An Approach for Modelling Relevance in Legal Ontologies

Cristiana SANTOS,[a,d] Víctor RODRIGUEZ-DONCEL,[b] Pompeu CASANOVAS,[a,c], Leon Van der TORRE[d], Livio Robaldo[d]
[a] *Autonomous University of Barcelona-Institute of Law and Technology*
[b] *Ontology Engineering Group, Universidad Politécnica de Madrid*
[c] *Faculty of Business and Law, Data to Decisions Cooperative Research Centre, Deakin University, Geelong, Australia*
[d] *University of Luxembourg*

Abstract. This paper presents an analysis on the concept of "relevance" in the legal domain. It uses a conceptual framework of relevance dimensions and tailors it to legal computational ontologies to represent relevant legal information aimed to help in legal decision-making processes. A case study in the consumer law domain is presented.

Keywords. Relevance, legal ontology, consumer law

Introduction

In legal decision-making processes, there is a need to retrieve the information that is most relevant for a dispute, in order to help consumers assessing the legal merit/non-merit of a dispute, discouraging unmeritorious or frivolous complaints. For example, a passenger may better identify if the delayed flight of one hour, to which he has been subjected to, is legally motivated and if the reported dispute is eligible for redress; accordingly, the relevant delay at departure is considered only from two hours or more. The air carrier also can better manage dispute avoidance and dispute containment.

Relevance[*] is assessed *in relation to* the matter at hand, a relation between two entities that corresponds to: (i) *the information resource*, such as a document, a bibliographical representation, or information, subsumed to the information need of the user; and (ii) *the information object retrieved by a system*, obtained upon the introduction of a query or a request, represented in the system language.

Relevant information can be represented by computational ontologies, modelling the legal information objects in an organized and structured way, within a conceptual model, subject to queries and assertions, in a coherent and consistent manner [4].

Therefore, the research question of this paper is how can relevance be modelled within ontologies. To answer to this research question, Saracevic's relevance framework was chosen [1] and tailored to legal domain, comprising: *algorithmic relevance*, referring to the way the information need (query or request) is represented within a

[*] Theoretical frameworks on relevance already exist. We will put aside in this paper other approaches in which relevance has played a major role: computational linguistics (to summarise texts, identifying and clustering information, or discovering synonyms) [29] [30] [31], logic (to solve deontic paradoxes) [28], discourse analysis (semantic and pragmatic markers) [32] and Information Retrieval (IR) [1].

system; (2) *situational relevance*, referring to the situation / work task at hand of the user, normally translated to a taxonomy of disputed cases; (3) *topical relevance*, referring to the topic of the legal information need; (4) *cognitive relevance*, referring to the cognitive needs of the user; (5) *legal salience*, referring to the most important legal sources recognized by the legal community.

This framework was already approached in the legal domain by legal scholars such as Van Opijnen, tailoring the five dimensions to the legal domain [2] [20]; and Geist, deepening mainly the algorithmic and domain relevance [3].

This quadrant of various interplaying relevance's is useful for various reasons: for avoiding ambiguities on which kind of relevance is to be considered to decide the core of a dispute; each specific dimension is made explicit and provides assessment criteria to understand how relevance is inferred; each dimension covers an information need of the user: the topic at hand, the cognitive need, the problem at hand, the domain authoritative sources, etc., which are important to the legal decision making process; and finally, each dimension is equally important.

This paper proposes a modelling of the various interplaying relevance's useful to make explicit which kind of relevance is to be considered to decide a legal case. Hence, relevance dimensions are formalized in legal ontologies in order to represent information in a structured and explicit way. In this work, relevance dimensions are tailored to legal ontology-building, in particular, in the Relevant Legal Information in Consumer Disputes Ontology (RIC) and its specialization, the Air Transport Passenger Incidents Ontology (RIC-ATPI).

The paper is organized as follows. Section 1 reports previous work on modelling relevance within ontologies. Section 2 describes the relevance framework and its five dimensions tailored to legal ontologies. Section 3 presents an example of such tailoring. Section 4 concludes the work.

1. Related Work

The ability to represent relevant information is an intrinsic feature of every ontology, as an ontology by definition includes relevant aspects of a certain domain. But acknowledging that relevance is a mutable relational property of information, and that axioms may become relevant or irrelevant depending on the contexts, is a rarer concern among ontologists. In this section, related ontologies modelling relevance are mentioned, emphasizing the dimensions covered.

Bobillo et al. proposed an ontology design pattern for representing relevance in OWL [5]. Concepts in the domain ontology and concepts in the context ontologies are related through the instances of a "relevance" class, reification of the relevance relations between relevant information and context. The changes which are necessary to implement Bobillo's pattern would consist of (i) defining a single Relevance class; (ii) defining five instances of relevance and declaring axioms that relate the actual classes for each case.

Stojanovic presented an IR ontology-based approach for determining relevance, exploiting the semantics of explicit links [6]. The querying process is supported by an ontology such that other important sources for determining the relevance of results can be considered: the structure of the underlying domain and the characteristics of the searching process. Only algorithmic and topical dimensions are covered. Saravanan et al. postulates that relevance in legal information retrieval is improved by using an ontological framework [7]. They apply a standard mathematical model for term patterns identification in judgments. The model only approaches relevance in its system

and topical dimensions. Schweighofer acknowledged the improvement of Boolean search with query expansion using lexical ontologies and user relevance feedback [8], testing it in a prototype in the area of European state aid law. It is the authors' cognition that legal information system providers store user's information on search practices, and using query logs to increase search engine performance would be easy to implement. This model depicts cognitive, topical and algorithmic relevance dimensions. Lame presented an IR-oriented legal ontology of French Civil Code Law. The ontology provides request extension and text categorization. NLP tools are used to detect relevant domain terms and lexical relations in French legislation [9]. The ontology is not formalized in any existing representation language, as OWL, to understand the intended conceptualization. Only topical relevance is envisioned. *BEST* ontologies consists in an IR based-ontologies with the aim of retrieving relevant knowledge regarding what the outcome of a court decision would be in a similar case. This information may offer laymen insights regarding their positions for negotiation. It has two main functions: (i) supporting the user to describe a specific legal situation [10]; and (ii) retrieving and ranking descriptions of court decisions on similar cases, making use of document retrieval system based on terms provided by the laymen that match terms provided in stored case law.

Both RIC-ATPI (described in Section 3) and the BEST ontologies have the same viewpoints: the ontology retrieves the relevant information to the end-user, convening algorithmic, topical and cognitive relevance. Even if topical relevance is addressed, the information retrieved does not consider the granular information according to cognitive, situational and domain relevance. Other sources of law are not consulted to provide insights regarding laymen's positions for negotiation. RIC and RIC-ATPI ontologies instead cover the five dimensions of relevance. Table 1 depicts the dimensions of relevance covered by each of the presented ontological models.

Table 1 Dimensions of relevance covered by ontological models

Ontological models	Algorithmic	Topical	Situational	Cognitive	Domain
Bobillo's ODP on Relevance	✓				
Stojanovic's IR	✓	✓			
Saravanan's LIR	✓	✓			
Schweighofer's Boolen Search with Query Expansion LIR	✓	✓		✓	
Lame's LIR		✓			
BEST Project LIR	✓	✓	✓		

2. Relevance Dimensions in Legal Ontologies

In this section, the five dimensions of relevance are defined by indicating the relation between some entity and the information object retrieved. In this work, a strict interpretation of the information object is followed, only comprising the textual information encountered at different levels: source, content and document level of legal information. As relevance presupposes a context, the chosen use-case consists in modelling relevant legal information of consumer disputes in the air transport passenger (hereinafter ATP) domain [11]. Figure 1 depicts each dimension; in figure 2 the different types of relevance are defined by the elements under comparison and Table 2 summarizes the relevance dimensions applied to RIC and RIC-ATPI ontologies.

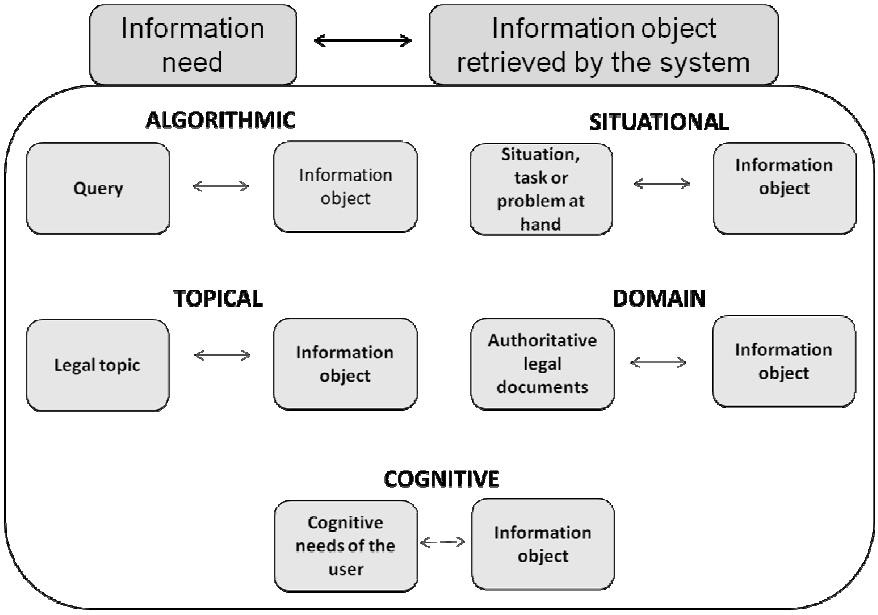

Figure 1 Dimensions of relevance

Figure 2 depicts the various interplaying relevance's. A user faces a legal case at hand (e.g., a passenger wants to claim for his rights); then an information need is perceived (e.g., minimal delay to be compensated) which is made explicit as an information request in the interface of an information system (e.g. a web form). The information system translates it into a query (e.g. SPARQL) and retrieves an information object (e.g. an excerpt of the applicable law).

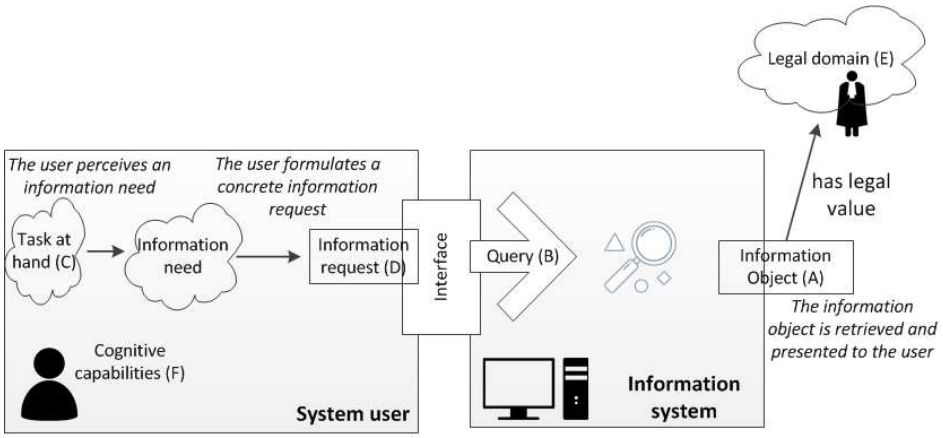

Figure 2. Different types of relevance are defined by the elements under comparison: algorithmic relevance (A and B), situational relevance (A and C), topical relevance (A and D), domain relevance (A and E), cognitive relevance (A and F).

2.1. Algorithmic Relevance

Following Saracevic's cognition, algorithmic relevance consists on a "*relation between the query and the information objects in the file of a system as retrieved by a given procedure or algorithm. Each system has ways and means by which given objects are represented, organized, and matched to a query. They encompass an assumption of relevance, in that the intent is to retrieve a set of objects that the system inferred (constructed) as being relevant to a query.*". In this relation, the query is to be understood as the computer processable translation of the request, as entered in the user interface or any other intermediary component, by the end-user. In this work, the information objects are represented through ontologies, in a specific ontology language, querying the information modelled therein. The intent is to retrieve the information objects the system deemed relevant to a query.

2.2. Topical Relevance

This dimension is both system-oriented, because the success of the relation depends on the system's modelling ability to retrieve relevant information, and also user-dependent on the formulation of the user request. The subjectivity is demarcated if human knowledge organisation is applied to the system [12]. This dimension entails the existence of a taxonomy classifying the documents, and the information needs corresponding to this taxonomy. Many types of legal information, particularly legal cases, are routinely categorized into a set of fixed or fluid categories [1]. In our work, the topical-cases were computed in the ontology, corresponding to the ATP domain and a case-taxonomy defines the relevant cases/incidents of this domain. Therefore, the topicality match relates the incident selected by the consumer and the retrieved information objects from the ontology-based system. The user is limited to formulate his request: he must align it with the available classification system.

2.3. Cognitive Relevance

Cognitive correspondence evokes the following questions: Who are the end-users? What are the cognitive information needs the users have? Is the document or information recovered really related to the underlying, maybe implicit, information need of the user? A cognitive state generally means the tacit knowledge of the user. Cognitive relevance is system and user-dependent: the features of an ontology-based system should take into account the user's background conceptual knowledge, preferences, and also his understanding or perception of his information need. The system features should be tailored to the past search experience, and should be able to explicitly or implicitly understand the information needs of the users. Within our case-study, both in the analysed consumer complaints and in the ATP literature analysis, it was observed that the description of the incidents was sufficiently conveyed and also the request explicitly addressed, which enabled the identification of the information needs. To model the ontologies with the cognitive relevance dimension, it was taken into account the user's search behaviour by consulting the Report from the European Consumer Centre Network on Air Passenger Rights (2015), studies on information-seeking behaviour of consumers, and the theory of behavioural economics embedded in consumer policy [13]. Accordingly, consumers need user-specific information that can provide the relevant benefits (ex. immediate exercise of rights, money savings, time savings), enabling consumers ''to know before they owe", as a "smart disclosure", detailed information in standard, machine-readable formats. The user formulating a request in the ontology will attempt to find the rights, conditions, exceptions, constraints, interpretations, requisites that matches the incident at stake, e.g. the rights of a passenger in case of a cancelation of a flight, or the exceptions to the right of compensation.

2.4. Situational Relevance

This relation involves both system and user, as the success of the relation depends on the system's modelling ability to represent the relevant texts reporting the problems of the user, but also depends on the user's ability to use the information objects for a certain purpose, within a given situation. According to such dimension, the information objects retrieved in the system should be deemed appropriate to solve the problem space of the end-user, or the end-user's legal problem, useful in decision-making, or at reducing uncertainty. The following questions may arise: does the document/information found really help the user to solve his (legal) problem? Discovering the actual problem of the users stems from a problem-oriented approach, or analysing the complaints and disputes (in our case-study). The context of legal decision-making requires access to facts derived from empirical observation analysed in our study. The information modelled in the ontologies hence correspond to real needs reflecting empirical evidence convened through complaints and also studies estimating passenger's patterns. To incorporate this dimension, the ontologies provide the most important disputed problems in the ATP domain and the correspective rights, amenable to a better decision-making.

2.5. Legal Salience

Marc Van Opijnen delineates this dimension of relevance as "domain relevance", also used as a synonym for "legal authority" and "legal importance". For the first time, this dimension was tailored to the legal field. It is defined as the relevance of information

objects within the legal domain itself, encompassing the general opinion of the legal community or "legal crowd" on the significance of a case for legal theory and practice [14]. The author attends to the specific features of legal information, as well as for modelling reasons. Within legal operators, controversy regarding divergent opinions and perceptions are a constant. Also, authority is generally related to a corresponding sanctioning power and liability. Instead, in this work, following the author's cognition and adding a pragmatic approach, it is used legal salience of information in the accounted operation or case. Even if independent from an information system and from any user request, once adapted to knowledge engineering, the ontology should present the most important legal information within the domain. Cosijn mentions that metadata could ameliorate this (socio-cognitive) relevance [12].

Table 2 Summary of the relevance dimensions applied to RIC and RIC-ATPI ontologies

Relevance dimensions	Describes a relation between	Assessment	Relevance in RIC and RIC-ATPI
Algorithmic	Query and information objects	Similarity (weighting and ranking functions)	Ontology modelling: class individuals of RIC-ATPI, in particular, by using the entry rdfs:comment entry provides the needed capability to annotate or describe an entity. The relevant information is retrieved by means of a predefined Sparql query
Topical	Topic in the request and topic covered	Aboutness Interpretation	ATP domain
Cognitive	Information needs (background or specific needs) of the user and information objects of the system	Correspondence, novelty, information quality, informativeness, Preferences, Background knowledge	Information needs of the end-user (consumers), refracted in RIC classes of rights, obligations, legal sources, constraints, exceptions, and its class instances of RIC-ATPI
Situational	Situation / work task at hand and information object	Usefulness in decision-making, appropriateness in problem-solving, reduction of uncertainty	Consumer disputes in the ATP domain, modelled as information objects in RIC-ATPI ontology. RIC:EnforcementProcedure approaches this relation
Legal Salience	Opinion of the legal community and information objects	Legal salience, Citations	Two-folded, requesting the most important documents, within the specific legal domain, such as case-law, legislation, a.s.o. The class RIC:FurtherInterpretation provides additional relevant information related to a right; rdfs:seeAlso entry relates a resource to another resource that might provide additional information about the subject resource

3. RIC and RIC-ATPI Ontologies

RIC represents relevant legal information in a domain-neutral manner, hence, able for reuse in other consumer domains (telecommunications, banking, utilities, etc.). It includes nine concepts describing the Rights emerged whenever an Incident occurs. Rights are drawn from a LegalSource. The entitlement of Rights depends on some

Requisite. The scope of the Rights may encompass relevant information, such as: Exception, Constraint, EnforcementProcedure and FurtherInterpretation, each of them referring to a specific LegalSource, respectively. It includes six object properties: isTriggeredBy, hasRequisite, subjectTo, hasEnforcementProcedure, hasLegalSource and isReportedIn. RIC-ATPI ontology extends RIC ontology, representing the incidents in the ATP domain, and the relevant information derived thereof as class-instances of RIC classes. The modelling of both ontologies and its evaluation are documented in http://ricontology.com [†]; the SPARQL queries used during the execution of the online application are available online.[‡§] Different types of relevance are measured differently: whereas the algorithmic relevance will be a numerical value, other relevance measures will be simply Booleans or complex appraisals.

4. Future work and conclusions

This work presented an approach for modelling relevant legal information using ontologies, contextualized within a use-case of consumer disputes. The five dimensions of relevance are intertwined with each other, and the information objects are modelled as instances of RIC-ATPI once querying the ontologies. We claim that we model in our ontologies the important dimensions for legal decision-making: the ontologies (systemic relevance) are able to answer, through SPARQL queries, what is the relevant information within the ATP incident (legal salience and topical relevance); or what does the user need to know (cognitive), according to his legal problem (situational).
Relevance is an important component in legal knowledge representation and reasoning and modelling relevance can improve the performance of knowledge-based legal systems with more precise and pertinent results in legal decision-making.

As future work, and following Bobillo's philosophy, a small ontology will be built to label every piece of information in the different dimensions of relevance and to describe the user context (including the user's cognitive abilities). Some of these values for relevance, cannot be pre-established, but must be dynamically calculated. However, we believe that having the terms for each of the relevance aspects will foster computer scientists to determine what to evaluate algorithmically, posing the challenges.

In the long run, this work will be analysed within the big data realm, cognitive computing, existing predictive algorithm and with the results offered by commercial solutions like Thomson Reuter's "eDiscovery point" or RossIntelligence based on IBM Watson. However, before making such comparison, a framework for evaluating every aspect of relevance in these heterogeneous approaches has to be laid out. A data-driven analysis could add an additional dimension to the five-dimensions mentioned in the Saracevic model. This new dimension is connected with the "Unconscious Bias"; such an analysis will be conceived as future work.

[†] The OWL files can be consulted in ric.owl and ric-atpi.owl.

[‡] For example, the following query determines which are the rights for a short delayed departure.

```
? SELECT (str(?lab) as ?label) (str(?com) as ?comment) (str(?sour) as ?sourcelabel) (str(?sc) as ?sourcecomment) (?r as ?uri) ?tipo{ ?r ric:isTriggeredBy ric-atpi:shortdelayedatdeparture . ?r rdfs:label ?lab . ?r rdf:type ?tipo . OPTIONAL { ?r rdfs:comment ?com . } OPTIONAL { ?r ric:hasLegalSource ?ls . ?ls rdfs:label ?sour . OPTIONAL { ?ls rdfs:comment ?sc . } } FILTER (?tipo != owl:NamedIndividual) . } ORDER BY ?label LIMIT 50
```

[§] http://www.ricontology.com/manager

Acknowledgements. This work has been supported by a Juan de la Cierva contract, the project TIN2013-46238-C4-3-R, and by the Erasmus Mundus Joint International Doctoral (Ph.D.) Program in Law, Science and Technology.

References

[1] 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, JASIST, 58(3), 1915-1933

[2] Van Opijnen M. (2014) Op en in het web. Hoe de toegankelijkheid van rechterlijke uitspraken kan worden verbeterd, Amsterdam UvA, Den Haag

[3] Geist A. (2009). Using Citation Analysis Techniques For Computer-Assisted Legal Research In Continental Jurisdictions, Graduate thesis, University of Edinburgh.

[4] Gruber, T.R. (1995). Toward principles for the Design of Ontologies Used for Knowledge Sharing, Int. Journal of Human-Computer Studies, Vol.43, pp.907-928.

[5] Bobillo, F., Delgado, et al., (2007). An Ontology Design Pattern for Representing Relevance in OWL, Aberer, K., et al, (Eds.), The Semantic Web: 6th ISWC Proceedings. Springer, Berlin, pp. 72–85.

[6] Stojanovic, N., 2005. An Approach for Defining Relevance in the Ontology-Based Information Retrieval, in: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, WI '05. IEEE Computer Society, Washington, DC, USA, pp. 359–365

[7] Saravanan M., Ravindran B. et al. (2009), Improving legal information retrieval using an ontological framework, AI&L, 17: 101-124.

[8] Schweighofer E.,Geist A. (2007) Legal Query Expansion using Ontologies and Relevance Feedback, LOAIT

[9] Lame, G. (2005). Using nlp techniques to identify legal ontology components: Concepts and relations. In Law and the semantic web. Legal ontologies, methodologies, legal information retrieval, and applications,V. R. Benjamins et al (Eds). LNCS, Vol. 3369, 169–184. Berlin/Heidelberg: Springer

[10] Klein, M., Van Steenbergen, et al. (2006). Supporting layman in finding relevant court decisions in the best-project. In NWO TOKEN Symposium 2006, Leiden.

[11] Santos C., Rodriguez-Doncel V. et al. (2016) Modelling Relevant Legal Information for Consumer Disputes, Electronic Government and the Information Systems Perspective, V. 9831,LNCS, pp 150-165

[12] Cosijn E. (2003). Relevance judgements in information retrieval, Phd Thesis, University of Pretoria.

[13] Micklitz H-W, Reisch, L. et al (2011). An Introduction to the Special Issue on Behavioural Economics, Consumer Policy, and Consumer Law, JCP, 34:271–276

[14] Van Opijnen M. (2013). A Model for Automated Rating of Case Law, ICAIL, NY, 2013, pp. 140-149

[15] Yates, A., Goharian, N., et al. (2014). Relevance-ranked domain-specific synonym discovery. M. de Rijke et al. (Eds.) ECIR 2014, LNCS 8416, Springer, Heidelberg, pp. 124–135.

[16] Alonso, O. (2013). Implementing crowdsourcing-based relevance experimentation: an industrial perspective, Information retrieval, 16 (2): 101-120.

[17] Teufel, S., Moens, M. (2002). Summarizing scientific articles: experiments with relevance and rhetorical status, Computational linguistics, *28* (4): 409-445.

[18] Blakemore, D. (2003). Relevance and Linguistic meaning. The Semantic and Pragmatics of Discourse Markers, Cambridge: Cambridge University Press

[19] Levinson, S. C. A (1989). Review of Relevance, *Journal of Linguistics* 25: 455-472

[20] van Opijnen M., Santos C., (2017). On the Concept of Relevance in Legal Information Retrieval (2017). Artificial Intelligence and Law (2017) 25:65-87