as they noted that "(L)awyers seemed to require the opportunity to locate information outside the keyword range in order to spark an idea that enabled them to formulate the issues in a case." This approach is problematic when lawyers have few or imprecise details about the area of which an overview is required. Although *aboutness* is assumed to be an objective property and hence independent of any particular user, manual indexing is inherently subjective, and even the same indexer may sort the same document under different terms depending on the context the document is presented in (Bing and Harvold 1977). "Manual indexing is only as good as the ability of the indexer to anticipate questions to which the indexed document might be found relevant. It is limited by the quality of its thesaurus. It is necessarily precoordinated and is thus also limited in its depth. Finally, like any human enterprise, it is not always done as well as it might be." (Dabney 1986).

3. *Semi-automated classification* For huge public databases manual tagging is hardly an option, but automated classification turns out not to perform better than human indexing (Mart 2010). A general drawback of such automated systems is the mandatory use of the classification scheme in the user interface. This forces the user to limit or to reformulate his request to align it with the classification system available. A problem that can only be solved by the time-consuming and tedious task of "Using a combination of automated and manual techniques, [constructing] a list of concepts and variations for expressing a concept" (Zhang 2015). This requires in-depth legal knowledge, analysis of search engine log files and continuous maintenance. Semi-automated classification using ontologies (Boella et al. 2016) is gaining popularity, and notwithstanding the current hype about legal AI applications like IBM's Ross (Beck 2014), scepticism about their performance seems to be a healthy attitude (Paliwala 2016; Remus and Levy 2016).

4. *Relation-based search* Meanwhile, developers of LIR systems should consider whether the investment is worth the effort: surveys have shown that classification systems are not very popular among users (Peoples 2005), contrary to searches by relationship (Lastres 2015). Many topics in law, at least in the juristic mindset and information seeking behaviour, have a strong connection (*chain*) to other legal documents. Typical requests may refer to a

search for (everything) about a specific paragraph of law or court decision. In such requests these information objects represent a specific legal concept, but the only reason lawyers rephrase it might be related to the fact that the search engine cannot cope with their actual request. For well-known acts and codes such aboutness information is structured in treatises or loose-leaf encyclopaedias, but they are optimized for browsing, not for search. Since such works do not cover the whole legal domain, performing searches on citations might in principle be the obvious choice. In common law countries citators are very popular for such 'topical citation search', like LawCite.org (Mowbray et al. 2016) in the public domain and Shepard's in the private domain (Spriggs and Hansford 2000). The latter is based on manual tagging and also contains qualifications of these relations. In continental Europe the importance of search by citation—as a type of aboutness—needs more attention from search providers. For example, in EUR-Lex, HUDOC and various national legislative databases, relations between documents are tagged and searchable/browsable, but especially in national case law databases citation search is extremely difficult. A first reason is that judges have lousy citation habits: research showed that only 36% of cited EU acts was in conformity with the prescribed citation style, the other citations were made with a wide range of other styles (van Opijnen 2010b). Comparable problems appear when searching for case law citations, where additional complexity is added by the fact that one decision can be cited by many different identifiers (van Opijnen 2010a), like—often ambiguous—case numbers, reporter codes, commercial references or judgment identifiers like the Europe Case Law Identifier (ECLI)[3] (van Opijnen and Ivantchev 2015). Case names—often containing the names of the parties to the case—are problematic since they have many different spelling variants and are less frequently used since court decisions are anonymized more often (van Opijnen 2016a). Moreover, slashes, commas and hyphens are essential elements of legal identifiers, but are out-of-the-box interpreted by search engines as specific search instructions (e.g. '/' means 'near' and '–' means 'not'). Manual tagging for large scale public databases is undoable, so reference parsers have to be developed (Agnoloni and Bacci 2016; van Opijnen et al. 2015); as explained in Sect. 3.2.3 they can be used for recognizing the citations in the information objects as well as for understanding user requests.

Search in multilingual legal repositories—e.g. in the ECLI Search Engine on the European e-Justice portal[4]—poses additional problems: the terms used in the request do not only have to be translated into the language of the information objects, but also into the specific legal terminology of the jurisdiction the information objects are about. Various building blocks to tackle this have been developed. EuroVoc[5] is a large multilingual vocabulary; although it is used for

---

[3] Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law, CELEX:52011XG0429(01).

[4] https://e-justice.europa.eu/content_ecli_search_engine-430-en.do.

[5] www.eurovoc.europa.eu.

tagging in the EUR-Lex database, it is too much policy-oriented and too little legal to be of practical use for LIR. Aligning legal vocabularies of different legal systems and/or languages has proven to be quite difficult (Francesconi and Peruginelli 2010); within the Legivoc project various national legal vocabularies have been mapped (Vibert et al. 2013), but it needs more elaboration to be of practical use.

3.2.3 Bibliographical relevance

Topical relevance, as discussed in the previous subsection, is about the relevancy relation between the topic as formulated in the user request and the topic of the information objects. For most information retrieval systems this topicality suffices to measure whether the documents retrieved match the information request as formulated by the user: 'aboutness' is used as the decisive criterion. But contrary to the information contained in many general information (retrieval) systems, the information in legal information (retrieval) systems is highly self-contained. Information retrieval systems on animals, aeroplanes or people contain information about those topics, but not the objects themselves. However, legal information retrieval systems do contain legislation, court decisions and parliamentary documents themselves—notwithstanding the fact they might also contain other documents *about* these objects (which might also be such legal sources themselves). The distinctive criterion for establishing this *bibliographical relevance* is 'isness': the degree to which the documents retrieved actually *are* those requested by the user. Probably because most academic research on information retrieval is about non-self-contained domains, bibliographical relevance is not considered to be a relevance dimension of its own [compare e.g. (Cosijn and Ingwersen 2000; Saracevic 1996)]. Contrary to topical queries or browsing, which are intended for surveying the unknown, bibliographical queries are intended for searching the known, at least from the user perspective: a specific act, court case, parliamentary document or scholarly article. Although this might look like an issue for *data* retrieval instead of *information* retrieval (Baeza-Yates and Ribeiro-Neto 1999) and hence a no-brainer (Harvold 2008), for various reasons in most legal information systems it is still a real brainteaser, and hence it is defendable to approach this as an information retrieval issue.

1. *The ontological Levels of FRBR*
   Before we elaborate this proposition, we first have to introduce the ontological topology developed within the functional requirements for bibliographical records (FRBR) of the International Federation of Library Associations and Institutions (IFLA 1998), which is also widely used for structuring, describing and identifying legal information (Boer 2009; CEN 2010). The four distinctive ontological levels of FRBR are work, expression, manifestation and item.
   The *work* is an abstract level, defined as: 'a distinct intellectual or artistic creation'. For e.g. a court decision the *work* is the judicial decision resolving the specific legal dispute brought before the court. This work level is addressed when one says: "The Google Spain decision of the Court of Justice of the European Union is a landmark decision in the realm of data protection."

The *expression* is also an abstract level, defined as: 'the intellectual or artistic realization of a work'. Note that the expression is also an intellectual or artistic product, but that it is always derived from a work. For legal documents different types of expressions exist: linguistic, temporal and editorial. Temporal expressions are especially relevant for legislation, since the law changes continuously. Editorial expressions are generally more relevant for court decisions: the authentic version of the judge, the anonymized version published on the court's web portal or an abridged expression edited by a legal publisher. The *manifestation* is a (specific) physical embodiment of an expression of a work. Printed documents, PDF-, XML- or Word versions are examples of manifestations. Apart from its non-abstract character the *manifestation* also lacks the intellectual or artistic effort to have it created.

Finally, the *item* is the single exemplar of a manifestation. It could e.g. be the digitally signed PDF version of a court decision residing in a specific directory on my computer or the most recent hardcover version of the Lithuanian criminal code lying on my desk.

2. *The FRBR Problem*

Bibliographical relevance poses three interrelated problems to retrieval systems, all of them supporting our proposition that this is in the realm of information retrieval and not in the domain of data retrieval. The first hurdle is in understanding whether the user poses an '*is* request' or an '*about* request'; the second issue is the identification problem and the third challenge is about retrieving the correct FRBR version(s) of a legal information object.

As to the first problem, information retrieval systems operating within non-self-contained domains can interpret a user request, written in natural language, always as an *about* request. They can process the request with the optimizations described in Sect. 3.2.1 on algorithmic relevance, but if asked 'Jaguar E-type' the system can be sure the user expects descriptions, pictures and manuals of the iconic car to be retrieved, but not the thing itself. But when asked for 'Dublin Regulation' the system must be able to understand that this might be a request for documents containing the two words, or for legal provisions applying to the Irish capital, but that first and foremost it must be understood as a request for the text of 'Regulation (EU) No 604/2013 of the European Parliament and of the Council of 26 June 2013 establishing the criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third-country national or a stateless person',[6] in which title the word 'Dublin' does not appear at all.

The second problem surfaces when one realizes that lawyers are not that precise in citing legal sources, and hence in formulating their search requests. The abovementioned regulation might also be cited as e.g. 'Regulation No 604-2013', 'EC-reg. No 604/2013' or 'Reg (EEC) 604.2013'. All of these styles are not compliant with the EU interinstitutional style guide (EU Publications Office 2011) or even incorrect, but when used in a citation they

---

[6] CELEX:32013R0604.

will be understood immediately by any legal professional. When used in a search engine though they will not lead to the desired result. For the reasons already discussed under *relation-based search* of Sect. 3.2.2, punctuation marks are interpreted as specific query instructions and the tens of different formatting variants are too difficult to be interpreted correctly during query execution.

For this reason, as well as to understand that a user is actually searching a legal document and not performing a topical search, many legal information retrieval systems offer a complex search screen, enabling the user to specify his request very precisely as to the title of the document, its (often compound) document identifier, publication reference, document date or abbreviation. The fact that such detailed screens are often offered as the default search mode or at least very prominently advertised, underlines the importance of bibliographical searches: such forms are still needed to achieve an acceptable performance on the *isness* criterion. At the same time though the existence of many different labels for a wide variety of identifiers and metadata with a lot of variations between the many legal information retrieval systems a user has at its disposal nowadays is a serious threat to findability of documents and hence to the usability of these systems. This problem is often multiplied by changes in identification systems or citation habits. An example can be drawn from the EUR-Lex advanced search—where one has to split the document number into a 'year' part and a 'number' part—even a trained user can be puzzled where to put which digits in case he is looking for 'Directive 96/95/EC', 'Regulation 98/2014' or 'Regulation 2015/2016'.[7]

One could say in general that such 'advanced' search forms for finding specific legal documents are too strict, while also here the adagium "Be lenient in what you accept and strict in what you produce" (Musciano and Kennedy 2006) should apply. Reference parsers that have been developed for detecting citations in documents themselves (van Opijnen et al. 2015)[8] may also be used for pre-parsing user requests, making obsolete most of all those specific input fields.

Even if the LIR system understands that isness will be the evaluation criterion and not aboutness, and even if it also understands which information object(s) might be requested for, it is confronted with the third problem: which FRBR version(s) of the document should be presented to the user. There is no clear-cut answer, but some aspects have to be taken into account. First, there might be a problem of ambiguity at the work level. Above, the Dublin Regulation was mentioned as an example, stating it is an alias for Regulation (EU) No 604/2013, but although this alias is used in daily legal language, it is not unambiguous. More precisely, this regulation is dubbed 'Dublin III Regulation', its predecessor, Regulation (EC) No 343/2003,[9] being the 'Dublin

---

[7] The year is 1996, 2014 and 2015 respectively. In a directive the year comes first, in a regulation the number comes first. But from 1 January 2015 onwards the year comes first in all acts www.eur-lex. europa.eu/content/tools/elaw/OA0614022END.pdf.

[8] Below, Sect. 3.2.6.

[9] Council Regulation (EC) No 343/2003 of 18 February 2003 establishing the criteria and mechanisms for determining the Member State responsible for examining an asylum application lodged in one of the Member States by a third-country national, CELEX:32003R0343.

II Regulation', which in turn was preceded by the Dublin Convention[10] (the namegiver of the legal doctrine all these instruments are about). Because of the amendments already made to the Dublin II Regulation by Regulation (EC) No 1103/2008[11] and additional changes that had to be made, it was decided the regulation had to be recast, making the Dublin III Regulation actually a distinct temporal expression of the same *work* ('Dublin Regulation') as the temporal expression Dublin II.[12] For Dublin II there is the promulgated expression (published in the Official Journal), the first consolidated expression,[13] and the consolidation after its amendment in 2008. Also Dublin III exists in its promulgated expression in the Official Journal, as well as in a consolidated expression.[14] EU regulations are equally authentic in all official (24)[15] languages, and most of these language expressions exist for all temporal and promulgated/consolidated expressions. And with regard to temporal expressions, also (possible) future versions should, if available, be retrievable.[16] Many of such documents exists in different manifestations; for end-users often (X)HTML and PDF are available, for computers sometimes also e.g. (RDF/)XML or JSON.

The problem of finding and presenting the bibliographically most relevant version can be addressed by a variety of methods., e.g. taking into account the language of the user, using the metadata (e.g. on the provider of the document and its authoritativeness), offering an option for specifying the temporal expression in the request form, or the possibility to compare different linguistic or temporal expressions after a first version of the document has been retrieved. An example of the former can be found on EUR-Lex, which can now display up

---

[10] Convention determining the State responsible for examining applications for asylum lodged in one of the Member States of the European Communities, CELEX:41997A0819(01).

[11] Regulation (EC) No 1103/2008 of the European Parliament and of the Council of 22 October 2008 adapting a number of instruments subject to the procedure laid down in Article 251 of the Treaty to Council Decision 1999/468/EC, with regard to the regulatory procedure with scrutiny—Adaptation to the regulatory procedure with scrutiny—Part Three, CELEX:32008R1103.

[12] It should be borne in mind that opinions differ on the question what actually constitutes a new work. Within the ELI framework (see footnote 21) both regulations are works on their own (ELI Task Force 2015), and labeled 'eli:LegalResource'.

[13] In the ELI framework promulgated version and consolidated version are considered to be separate works, see footnote 12.

[14] Although the act 'Dublin III' has not been amended yet, the first consolidated version is generally regarded as a separate expression.

[15] Sometimes 23, the Gaelic version does not exist for all documents.

[16] For the Dublin Regulation e.g. a future consolidated version could come into being after adoption of the 'Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EU) No 604/2013 as regards determining the Member State responsible for examining the application for international protection of unaccompanied minors with no family member, sibling or relative legally present in a Member State' (CELEX:52014PC0382) and/or the 'Proposal for a Regulation of the European Parliament and of the Council establishing a crisis relocation mechanism and amending Regulation (EU) No 604/2013 of the European Parliament and of the Council of 26 June 2013 establishing the criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third country national or a stateless person' (CELEX:52015PC0450).

to three language versions at the same time.[17] Also time-travelling in legislative databases is improving: jurists often need to know the delta between the temporal version T of an act and version T + 1. Some legislative databases nowadays not only serve version T and $T + 1$ in parallel, but also actually show the delta in a user-friendly way.[18] On the server-side, specific 'FRBR resolvers' like the Akoma Ntoso resolver might be of aid for finding the best match for a given set of input parameters, even if this best match is on distinct server (Palmirani et al. 2014).

3.2.4 Cognitive relevance

Cognitive relevance concerns the extent to which an information object matches the cognitive information needs of the user: the information needs as he experienced them before he had to translate them into a request in the user interface. This relevance dimension is of a subjective nature: do the retrieved documents fit to the user's state of knowledge? Are there any characteristics regarding the information objects retrieved he should be aware of?

Since this dimension is of a subjective nature, the cognitive relevance performance of a LIR system depends on the ability of the system to explicitly or implicitly understand the information needs of each individual user; the many contexts in which the term 'personalized search' is used all have in common that they are about cognitive relevance.

Especially the possible use of recommender systems should be mentioned here. Recommender systems rely on intelligent filtering by comparing and combining document metrics, search results and user-generated data. Two types of filtering can be distinguished. 'Collaborative filtering' recommends documents by making use of the user's past search behaviour and/or that of a peer group. 'Content-based filtering' on the other hand uses shared features of the document at hand and other documents, based on e.g. topical resemblance, comparable metadata or closeness in a citation network. Of course, collaborative filtering and content-based filtering can also be combined. Recommender mechanisms can be used to limit the number of documents retrieved (e.g. because the system knows a given user is only interested in tax law and not in criminal law) or to increase the number of documents: by offering 'more like this' buttons or navigable citation graphs users can be supported in serendipitous information discovery (Toms 2000). Being tailored to the individual needs of the user, recommender system can also be used for pro-active search: notification systems informing a user about information objects that have been added to the repository and might be of interest for him, because he explicitly expressed the wish to be informed about data with those specific characteristics, or because the system reaches this conclusion based on past search behaviour. Within

---

[17] E.g. www.eur-lex.europa.eu/legal-content/EN-CS-ET/TXT/?uri=CELEX:32013R1316, showing the English, Czech and Estonian version of Dublin III Regulation in one screen.

[18] E.g.   www.wetten.overheid.nl/BWBR0006368/2016-01-01?VergelijkMet=BWBR0006368%3fg%3d2010-02-01.

legal information retrieval recommender systems have not had too much attention yet (Boer and Winkels 2016; Winkels et al. 2014).

3.2.5 Situational relevance

While cognitive relevance is associated with *search* task execution, situational relevance pertains to *work* task execution; the relevance of documents is measured by their usefulness for the task at hand, e.g. decision-making or problem-solving (Cosijn and Bothma 2005). "The judgement of situational relevance embraces not only the user's evaluation of whether a given information object is capable of satisfying the information need, it offers also the potential of creating new knowledge which may motivate change in the decision maker's cognitive structures. The change may further lead to a modification of the perception of the situation and the succeeding relevance judgement, and in an update of the information need." (Borlund 2000) It should be noted that the system is not asked to solve the problem itself—then it would be a legal expert system, not a legal information retrieval system.

Situational relevance in legal information retrieval comes close to—but should not be confused with—'legal relevance', which usually means that information is relevant to a proposition when it affects, positively or negatively, the probability that the proposition is true (Cross and Wilkins 1964).[19]

The difference between 'legal relevance' and situational relevance can be understood with the help of the following definition by Jon Bing:

A legal source is relevant if:

1. The argument of the user would have been different if the user did not have any knowledge of the source, i.e. at least one argument must be derived from the source; or
2. legal meta-norms require that the user considers whether the source belongs to category (1); or
3. the user himself deems it appropriate to consider whether the source belongs to category (1). (Bing 1991)

In this definition (1) pertains to the strict notion of 'legal relevance', while situational relevance in legal information retrieval also covers (2) and (3).

Probably because of the relative importance of case law in the United States and other common law countries, much LIR research has concentrated on finding the (most) relevant court decisions relating to a case at hand. This can be pursued using a variety of (sometimes combined) technologies, like argumentation mining (Mochales and Moens 2011) and natural language processing (NLP) (Maxwell and Schafer 2008).

---

[19] Next to this 'logical' or 'probablistic' definition often also a 'practical' concept is used, meaning 'worth hearing' (Woods 2010).

3.2.6 Domain relevance

We defined 'domain relevance' as the relevance of information objects within the legal domain itself. It is independent from any information system and independent from any user request. As can be understood from the previous paragraph we prefer to avoid the term 'legal relevance', but 'legal importance' is safe to use as a synonym for 'domain relevance within the legal domain' (van Opijnen 2016b).

Domain relevance can be applied in LIR systems in different ways.

1. *Legal importance of classes of information objects.*
   This concerns categories of information objects that can be classified as to their legal importance: constitutions outweigh ordinary acts, which in turn are more important than by-laws or ministerial degrees. In a comparable way, opinions of supreme courts can be expected to have more authority than district court verdicts, but in turn are surpassed by decisions of the European Court of Human Rights. For many categories of information objects their relative legal importance can be derived from basic metadata.
2. *Legal importance of individual information objects.*
   The concept of domain relevance can be used to classify individual information objects as to their legal importance as well. In vast repositories, separating the wheat from the chaff has for long been the territory of domain experts: as publication/storage was expensive, and adding documents itself labour-intensive, a selection was made on the input side of any paper or early digital repository. The ease with which information can be published on the internet nowadays has shifted the selection process—at least partially—from the input side to the output side: 'selection' has evolved from a publisher's issue into a challenge for information retrieval. Case law publication in the Netherlands could serve as an illustration: the public case law database in the Netherlands[20] contains a small percentage ($<1\%$) of decided cases, but in 15 years has accumulated 370,000 documents. More than 75% of those is not considered important enough to be published in legal magazines (van Opijnen 2014).
   An example of domain relevance applied at the document level can be observed in the HUDOC database, containing all case law documents produced by the European Court of Human Rights. To aid the user in filtering the nearly 57,000 documents as to their legal authority, four importance levels have been introduced. Except for the highest category, containing all judgments published in the Court Reports, all documents have been tagged manually. Since this importance level is an attribute of each individual document, it can easily be used in combination with other relevance dimensions.
   Since manual tagging is labour-intensive, for more massive repositories a computer-aided rating is indispensable. Given the abundant use of citations between court decisions, network analysis is an obvious methodology to assess case law authority (Fowler and Jeon 2008; Winkels et al. 2011). In the 'Model for Automated Rating of Case law' (van Opijnen 2013) the 'legal crowd'—the

---

[20] www.uitspraken.rechtspraak.nl.

domain specialists that rate individual court decisions as to their importance by citing them or not—is extended to legal scholars, while it also uses other variables within a regression analysis to predict the odds of a decision rendered today for being cited in the future. One of these variables is the changing perceptions over time regarding the importance of a singular court decision [see e.g. also (Tarissan and Nollez-Goldbach 2015)]. If court decisions are well-structured and citations are made to the paragraph level, importance can be calculated for the sub-document level as well (Panagis and Šadl 2015). Comparable techniques can be used for the relevance classification of legislative documents (Mazzega et al. 2009).

Network analysis is supported by the use of common identifiers, like the European Legislation Identifier (ELI),[21] the European Case Law Identifier (ECLI) and possibly in the future a European Legal Doctrine Identifier (ELDI) (van Opijnen 2017) or a global standard for legal citations.[22]

Apart from establishing the bare relationship between legal information objects as can be derived from citations, added value can be created by establishing and assessing the nature of the relationship. Shepard's citations (Spriggs and Hansford 2000) offers an example, but it is only available on subscription and since the classification itself is done manually large public datasets need automated solutions (Winkels et al. 2014).

4 Conclusions and further work

Relevance, the basic notion of information retrieval "Is a thoroughly human notion and as all human notions, it is somewhat messy." (Saracevic 2007) As upheld in this paper, 'relevance' within legal information retrieval deserves specific attention, due to rapidly growing repositories, the distinct features of legal information objects and the complicated tasks of legal professionals.

Because most LIR systems are designed by retrieval specialists without comprehensive domain knowledge, sometimes assisted by domain specialists with too little knowledge of retrieval technology, users are often disappointed by their relevance performance.

Four main conclusions can be highlighted. First of all, retrieval engineering is focused too exclusively on algorithmic relevance, but it has been proven sufficiently that without domain specific adaptations every search engine will disappoint legal users. By unravelling the holistic concept of 'relevance' we hope to stimulate a more comprehensive debate on LIR system design. All dimensions of relevance have to be considered explicitly while designing all components of LIR systems: document pre-processing, (meta)data modelling, query building, retrieval engine and user interface. Within the user interface, legal information seeking behaviour,

---

[21] Council conclusions inviting the introduction of the European Legislation Identifier (ELI), CELEX: 52012XG1026(01).

[22] LegalCiteM: www.oasis-open.org/committees/legalcitem/.

including searching, chaining, filtering and browsing should take full advantage of the various relevance dimensions, of course in a way that fits the legal mindset and acknowledging that relevance dimensions are continuously interacting in the process of information search.

Secondly, the 'isness' concept is overlooked too often. Finding (the expressions of) a work is—and not (just) the related works—is an often-used functionality for jurists, but misunderstood by system developers.

Thirdly, also domain relevance is an underdeveloped concept. While there is a tendency to publish ever more legal information, especially court decisions, without tagging it as to its juristic value, information overkill will become a serious threat to the accessibility of such databases. Performance on other relevance dimensions will suffer if the problem of domain relevance isn't tackled sufficiently.

Finally, given the importance of digital information for legal professionals—lawyers easily spend up to 15 h per week on search, most of it in electronic resources (Lastres 2015) although the abandonment of paper does not always seem to be a voluntary choice (Kuhlthau and Tama 2001)—the gap between LIR systems and user needs is still substantial. For a full understanding of their search needs just taking stock of their wishes is not going to suffice, since legal professionals are not capable of describing the features of a system that does not yet exist. To understand the juristic mindset, it is of the utmost importance to follow meticulously their day-to-day retrieval quests. It will for sure reveal interesting insights that can be used to improve the relevance performance of LIR systems.

References

Agnoloni T, Bacci L (2016) Linking data. Analysis and existing solutions. www.bo-ecli.eu/uploads/deliverables/DeliverableWS2-D1.pdf. Accessed 5 Feb 2017

Araszkiewicz M (2014) Time, trust and normative change. On certain sources of complexity on judicial decision-making. In: Casanovas P, Pagallo U, Palmirani M, Sartor G (eds) AI approaches to the complexity of legal systems: AICOL 2013. Springer, Berlin, pp 100–114

Baeza-Yates R, Ribeiro-Neto B (1999) Modern information retrieval. Pearson Education, Essex

Beck S (2014) The future of law. The American Lawyer, New York

Bench-Capon T et al (2012) A history of AI and law in 50 papers: 25 years of the international conference on AI and law. Artif Intell Law 20:215–319. doi:10.1007/s10506-012-9131-x

Bing J (1991) Handbook of legal information retrieval. Norwegian Research Center for Computers and Law, Oslo

Bing J, Harvold T (1977) Legal decisions and information systems. Universitets Forlaget, Oslo

Blair DC, Maron ME (1985) An evaluation of retrieval effectiveness for a full-text document-retrieval system. Commun ACM 28:289–299. doi:10.1145/3166.3197

Boella G, Di Caro L, Humphreys L, Robaldo L, Rossi P, van der Torre L (2016) Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. Artifici Intell Law 24:245–283

Boer AWF (2009) Legal theory, sources of law and the semantic web. University of Amsterdam, Amsterdam

Boer A, Winkels R (2016) Making a cold start in legal recommendation: an experiment. Paper presented at the legal knowledge and information systems—JURIX 2016: The twenty-ninth annual conference, Nice

Borlund P (2000) Evaluation of interactive information retrieval systems. Doctoral thesis, Aalborg

Casanovas P, Palmirani M, Peroni S, van Engers T, Vitali F (2016) Special issue on the semantic web for the legal domain guest editors' editorial: the next step. Semant Web J 7:213–227

CEN (2010) CEN workshop agreement Metalex (open XML interchange format for legal and legislative resources) vol CWA 15710:2010 E. CEN, Brussels

Cole C, Kuhlthau C (2000) Information and information-seeking of novice versus expert lawyers: how experts add value. New Rev Inf Behav Res 1:103–115

Cosijn E, Bothma T (2005) Contexts of relevance for information retrieval system design. Paper presented at the proceedings of the 5th international conference on context. Conceptions of Library and Information Sciences, Glasgow

Cosijn E, Ingwersen P (2000) Dimensions of relevance. Inf Process Manag 36:533–550

Cross R, Wilkins N (1964) An outline of the law of evidence. Butterworths, London

Dabney DP (1986) The curse of thamus: an analysis of full-text legal document retrieval. Law Libr J 78:5–40

Dervin B (1992) From the mind's eye of the user: the sense-making qualitative–quantitative methodology. In: Glazier JD, Powell RR (eds) Qualitative research in information management. Libraries Unlimited, Englewood, pp 61–84

Devadason FJ, Lingam PP (1997) A methodology for the identification of information needs of users. IFLA J 23:41–51

EU Publications Office (2011) Interinstitutional style guide. Publications Office, Luxemburg

ELI Task Force (2015) ELI a technical implementation guide. Publications Office. doi:10.2830/935692, 10.2830/74251, 10.2830/90802

Fowler JH, Jeon S (2008) The authority of Supreme Court precedent. Soc Netw 30:16–30. doi:10.1016/j.socnet.2007.05.001

Francesconi E, Peruginelli G (2010) Semantic interoperability among thesauri: a challenge in the multicultural legal domain. In: Abramowicz W, Tolksdorf R, Węcel K (eds) Business information systems workshops: BIS 2010 international workshops, Berlin, Germany, May 3–5, 2010. Springer, Berlin, Heidelberg, pp 280–291. doi:10.1007/978-3-642-15402-7_34

Geist A (2016) Rechtsdatenbanken und Relevanzsortierung. Dissertation, Universität Wien

Harvold T (2008) Is searching the best way to retrieve legal documents? Paper presented at the e-Stockholm'08 legal conference, Stockholm, 14/19-11-2008

Humphreys L, Santos C, Caro LD, Boella G, Torre Lvd, Robaldo L (2015) Mapping recitals to normative provisions in EU legislation to assist legal interpretation. Paper presented at the legal knowledge and information systems—JURIX 2015: the twenty-eighth annual conference, Braga, 10/11-12-2015

IFLA (1998) Functional requirements for bibliographic records. UBCIM Publications—New Series, vol 19

Koniaris M, Anagnostopoulos I, Vassiliou Y (2016) Multi-dimension diversification in legal information retrieval. Paper presented at the web information systems engineering—WISE 2016, 17th international conference, Shanghai, China

Kuhlthau CC, Tama SL (2001) Information search process of lawyers: a call for 'just for me' information services. J Doc 57:25–43

Lastres SA (2015) Rebooting legal research in a digital age. https://www.lexisnexis.com/documents/pdf/20130806061418_large.pdf

Leckie G, Pettigrew K, Sylvain C (1996) Modelling the information-seeking of professionals: a general model derived from research on engineers. Health Care Prof Lawyers Libr Q 66:161–193

Mart SN (2010) The relevance of results generated by human indexing and computer algorithms: a study of West's Headnotes and Key Numbers and Lexis Nexis's Headnotes and Topics. Law Libr J 102:221–249

Maxwell KT, Schafer B (2008) Concept and context in legal information retrieval. Paper presented at the legal knowledge and information systems—JURIX 2008: The twenty-first annual conference, Florence

Mazzega P, Bourcier D, Boulet R (2009) The network of French legal codes. Paper presented at the 12th international conference on artificial intelligence and law, New York

Mochales R, Moens M-F (2011) Argumentation mining. Artif Intell Law 19:1–22

Mowbray A, Chung P, Greenleaf G (2016) A free access, automated law citator with international scope: the LawCite project. UNSW law research paper no 2016-32

Musciano C, Kennedy B (2006) HTML & XHTML—the definitive guide. O'Reilly

Otike J (1999) The information needs and seeking habits of lawyers in England: a pilot study. Int Inf Libr Rev 31:19–39. doi:10.1080/10572317.1999.10762486

Paliwala A (2016) Rediscovering artificial intelligence and law: an inadequate jurisprudence? Int Rev Law Comput Technol 30:107–114. doi:10.1080/13600869.2016.1229651

Palmirani M (2012) Legislative XML: principles and technical tools. ARACNE, Rome

Palmirani M, Brighi R (2006) Time model for managing the dynamic of normative system. In: Wimmer M, Scholl H, Grönlund Å, Andersen K (eds) Electronic government; lecture notes in computer science, vol 4084. Lecture notes in computer science. Springer, Heidelberg, pp 207–218. doi:10.1007/11823100_19

Palmirani M, Vitali F, Bernasconi A, Gambazzi L (2014) Swiss federal publication workflow with Akoma Ntoso. Paper presented at the 27th international conference on legal knowledge and information systems (JURIX 2014), Krakow

Panagis Y, Šadl U (2015) The Force of EU case law: a multidimensional study of case citations. Paper presented at the legal knowledge and information systems—JURIX 2015: the twenty-eighth annual conference, Braga, 10/11-12-2015

Peoples LF (2005) The death of the digest and the pitfalls of electronic research: what is the modern legal researcher to do? Law Libr J 97:661–679

Remus D, Levy FS (2016) Can robots be lawyers? Computers, lawyers, and the practice of law. doi:10.2139/ssrn.2701092

Rissland EL, Daniels JJ (1995) A hybrid CBR-IR approach to legal information retrieval. Paper presented at the proceedings of the 5th international conference on artificial intelligence and law, College Park, MD, USA

Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18:613–620. doi:10.1145/361219.361220

Saracevic T (1996) Relevance reconsidered. Paper presented at the information science: integration in perspectives. Second conference on conceptions of library and information science, Kopenhagen

Saracevic T (2007) Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. J Am Soc Inf Sci Technol 58:1915–1933. doi:10.1002/asi.20682

Saravanan M, Ravindran B, Raman S (2009) Improving legal information retrieval using an ontological framework. Artif Intell Law 17:101–124

Shapiro FR (1991) The most-cited articles from the Yale Law Journal. Yale Law J 100:1449

Spriggs JF, Hansford TG (2000) Measuring legal change: the reliability and validity of Shepard's citations. Polit Res Q 53:327–341

Susskind R (2013) Tomorrow's lawyers: an introduction to your future. Oxford University Press, Oxford

Sutton S (1994) The role of attorney mental models of law in case relevance determinations: an exploratory analysis. J Am Soc Inf Sci 45:186–200

Tarissan F, Nollez-Goldbach R (2015) Temporal properties of legal decision networks: a case study from the international criminal court. Paper presented at the legal knowledge and information systems—JURIX 2015: the twenty-eighth annual conference, Braga, 10/11-12-2015

Toms E (2000) Serendipitous information retrieval. Paper presented at the DELOS workshop: information seeking, searching and querying in digital libraries

Turtle H (1995) Text retrieval in the legal world. Artif Intell Law 3:5–54

van Opijnen M (2010a) Canonicalizing complex case law citations. Paper presented at the legal knowledge and information systems—JURIX 2010: the twenty-third annual conference, Liverpool

van Opijnen M (2010b) Searching for references to secondary EU legislation. Paper presented at the fourth international workshop on Juris-informatics (JURISIN 2010), Tokio, 18/19-11-2010

van Opijnen M (2013) A model for automated rating of case law. Paper presented at the fourteenth international conference on artificial intelligence and law, Rome, 11/13-06-2013

van Opijnen M (2014) Open in het web. Hoe de toegankelijkheid van rechterlijke uitspraken kan worden verbeterd. Dissertation, Amsterdam UvA

van Opijnen M (2016a) Court decisions on the internet, development of a legal framework in Europe. J Law Inf Sci 24:26–48

van Opijnen M (2016b) Towards a global importance indicator for court decisions. Paper presented at the legal knowledge and information systems—JURIX 2016: The twenty-ninth annual conference, Nice

van Opijnen M (2017) The European legal doctrine identifier—a missing link? In: Peruginelli G, Faro S (eds) Access to legal scholarship: tools, approaches, technologies (Forthcoming). Giappichelli, Italy

van Opijnen M, Ivantchev A (2015) Implementation of ECLI—state of play. Paper presented at the legal knowledge and information systems—JURIX 2015: The twenty-eighth annual conference, Braga, 10/11-12-2015

van Opijnen M, Verwer N, Meijer J (2015) Beyond the experiment: the eXtendable legal link eXtractor. Paper presented at the workshop on automated detection, extraction and analysis of semantic information in legal texts, held in conjunction with the 2015 international conference on artificial intelligence and law (ICAIL), San Diego, 12-06-2015

Vibert H-J, Jouvelot P, Pin B (2013) Legivoc—connecting laws in a changing world. J Open Access Law 1

Winkels R, de Ruyter J, Kroese H (2011) Determining authority of Dutch Case law. Paper presented at the legal knowledge and information systems. JURIX 2011: the twenty-fourth international conference, Vienna

Winkels R, Boer A, Vredebregt B, Someren AV (2014) Towards a legal recommender system. Paper presented at the 27th international conference on legal knowledge and information systems (JURIX 2014), Krakow

Woods J (2010) Relevance in the Law: a Logical Perspective. In: Gabbay DM, Canivez P, Rahman S, Thiercelin A (eds) Approaches to Legal Rationality. Springer, Dordrecht

Zhang P (2015) Key-note speech on ICAIL 2015 workshop on automated detection, extraction and analysis of semantic information in legal texts. www.lrdc.pitt.edu/ashley/icail2015nlp/