

The Flood, the Channels and the Dykes: Managing Legal Information in a Globalized and Digital World

Joost BREUKER^{a1}, Pompeu CASANOVAS^b, Michel C.A. KLEIN^c,
Enrico FRANCESCONI^d

^a*Leibniz Center for Law (ICL), University of Amsterdam, The Netherlands*

^b*Institute of Law and Technology (IDT), Universitat Autònoma de Barcelona, Spain*

^c*Department of Computer Science, VU University Amsterdam, the Netherlands*

^d*Institute of Legal Information Theory and Techniques, Italian National Research Council (ITTIG-CNR), Florence, Italy.*

Abstract: Information search and retrieval are part of daily routines of the legal profession. Lawyers, judges, prosecutors, and legal clerks usually access a number of electronic resources to browse, search, select, or update legal contents. Legal databases have currently become large digital libraries where the tasks related to information-seeking may sometimes be cumbersome. Adding semantics to support information search may provide significant results in terms of efficiency, efficacy, and user satisfaction. Semantic technologies may be able to improve legal information search in the judicial and lawyers' domains. However, legal professionals sometimes prefer following routines than changing their information search behavior. New trends in legal ontologies and Semantic Web technologies may help to improve both professional and laymen's skills.

Keywords: Information overload, legal ontologies, Semantic Web, information retrieval, lawyers, globalization, e-court, e-discovery, e-governance, e-administration

¹ Corresponding author, breuker@science.uva.nl

1. The flood: legal information overload

There is a paradoxical situation in the modern world: although there is an overabundance of available information, it is often difficult to obtain relevant information when it is needed [21]. In addition, researchers in organization and knowledge management have found that the quality and efficiency of decision-making vary with the amount of information people are exposed to. The performance correlates positively with the amount of received information, *but up to a certain point*. If further information is provided beyond this point, the performance of the individual will rapidly decline [22].

These kind of related phenomena have dysfunctional consequences such as stress or anxiety, and have been termed in several ways in the literature: *cognitive overload*, *sensory overload*, *communication overload*, *knowledge overload*, *information fatigue syndrome*, *data smog* or *analysis paralysis* [21, 22]. However, the most common and generic term is *information overload*. Information overload “occurs when information received becomes a hindrance rather than a help when the information is potentially useful” [5, 30].

Information overload is not a new phenomenon. As early as 1545 Conrad Gesner complained in his *Bibliotheca Universalis* (Zurich, 1545) of “that confusing and harming abundance of books” and Adrien Baillet wrote in the *Jugemens des sçavans sur les principaux ouvrages des auteurs* (Paris, 1685, quoted by [10]):

We have reason to fear that the multitude of books which grows every day in a prodigious fashion will make the following centuries fall into a state as barbarous as that of the centuries that followed the fall of the Roman Empire.

This excess of information is also prominently present in the legal domain, which started already centuries ago. In Europe, *la raison écrite* that led to the development of Civil law, and legislation and case law records in the Common law, produced a massive amount of documents since the 16th century. It seems that what Jack Goody [28] called the “domestication of savage mind” —the modern process of thought through technological means— had indeed a strong legal side. At the end of the eighteenth century the English Parliament had quadrupled its legislative output. The Houses of Commons of George III (1760-1820) legislated four times more than those of Wilhelm III (1689-1702) [35, 49]. In USA, the growth of case law was fast (from 18 volumes of legal reports in 1810 to nearly 3800 in 1885). Therefore, at the end of

the nineteenth century the number of caseloads had increased by forty since the beginning of the century [21].

Nowadays, the amount of legal information grows even further because of the ongoing legalization of the society. On the one hand, much new technology requires new and specific law, e.g. around online purchases, security and data protection. On the other hand, law is becoming a suitable application domain for technological developments, as technology can be used to automatically enforce law or act upon it. For example, in case of *digital rights management* the technology guarantees the respect of the copyrights, while the technology in modern *traffic toll systems* automatically charges the driver. In spite of the differences between legal cultures, this creates an increasingly *legalized* society. Today, the legal database of the Publication Office of the European Union, EUR-Lex, contains 1.800.000 documents in 22 European languages. The average number of visits per working day at mid 2006 were 170.000 [8], at mid 2008 approximates 175.000.²

Table 1. Legal markets size. Source: Euromonitor.

Country	Value in 2003	Growth 2002-03	Expected value in 2008
USA	140.3 billion \$	5.6 percent	174.1 billion \$
UK	28	3.6	31.6
France	14.7	14	16.7
Australia	5.9	11.6	9.4
South Korea	1.35	-3.3	1.9
China	1.34	7	1.9
Japan	0.9	8	1.6

The growth of the legal profession is another factor that contributes to flood of the legal information. Especially law firms in the USA and Europe have experienced a permanent growth since the seventies. Table 1 and Figure 1 may show how the situation looks like. The size that legal markets have reached implies that law firms have becoming transnational corporations, competing for revenues and income (Table 1). According to the *The American Lawyer 100 Report* —the top-grossing law firms in the United States— the total revenues reached \$ 64.6 billion in 2007, an increase

² We thank Pascale Berteloot for this updated information.

of 13.6 %. Lawyers themselves are starting to wonder whether such numbers are sustainable [44]. Even cultures that are traditionally less-prone to litigation —such as the Netherlands and other North-European countries— have showed an increase in the number of lawyers per 100.000 inhabitants (see Figure 1 for the numbers of 2004). In 2006, there were 833,763 lawyers registered in the national bars of 31 European countries [17].

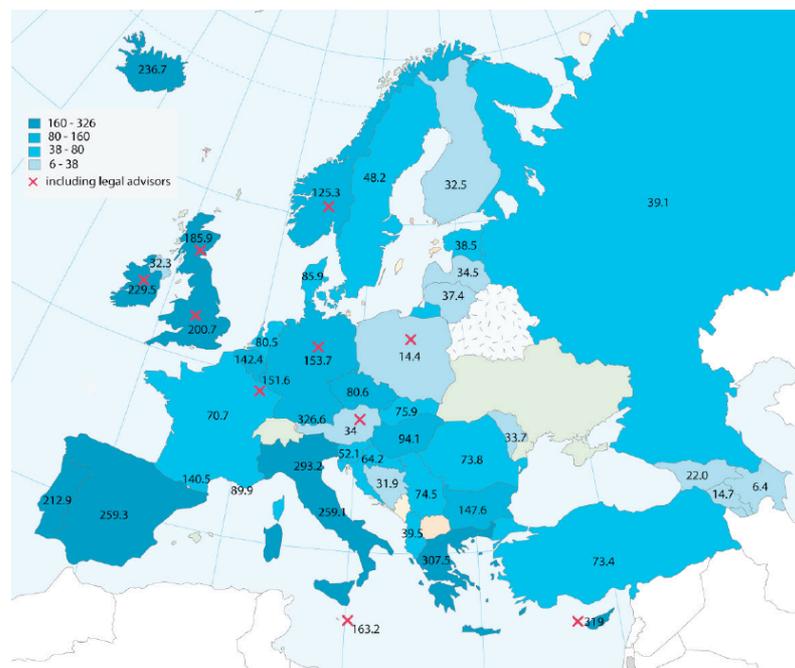


Figure 1. Lawyers in Europe per 100,000 inhabitants per country (2004). Source: Council of Europe. 2006.

Eventually, the legal information overload might lead to a reduced access for citizens to the judicial system. If the right information is not available at the right moment for the right person, it hinders the accessibility of justice. The mere fact that it can take too long to find the right information could already be seen as problematic, according to the classical quotation “*justice delayed is justice denied*” attributed to William Gladstone (1809–1898). In the past, financial constraints hindered the access

to justice, in later years time constraints formed a bottle-neck. It should be prevented that the information overload forms a new barrier for the access to justice in the future. Therefore, measures are needed to channel the legal information flood.

2. Channelling: Semantic Web technology and XML

The Semantic Web may be defined as a large scale, heterogeneous collection of formal, machine processable, ontology-based statements (semantic metadata) about web resources and other entities in the world, expressed in a XML-based syntax. As such, Semantic Web technology and ontologies can help to organize the information overload. They play the following roles:

1. **giving meaning:** ontologies define unambiguously the intended meaning of information;
2. **structuring:** semantic mark-up together with ontologies give structure to documents.

Together, ontologies and semantic web technology facilitates more precise retrieval of legal documents, partial automatic integration and exchange of information, and to some extent automatic reasoning over problems (e.g. question answering, problem solving).

The benefits of Semantic Web technology for the legal domain is leveraged by the increasing availability of legal data in XML format, as such data can be more easily annotated with semantic information. Currently, the adoption and development of standards for legal information, electronic court filing, court documents, transcripts, criminal justice intelligence systems, etc. has become the core activity of a number of projects in several different countries (LegalXML, LEXML, Norme in Rete, MetaLex, LexDania, CHeXML, eLaw, among other initiatives) [9, 26].

All these institutional and research initiatives derive from the awareness of the peculiarities of legal users' information needs (be them law makers, legal experts or citizens) which are increasingly pushing towards the use of advanced Information Technologies in the legal field. Legal information systems should be aimed not only at providing advanced search and retrieval services to the users, but also at maintaining and up-keeping the legal order, at monitoring the impact of new norms, at handling document timeline and versioning.

Moreover, users are mainly interested in accessing norms rather than simply documents; they are particularly interested in knowing the relations between norms and having support to legal reasoning. In this perspective the annotation of legal documents, in particular legislative ones, with shared document standards is particularly desirable to describe their well defined structures and to provide them with metadata able to manage production, preservation and workflow, which involve Parliaments and Public Administrations.

The use of document open standards and Semantic Web technologies represents, therefore, a pre-condition for the development of services for legislators, legal information providers, legal experts as well as citizens. Recently EU Member States and Institutions have considered the use of Information and Communication Technologies in the legal domain of paramount importance to achieve better quality in legislation as well as to improve legal information management and accessibility across the EU. Institutional initiatives in legal document standards have been undertaken exactly to cope with these requirements, with the aim of providing high quality legal information integrated services.

This development is in line with the trends identified by E. Motta and M. Sabou [41]. In a comparison between the first and the next generation of Semantic Web applications, they identify several features of the new orientations: (i) reuse (vs. semantic data generation); (ii) multi-ontology systems (vs. single-ontology systems); (iii) openness with respect to semantic resources, (iv) scale as important as data quality, (v) openness with respect to Web (non-semantic resources), (vi) compliance with the Web 2.0 paradigm, (vi) openness to services.

3. ICT impact on the law field

Before we can see how the law domain can benefit from Semantic Web technology, we first discuss the impact of ICT in general on the field of law. Researchers in Artificial Intelligence and Law used to separate ICT (Information Communication Technologies) and Law into two big domains: (i) ICT law (data protection, copyright, security, domain names...), (ii) ICT for lawyers (e-government, e-court, Online Dispute Resolution, Multi-Agent Systems, etc.) [37, 38]. The first area would cover regulations and protocols. The second one refers to all the languages, tools, software,

etc. that bring support to legal activities at the workplace. From a legal point of view this seems quite reasonable.

However, recent developments in semantic technologies, Natural Language Processing (NLP), legal ontologies, information retrieval technologies, and the Web 2.0 contribute to the convergence of the two approaches into a single techno-legal one. A lawyer seriously interested in meta tagging litigation cannot ignore OWL. A computer scientist developing legal ontologies for procedural legal knowledge must have a clear picture of court proceedings. This may also challenge the traditional “technological gap” or “computer divide” that researches in AI and Law have pointed out when describing the traditional diffidence of the legal field towards technology [34, 37, 38]. Barriers to ICT developments and AI applications certainly persist [33].

It is true that very little AI has been applied to the legal field so far. Detailed descriptions of working e-Court technologies in Europe lead to the same conclusion [24, 25, 42]. It may be true as well that, compared to other kind of company organizations, the legal industry “is a surprisingly fragmented, undercapitalized and inefficient sector” [34]. Very likely, lawyers may do better. However, looking at the available data, there is no doubt that law firms have put effort in improving their skills and efficiency through ICT investments.

Annual technology surveys are part of the legal marketplace, and law firms spend a substantial part of their budget specifically for technology, according to the American Bar Association Tech reports. The average law firm spends 6% to 7% of gross revenue on technology-related expenses. This is correlated with firm size (from 2% to 7%). The 2007 and 2008 ILTA surveys [31, 32] show that small firms (under 200 attorneys) have higher implementation rates for case management, courtroom technology, docketing software, imaging/scanning/OCR, patch and records management software, while large law firms are more interested for remote access technology, voicemail upgrades, wireless connections and workflow automation.

We may distinguish the following ICT domains that have impact on the legal market [15, 46]: (i) Legal Information Research (LIR); (ii) Electronic Data Discovery (e-Discovery); (iii) Web-based communications; (iv) Collaborative tools; (v) Metadata; (vi) XML technologies; (vii) Technologies in Court-rooms and Judicial and Prosecutors’ offices (e-Court); (viii) Technologies in Administration offices (e-Government, e-Administration); (viii) Multimedia and law.

Some fields such LIR are well trodden-paths, with three big legal information providers (Thomson-Westlaw, Reed Elsevier-Lexis-Nexis, and Wolters Kluwer) controlling about 85% of the market. Others, such as e-Discovery —litigation support software that process, collect, preserve, review and produce electronically stored information, including e-mails— are still emerging. However, the market has gone up to 150 million US dollar. As recently reported in *The Wall Street Journal*, e-discovery conflicts have aroused in recent times between tech-law firms and software companies in the Common Law areas, because this is reducing the need for attorneys doing this kind of work [47].

Perhaps investments in technology could be considered one of the ways firms can reduce costs and improve services. This is related to outsourcing services as well. The December 2007 ILTA White Paper on Law Firm Staffing notes: “There are simply too many IT functions to be performed in today’s law firm and too many different skill sets required to go it alone; outsourcing some IT functions, either in whole or in part, is the norm for firms of all sizes” [27].

However, there is more needed than outsourcing IT functions. We think that adopting new ICT techniques is the only way of coping with legal information overload and with the changes in customers’ expectations coming up with the Web 2.0 and 3.0. Knowledge Management in law firms may be defined as “the way in which lawyers optimize the relation between knowledge and knowledge processes with the help of Information Technology” [3]. We will show in the next section that this perspective is not only a professional issue, but an economic, political and cultural one. Focusing on the individual attorney perspective and combining different types of knowledge, as Apistola and Lodder [3] suggest, constitutes a good strategy to face the transformation which is taking place in the legal field.

4. The need for Semantic Web technology in the legal domain

The traditional fields and practices of law are changing fast. Legal drafting, private contracting, judicial sentencing and administrative management have been enlarged with online dispute resolution initiatives and new forms of self-regulation and access to justice. Citizens, customers and consumers require a greater participation and faster and more effective ways of facing their legal activities. The emergence of auto or self-made law within the Internet out of the law firms influence has been already noticed and checked by some market analysts [50].

Therefore, there is a clear need for less expensive lawyering, less adjudication procedures, more dialogue, more participation, and more flexibility and autonomy. These also seem to be the aims of new legal forms of *relational administration* and *relational justice* [14]. The urgency of this is clearly illustrated by a quote of a forthcoming paper by Colin Rule³ [45], affiliated with e-Bay:

“If you have any doubt that consumers are moving to online commerce, take a look at eBay, the online auction company. In the 13 years since it was founded, eBay has grown into the largest marketplace in the world. In the first half of 2008, there were more than one billion product listings added to eBay worldwide. At any given moment, there are more than 100 million listings around the world, and approximately 7.1 million listings are added each day. eBay users trade almost every kind of item imaginable, in more than 50,000 categories. On eBay, a pair of shoes sells every 7 seconds, a cell phone sells every 7 seconds, and a car sells every 56 seconds. The daily volume of trade on eBay is greater than the daily volume of the NASDAQ.

Unsurprisingly, all of these transactions generate a lot of consumer disputes. Even though less than 1 percent of purchases generate a problem, the incredible volume on the site means eBay handles more than 40 million disputes a year, in more than 16 different languages.”

Modern Web 2.0 developments [40] and Multi Agent Systems technology [12, 13] seems to be able to provide (partial) answers to these needs. It has been highlighted that the Web 2.0 implies a *democratic* model. People can cooperate and jointly build their ideas. Enriching this process in the direction of Web-mediated dispute resolutions seems a quite natural move from the Semantic Web perspective. However, it is not that easy. Semantic Web developers themselves have pointed out some obstacles or limitations of the original purposes of the Semantic Web [7]. For example, search in the World Wide Web is the great unfulfilled promise. Bridging Semantic Web and information retrieval technologies face scientific problems on knowledge representation and natural language understanding that remain still unsolved. In Baeza-Yates, Mika and Zaragoza’s words “IR research is strongly driven by a problem, whereas Semantic Web is driven by a solution” [4].

³ We thank the author for allowing us to quote his still unpublished paper.

Although this may be true, hybrid approaches and perspectives seem to guide the ontological work, as long as folksonomies, wikis, data mining, NLP techniques, and upper, middle and domain ontologies develop [2, 11, 48, 55]. Moreover, to optimize the possibilities of success, SW strategies should be grounded on detailed economic, sociological and cognitive studies on the daily behavior and real needs of professional ICT users and consumers.

Another need for semantic technology follows from the usage of the web by legal professionals. Not only attorneys, but even judges and magistrates are currently browsing the web when they gather information to build their legal strategy to construct and solve the cases they have in front of them. According to the recent American Bar Association *2008 Legal Technology Survey Report* legal research online has grown up from 79% in 2003 to 96 % in 2008 [1]. In 2008 the number of lawyers using free online research services has overtaken for the first time the number using for-fee services (89% vs. 83 %). Lawyers receive information mainly through news websites (79%) and e-mail newsletters (59%), and 72% of respondents report that they or their staff file court documents electronically, up from 55% in the 2007 survey.⁴ In addition, legal chats and blogs can be found along with public or private legal databases. Moreover, e-mail boxes are currently being used as personal databases.

We do not argue that common sense knowledge plays the same role as expert knowledge does when facing a legal case. However, lay people and experts tend to rely on the web for their information needs, and the web offers increasingly accessibility to documents containing legal rules and procedures, past cases and accumulated experiences.

This goes far beyond the practice to look and seek for information through the existing legal databases. We think that the reason for what lawyers do this is because they save time and effort. Therefore, expert knowledge, personal and professional experiences, and common knowledge have to be combined in a new way that reduces the differences among experts and lay people or, at least, approaches the legal perspective to the social perspective of non-legal users. Thus, lawyers (and magistrates) practice law *through* and *within* the web. This integration can clearly benefit from semantic technology.

⁴ The Survey report is based on 789 completed Baseline and Budgets questionnaires: <http://www.abanet.org/abastore/index.cfm?fm=Product.AddToCart&pid=2680079PDF>

In short, both the amount and character of legal activities, and the increased use of online information by lay and professional users require new ways of handling legal information. More and more people rely on web accumulated information first to find a solution for their administrative or legal problems. This situation makes the legal activity of citizens and the daily routine of experts suitable to be treated with the Semantic Web techniques.

5. Overview of the status on legal ontologies

Given the expected increasing impact of Semantic Web technology on the legal domain, it is worthwhile to have a look at the ontologies for the legal domain that are available. After all, ontologies are often the core of all applications that exploit Semantic Web technology. In this section, we will discuss the different types of legal ontologies and give an overview of a large number of existing ontologies and their role.

The term ‘ontology’ may have different meanings: (i) philosophical discipline, (ii) informal conceptual system, (iii) a formal semantic account, (iv) a specification of a conceptualization, (v) a representation of a conceptual system via logical theory, (vi) the vocabulary used by a logical theory, (vii) a meta-level specification of a logical theory [29].

For Semantic Web technologies, an ontology may be defined as a 4-tuple $\langle C, R, I, A \rangle$, where C is a set of concepts, R a set of relations, I a set of instances and A a set of axioms [55]. Ontologies consist of concepts, relations, instances and axioms. They represent knowledge in such a way that can be understood and processed by a machine.

Legal ontologies represent legal knowledge. The problem, then, is to define in a more precise way what “legal knowledge” means.

In the mid-nineties, pioneering work by McCarty, Stamper, van Kralingen, Visser, Breuker, Winkels and Valente tried to bridge the gap between computational and legal concepts. Ontologies were faced as “the missing link between legal theory and AI & Law” [51]. Therefore, concepts traditionally viewed as belonging to the legal theory field such as rights, duties, norms and actions were modeled, in

addition to more specific legal terms stemming from a particular domain (criminal, commercial or maritime Dutch law). These early efforts have recently lead to more developed legal-core ontologies, such LKIF-ontology, e.g.

There were four main directions in legal ontology building as identified by Visser and Bench-Capon [53]: (i) legal discourse (McCarty); (ii) legal norms (Stamper); (iii) frame-based ontology of law (Visser and van Kralingen); (iv) functional ontology of law (Breuker and Valente). More general upper and top ontologies lexically-based in Wordnet (Gangemi, Tiscornia, and Sagri) have to be added to these trends. And, to complete the whole picture, fundamental legal concepts and common-sense categories have been recently merged in a general ontology (LKIF) based on the LRI-Core Ontology developed in the nineties.

Ontologies keep growing in the legal field. Table 2 summarizes twenty-three already existing legal ontologies. It should be noticed that other types of legal knowledge are being added to the fundamental ones: legal professional knowledge, multimedia and global contexts (in which digital rights operate e.g), negotiation (ODR), and laymen legal conceptualization set up complementary scenarios.

As stated above, these types of legal knowledge fit into the new trends of the Semantic Web and the development of the Web 2.0. These ontologies are user-centered and even those built for information retrieval purposes are web service-oriented. They intend to operate through the Internet.

Table 2. Extension of André Valente's table of existing legal ontologies [6, 52].

Ontology or Project	Application	Type	Role	Character	Construction	Language
McCarty's Language of Legal Discourse	General language for expressing legal knowledge	Knowledge representation, highly structured	Understand a domain	General	Manual	English
Valente & Breuker's Functional Ontology of Law	General architecture for legal problem solving	Knowledge base in Ontolingua, highly structured	Understand a domain, reasoning and problem solving	General	Manual	English

Ontology or Project	Application	Type	Role	Character	Construction	Language
Van Kralingen & Visser's Frame Ontology	General language for expressing legal knowledge, legal KBSs	Knowledge representation, moderately structured (also as a knowledge base in Ontolingua)	Understand a domain	General	Manual	English
Mommer's Knowledge-based Model of Law	General language for expressing legal knowledge	Knowledge base in English very highly structured	Understand a domain	General	Manual	English
Breuker & Hoekstra's LRI-Core Ontology	Support knowledge acquisition for legal domain ontologies	Knowledge base in DAML+OIL/RDF using Protege (converted into OWL)	Understand a domain	General	Manual	English
Hoekstra & Breuker's LKIF-Core Ontology	Support knowledge acquisition for legal domain ontologies	Knowledge base in OWL, highly structured	Understand a domain	General	Manual	English
Gangemi, Sagri & Tiscornia's JurWordNet	Extension to the legal domain of WordNet	Lexical Knowledge base in DOLCE (DAML), lightly structured	Organize and structure information	General	Manual	Italian
Benjamins, Casanovas et al. Ontology of Professional Legal Knowledge (OPLK)	Intelligent FAQ system (information retrieval) for judges (luriservice)	RDF.. Knowledge base in Protégé, highly structured (converted in OWL)	Semantic indexing and search	Domain	Semi-automated	Spanish
Casellas, N. et al. Ontology of Professional Judicial Knowledge (OPJK)	i-FAQ for judges (luriservice, second version)	Last version in OWL. Knowledge base in Protégé, highly structured	Semantic indexing and search	Domain	Manual	Spanish

Ontology or Project	Application	Type	Role	Character	Construction	Language
Lame's ontologies of French Codes	Legal information retrieval	NLP oriented (lexical), knowledge base, lexical, lightly structured	Semantic indexing and search	Domain	Automated	French
Leary, Vanderverghe & Zeleznikow's Financial Fraud Ontology	Ontology for representing financial fraud cases	Knowledge base (schema) in UML, lightly structured	Semantic indexing and search	Domain	Manual	English
Asaro et al.'s Italian Crime Ontology	Schema for representing crimes in Italian law	Knowledge base (schema) in UML, lightly structured	Organize and structure information	Domain	Manual	Italian
Boer, Hoekstra & Winkel's CLIME Ontology	Legal advice system for maritime law	Knowledge base in Protégé and RDF, moderately structured	Reasoning and problem solving	Domain	Manual	English
Lehman, Breuker & Brouwer's Legal Causation Ontology	Representation of causality in the legal domain	Knowledge base lightly structured	Understand a domain	Domain	Manual	English
Delgado et al.'s IPROnto (Intellectual Property Rights Ontology)	Integrating XML DTDs and Schemas that define Rights Expression Languages and Rights Data Dictionaries	Knowledge base: first version in DAML+OIL (2001), current version OWL (2008)	Interoperability between Digital Rights Management (DRM) systems	Domain	Manual	English
Teodoro, Binefa et al. e-Sentencias (Procedural Ontology for Multimedia in Courts)	Ontology for Representing Procedural Stages of Spanish Civil Hearings	RDF. Procedural Knowledge within Spanish Civil Hearings (typology)	Diarization and Content Classification of the Official Video Recordings (image and audio)	Domain	Manual	Spanish

Ontology or Project	Application	Type	Role	Character	Construction	Language
J. Saias, P. Quaresma, Portuguese Attorney Office Ontology	Ontology to semantically enriching legal texts	OWL and logic programming (ISCO and EVOLP)	Organize and structure information	Domain	Automated	Portuguese
M.Klein, E.Uijttendbroek, A. Lodder, Laymen Ontology	Ontology to represent laymen knowledge on liability cases	OWL and NLP. Knowledge base in laymen natural language	Understand a domain (tort law) and interoperability between NL and legal concepts	Domain	Semi-automated	Dutch
J. Breuker, A. Elhag's Crime.NL	Ontology of Dutch Criminal Law	OKBC	Main structure of (Dutch) criminal law; for comparing European CL	Domain/General	Manual	Dutch / English
Tiscornia, Francesconi, Spinosa et al. - DALOS	Legal drafting	OWL	Support for multilanguage legal drafting and document indexing	Domain	Manual/automatic	English, Spanish, Italian, Dutch
S Despres, S. Szulzman Micro-ontology	Ontology to represent concepts in European Directives	OWL and NLP (TERMINAE method)	Understand a domain	Domain	Semi-automated	French / English
UCC Ontology J. Shaheed, A. Yip, J. Cunningham	Ontology to represent top-level concepts (e.g. ownership)	NML Top-level ontology based on NM	Organize and structure information	Domain (top-level)	Manual	English
E.Schweighofer, D. Liebwald's CLO (Comprehensive Legal Ontology)	Ontology for information management	Some frame representation		General	Manual with support of legal core ontologies	English?
E. Melz & A. Valente's IRC ontology	Ontology of Internal Revenue Code (USA)	OWL	Reasoning about tax cases	Domain	Manual	English

6. About the content of this book

The articles in this book are sorted in two main categories: 1) those which focus on the modeling of legal ontologies and the resulting ontology itself, and 2) those presenting applications, which use ontologies and Semantic Web technology. Within each category one will find a high diversity in kinds, roles and applications of legal ontologies. The message should be clear: in the legal domain – and probably in other domains as well – the nature and use of ontologies is highly diversified. This is no so much due to the fact that there are still differences in believing what an ontology should be about, but rather to the discovery that ontologies play a pivotal role in applications that require some semantic understanding of terms used by these applications. This understanding may range from automated reasoning to semantic support in legal information and knowledge management.

The first article by **Hoekstra, Breuker, Di Bello and Boer** presents LKIF-Core, a core ontology that contains definitions of concepts that are general and typical for legal domains. The ontology is cast in OWL-DL and its main distinctions are between physical and mental concepts. LKIF-Core is particularly worked out on notions about documents and (legal) roles: roles being the basic terms for social structures. The ontology serves a number of purposes: the usual and main one being a template for modeling legal domain ontologies: an example modeling a EU Directive on driver's licenses.

The article by **Gangemi** discusses another way in which ontological definitions can act as templates in knowledge acquisition for legal domains: by using 'design patterns': recurring, abstract structures of defined terms. They can be reused as "core components" in ontology construction in (legal) domain. An overview of the typical tasks and services for legal knowledge is presented, the notion of ontology design pattern is introduced, and some excerpts of a reference ontology (CLO) and its related patterns are included, showing their utility in a simple legal modeling case.

If these two articles provide top-down support for developing legal domain ontologies, the article by **Lenci, Montemagni, Pirrelli and Venturi** explains how legal ontologies can be modelled in a bottom-up way by using a tool that enables learning from text (T2K) that combines a full range of technologies from Artificial Intelligence (AI): natural language processing, statistical text analysis and machine learning. It is applied to two legal text corpora and it is shown that the interleaving of

these technologies provide a valid harvest of terms, which form semantic clusters which can be further processed manually as to construct an ontology.

The next article is complementary to the previous one, as it also combines linguistic extraction and statistical techniques to arrive at ontology population. **Walter and Pinkal** describe how they trace (legal) definitions from 6000 German court verdicts: they observe that definitions are an important element of legal text, and apply linguistic tools for their extraction. These definitions lend themselves easily to ontological modeling. The tools used enable the extraction of these definitions with high precision, although recall is to be improved. Also, here, the actual ontology is a next step, not reported.

The article by **Mochales and Moens** discusses also natural language processing technologies. However, in this case the aim is not to extract terms for an ontology, but to detect (legal) argument structures in text. It differs from the previous two articles that it is not so much concerned with definitions of terms – the basic stuff of ontologies – but of reasons. Legal reasoning uses (many) definitions of terms, but they are only components in the reasons for justifying decisions. The article provides an overview of technologies for automated argument detection, and shows that argumentative text parts can be separated from non-argumentative ones. They demonstrate that many applications are available for use on the (Semantic) Web, in particular for legal domains.

Argumentation is also a central issue in the article by **Trojahn, Quaresma and Vieira**. As ontologies become more and more available on the web with similar or overlapping terms, it becomes essential to assess how these overlaps match. In this article a method is presented based upon Value-based Argumentation Framework (VAF) in a multi-agent paradigm. An example is presented, where LKIF-Core (cf. Hoekstra et al., this volume) and LCO (cf. Gangemi, this volume) are compared.

The sharing and reuse of content from the web poses many (new) issues on copyrights. They are discussed in the context of digital right management systems, but it is pointed out by **Garcia and Gil** that these approaches only make sense if supported also by digital tools. To enable (semantic) interoperability the authors have developed an ontology cast in OWL-DL, which takes as main concepts the notions of works, rights, actions, and copyright licenses, which require the modeling of roles, events and deontic modalities.

Casanovas, Casellas and Vallbé present an ontology based upon the professional experiences of judges in providing legal support in day-to-day police activities. This ontology (Ontology of Professional Judicial Knowledge (OPJK)) drives Iuriservice, a FAQ advisory system, intended for novice judges. OPJK, expressed in OWL, provides the semantics for accessing Iuriservice. The development of OPJK which covers practical rather than formal legal knowledge is described in detail.

The contribution of **Agnoloni, Bacci, Francesconi, Peters, Montamegni and Venturi** to this book reports the use of ontologies in legal drafting, and in particular the multi-language aspects of the European context (Directives) which require more than simple dictionary mappings, but also a representation of the underlying semantics. The lexical and semantic knowledge form two connected layers.

The last article by **Casanovas, Binefa, Gracia et al.** describes how a legal ontology (e-Sentencias) enables the management of multi-medial information, obtained in recording hearings of Spanish civil courts. The ultimate goal is to obtain an automatic classification of the episodes of the audio-visual records. The focus in the article is on the knowledge acquisition process.

All together, this book can be read as an overview of ongoing attempts to manage the legal information flood. Ontologies provide the channels that separate the streams on semantic grounds, while Semantic Web technologies take the high ground of dykes to control and prevent overflow.

References

- [1] American Bar Association Technology Resource Center. 2007, 2008 ABA Legal Technology Survey Report: Online Research <http://meetings.abanet.org/ltrc/index.cfm?data=20080909> [accessed September 18th 2008].
- [2] S. Angeletou, S. Specia, M. Sabou, E. Motta. Bridging the Gap between Folksonomies and the Semantic Web: An Experience Report, ESWC-07, 2007, 30-43.
- [3] M. Apistola and A. Lodder. Law Firms and IT. Towards Optimal Knowledge Management, *Journal of Information, Law and Technology* 2(3) (2005), 1-28.
- [4] R. Baeza-Yates, P. Mika, H. Zaragoza. Search, Web 2.0, and the Semantic Web, in V.R. Benjamins, V.R. (Ed.) (2008), 80-82.
- [5] D. Bawden; C. Holtham, N. Courtney: Perspectives on Information Overload, *Aslib Proceedings* 51 (September 1999), 249-255.

- [6] V.R. Benjamins, P. Casanovas, J.Breuker, A. Gangemi (Eds.). *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*. LNAI 3369, Springer Verlag, Berlin, Heidelberg, 2005.
- [7] V.R. Benjamins (Ed.). Near-term Prospects for Semantic Technologies, *IEEE Intelligent Systems*, January/February, (2008), 76-88.
- [8] H. Bernet, and P. Berteloot. EUR-Lex: A Multilingual On-line Website for European Union Law, *International Review of Law, Computers and Technology*, **3**, (2006), 337-339.
- [9] C. Biagioli, E. Francesconi, G. Sartor (Eds.). *Proceedings of the V Legislative XML Workshop*, European Publishing Academic Press, Florence, 2006.
- [10] Blair. Reading Strategies for Coping With Information Overload ca. 1550-1700. *Journal of History of Ideas*, **64**, 1 (Jan. 2003), 11-28.
- [11] J. Breuker, A. Elhag, E. Petkov, R. Winkels. Ontologies for legal information serving and knowledge management. In *Legal Knowledge and Information Systems, Jurix 2002: The Fifteenth Annual Conference*, IOS Press, Amsterdam 2002, 1-10.
- [12] P. Casanovas and P. Noriega. Introduction: Legal Knowledge, Electronic Agents, and the Semantic Web, in in P. Casanovas et al (Eds.) (2007), pp. 15-40.
- [13] P. Casanovas, P. Noriega, D. Bourcier and F. Galindo (Eds). *Trends in Legal Knowledge. The Semantic web and the Regulation of Electronic Social Systems.*, Florence: European Academic Press, Florence, 2007.
- [14] P. Casanovas and M. Poblet. Concepts and fields of relational Justice, in P. Casanovas et al. (2008), 323-339.
- [15] P. Casanovas, G. Sartor, N. Casellas, R. Rubino (Eds.) *Computable Models of the Law: Languages, Dialogue, Games, Ontologies*. LNAI 4884, Springer Verlag, Berlin, Heidelberg, 2008.
- [16] Council of Europe. 2006 European Judicial Systems 2006 (2004 data). http://www.coe.int/t/dg1/legalcooperation/cepej/evaluation/2006/CEPEJ_2006_eng.pdf [accessed September 18th 2008]
- [17] Council of Bars and Law Societies of Europe, *CCBE Statistics 2006*, <http://www.ccbe.org/index.php?id=29&L=0> [accessed September 18th 2008]
- [18] M. D'Aquin, E. Motta, M. Sabou, S. Angeletou, L. Gridinoc, V. Lopez, D. Guidi. Toward a new Generation of Semantic Web Applications, *IEEE Intelligent Systems*, May/June (2008), 20-28.
- [19] J. Davies, M. Grobelnik, D. Mladenic. Challenges of Semantic Knowledge Management, in J. Davis et al. (Eds) (2008), 248-251.
- [20] J. Davies, M. Grobelnik, D. Mladenic (Eds). *Semantic Knowledge Management. Integrating Ontology Management, Knowledge Discovery, and Human Language Technology*, Springer Verlag, Heidelberg, Berlin, 2008.
- [21] Edmunds and A. Morris. The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management* **20** (2000), 17-28.
- [22] M.J. Eppler and J. Mengis. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society* **20** (2004), 325-344.
- [23] P. Fabra and A. Cerrillo (Eds). *E-Justice. Using Information Communicatoin technologies in the Court System*, Information Science Reference, Hershey, New York, 2008.
- [24] M. Fabri and F. Contini (Eds). *Justice and Technology in Europe: How ICT is Changing the Judicial Business*, Kluwer Law International, The Hague, 2001.
- [25] M. Fabri. The Italian Perspective of E-Justice in a Comparative Perspective, in A. Cerrillo and P. Fabra (Eds.) (2008), 1-20.
- [26] Francesconi,. *Technologies for European Integration. Standards-based Interoperability of Legal Information Systems*, European Publishing Academic Press, Florence, 2007.

- [27] R. Friedman. Why and What Lawyers Should Consider Outsourcing. Law and Technology resources for Legal professionals. <http://www.llrx.com/features/legaloutsourcingoptions.htm> [Accessed: September 12th 2008].
- [28] J. Goody. *The Domestication of Savage Mind*, Cambridge University Press, Cambridge, 1977.
- [29] N. Guarino and P. Giaretta. Ontologies and knowledge bases: towards a terminological clarification. In N. Mars (Ed.) *Towards very large knowledge bases*, IOS Press, Amsterdam, 1995, 25-32.
- [30] S. Hardesty and T. Sugarman. Academic Librarians, Professional Literature, and New Technologies: A Survey. *The Journal of Academic Librarianship*, **23**(2) (2007), 196-205.
- [31] *ILTA 2007 Technology Purchasing Survey* [Online]. Available at: <http://iltanet.org/pdf/2007PurchasingSurvey.pdf> [accessed September 10th 2007].
- [32] *ILTA 2008 Technology Purchasing Survey* [Online]. Available at http://www.iltanet.org/communications/pub_detail.aspx?nvID=00000011205&h4ID=000001315505 [accessed September 15th 2008]
- [33] J. Jenkins. What Can Information Technology Do for Law? *Harvard Journal of Law & Technology*, **21**(2), Spring (2008), 589-607.
- [34] M. Lauritsen. Artificial Intelligence in the Legal Real Workplace, in A. Lodder, A. Oskamp (Eds.) (2006), 165-176.
- [35] Lieberman, David. *The Province of Legislation Determined. Legal Theory in Eighteenth-century Britain*, Cambridge: Cambridge University Press, 1989.
- [36] D. Lieberman. *The Province of Legislation Determined. Legal Theory in Eighteenth Century Britain*. Cambridge University Press, Cambridge, 1989.
- [37] Lodder and A. Oskamp. Law, Information Technology and Artificial Intelligence, in A. Lodder, A. Oskamp (Eds), 2006, 1-22.
- [38] Lodder and A. Oskamp (Eds). *Advanced Technology in the Legal Domain. From Challenges to Daily Routine*. Springer, Dordrecht, 2006.
- [39] R. Mizoguchi, Z. Shi, and F. Giunchiglia (Eds). [ASWC 2006] *The Semantic Web*, LNCS 4185, Springer Verlag, Heidelberg, Berlin, 2006.
- [40] E. Motta. Knowledge Publishing and Access on the Semantic Web: A Sociotechnological Analysis. *IEEE Intelligent Systems*, May/June 2006, 88-90.
- [41] E. Motta and M. Sabou. Next Generation Semantic Web Applications, in R. Mizoguchi et al. (2006), 24-29., pp. 24-29.
- [42] Oskamp, A. Lodder, M. Apistola (Eds). *IT Support of the Judiciary. Australia, Signapore, Venezuela, Norway, the Netherlands and Italy*. TMC Acer Press, The Hague, 2004.
- [43] M. Poblet, P. Casanovas, V.R. Benjamins. Technology and Law: The Not So Odd Couple, in P. Warren et al. (2008), 201-211.
- [44] Press and J. O'Connor. Lessons of the Am Law 100: Is the Golden Age Over?, *The American Lawyer*, April, 2008.
- [45] Rule. Making Peace on eBay: Resolving Disputes in the World's Largest Marketplace. *ACResolution Magazine*, Fall 2008 (forthcoming).
- [46] Sartor, P. Casanovas, N. Casellas, R. Rubino. Computable models of the law and ICT: State of the Art and Trends in European Research, in P. casanovas et al. (Eds) (2008), 1-20.
- [47] J. Scheck. Tech Firms Pitch Tools for Sifting Legal Records. *Wall Street Journal*, 22 August, 2008.
- [48] R. Studer, A. Ankolekat, P. Hitzler, Y. Sure. A Semantic Future for AI, *IEEE Intelligent Systems*, July/August 2006, 8-9.
- [49] P. D. Thomas. *The House of Commons in the Eighteen Century*, Oxford University Press, Oxford, 1971.
- [50] G.L. Urban. Online Trust and Customer Power: The Emergence of Customer Advocacy, in P. Warren et al. (Eds.) (2008), 39-54.

- [51] Valente and J. Breuker. Ontologies: the missing link between Legal Theory and AI & Law, in A. Soeteman (Ed.) JURIX 94, Koninklijke Vermande, Lelystad, 1994, 138-150.
- [52] Valente. Types and Roles of Legal Ontologies, in R.V. Benjamins et al. (2005), 65-76.
- [53] P. R. S. Visser and T.J.M. Bench-Capon. A Comparison of Four Ontologies for the Design of Legal Knowledge Systems”, *Artificial Intelligence and Law* 6 (1998), 27-57.
- [54] P. Warren, J. Davies and D. Brown (Eds.). *ICT Futures. Delivering Pervasive, Real-time and Secure Services*. John Wiley & Sons Ltd, Chichester, 2008.
- [55] P. Warren, R. Studer, J. Davies. The goal of the Semantic Web, in J. Davies et al. (Eds.) (2006), 1-18.
- [56] P. Warren, J. Davies, D. Brown. Conclusions, in P. Warren et al. (2008), 229-233.