Eunomos, a legal document and knowledge management system for the Web to provide relevant, reliable and up-to-date information on the law

Undefined 1 (2012) 1
IOS Press

Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the Law

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Guido Boella [a,*], Luigi Di Caro [a], Llio Humphreys [b], Livio Robaldo [a], Piercarlo Rossi [c] and Leendert van der Torre [d]

[a] *Università di Torino, Italy. E-mail: {guido,dicaro,robaldo}@di.unito.it*
[b] *Università di Torino, Italy and University of Luxembourg, Luxembourg. E-mail: humphreys@di.unito.it*
[c] *Università del Piemonte Orientale, Italy. E-mail: piercarlo.rossi@unipmn.it*
[d] *University of Luxembourg, Luxembourg. E-mail: leon.vandertorre@uni.lu*

Abstract This paper describes the Eunomos software, an advanced legal document and knowledge management system, based on legislative XML and ontologies. We describe the challenges of legal research in an increasingly complex, multi-level and multi-lingual world and how the Eunomos software helps users cut through the information overload to get the legal information they need in an organized and structured way and keep track of the state of the relevant law on any given topic. We describe the core system from workflow and technical perspectives, and discuss applications of the system for various user groups and our long term vision towards an Internet of Social Things, where laws can have an identity and be manipulated adding interpretation and can proactively inform interested users of their changes over time.

Keywords: Legal document management, legal ontologies, classification, knowledge acquisition and concept representation on annotations and legal texts

1. Introduction

1.1. Goal of the paper

We live in a complex regulatory environment. The body of law to which citizens and businesses have to adhere to is increasing in volume and complexity as our society continues to advance. Laws become more dynamic, more specialized and cover more and more areas of our lives. Paper-based methods of dealing with laws and regulations are no longer fit for purpose, but making them accessible online is not sufficient either.

This paper presents Eunomos, a legal document and knowledge management system. Differently than other systems, it firstly recognize the need for a stricter coupling between legal knowledge and its legislative sources, associating the concepts of its legal ontology with the part of regulations defining them, structured using legislative XML. On the one hand, this solution faces the utopia of pretending that the simple availability of the text of laws online solves the practical problems of citizens and business. On the other hand, it

*Corresponding author. E-mail: guido@di.unito.it

allows to ground concepts of legal ontologies to their sources, making ontologies more acceptable to practitioners and synchronizing their meaning with the evolution of the text of the law across its modifications.

First of all, we present in this section the two sides of the problem: the increasing burden of dealing with regulations and the complexity of the meaning of laws.

1.2. Growth of the law

The body of law to which citizens and businesses have to adhere to is increasing in volume and complexity as our society continues to advance. Laws become more specialized and cover more and more areas of our lives.

The law is increasing in level of specialisation as advanced multi-level societies require domain-specific laws for different areas of our lives. But in most legal systems, laws are not clearly classified, and some laws contain norms on more than one legal domain. The extent of the law over our lives is also increasing as the administrative and technological instruments at the disposal of the State allows for more control of individual and business behaviour. Another development is that we are becoming increasingly subject to multi-level jurisdictions. In the United States, "large corporations operating in multiple jurisdictions often need to conduct a so-called '50 state survey of the law' to identify and analyze different legal requirements on different topics." [31]. In Europe, due to subsidiarity, laws are applicable from European, national, regional and municipal levels. Italy now produces thousands of laws every year, with many pieces of legislation containing a number of norms on a range of different topics. Meanwhile, the European legislation is estimated to be 170,000 pages long. To these figures we must add internal regulations of firms. In Italy each bank employee is expected to know 6,000 pages of internal regulations[1].

Paper-based methods of researching laws and regulations are no longer fit for purpose. In many regions in Europe and beyond, there are now official online portals making laws and decrees available to all, due in no small part to the momentum gained by Open Government Data and Linked Open Data initiatives. Sartor envisages a future legal semantic web where legal contents on the web will be enriched with machine processable information. "This information will then be automatically presented in many different ways, according to the different issues at stake and the different roles played by its users (legislators, judges, administrators, parties in economic or other transactions)" (p7, [46]). However the heterogeneous ways in which legal data are published by public sector organisations - in terms of formats, structure, and language - inhibit this development. The current reality is that much time and effort can be spent searching multiple portals for regulatory provisions.

The laws are usually not classified in an intuitive way (for example, the Normattiva website of Italian national legislation will classify laws according to the Eurovoc scheme, which is based on the administrative structures of the European Commission). And some legislation portals do not contain clickable links to other referenced legislation. Legislations are full of cross-references, so this makes navigating laws most difficult. Lord Justice Toulson in R v Chambers (2008) (as quoted in Holmes [28]) expressed grave concern about accessibility of UK legislation: "To a worryingly large extent, statutory law is not practically accessible today, even to the courts whose constitutional duty it is to interpret and enforce it. There are four principal reasons. ... First, the majority of legislation is secondary legislation. ... Secondly, the volume of legislation has increased very greatly over the last 40 years ... Thirdly, on many subjects the legislation cannot be found in a single place, but in a patchwork of primary and secondary legislation. ... Fourthly, there is no comprehensive statute law database with hyperlinks which would enable an intelligent person, by using a search engine, to find out all the legislation on a particular topic."

Another problem is legislative updates. Some laws state explicitly which articles of other legislation are modified, others don't. This resulted in the parliamentary practice of 'implicit abrogation' of norms with regard to the temporal succession of laws. According to this principle, the more recent legislative norms will prevail, if it applies to same subject, whether or not they mention the overruled norms. In the end, the application of norms is subject to judicial interpretation on a case by case basis. Enrico Seta commented on this issue in World e-Parliament Reports 2008:[2] "In the Italian legal system what is really difficult for citizens, as well as for the interpreter (the judge), is to recognize the final legislation resulting from the con-

[1] Source: ABILab.

[2] United Nations, World e-parliament report 2008: http://www.ictparliament.org/es/node/687

tinuous, fragmentary and sometimes dispersed law-making process. This activity may involve the comparison of many acts and of explanatory notes, given that in the Italian legislation only very few consolidated codes are present." Delegification (attributing power to amend legislation to other institutions besides the parliaments on some topics) makes the situation even worse. The Italian Parliament occasionally does produce official consolidated codes. But most of the time, this work is left to independent agencies, whose interpretation does not have official status.

Meanwhile, also due to the above difficulties, failures in the legislative drafting process have resulted in legislation that continue to refer to norms that have since been overridden: e.g., in the US, "ADAAG references the A17.1 elevator code for conformance. Since 2000 there has been no section of the A17 that references lifts for the disabled. Therefore ADAAG references a non-existent standard" (an example by Lau [31]).

1.3. Understanding the law

Many of the above problems are intrinsically connected to the functioning of legislative rules, and can be seen as problems of accessibility and retrieval. Once legislation is retrieved, there are then issues of understanding. Legislative language is notoriously difficult to understand. Some terms, understood as "terms of art" have acquired meanings from statutory definitions and scholarly or judicial interpretations that differ from their meaning in ordinary language. It is not always clear where to find the correct meaning for the term because legal interpretations often gain acceptance with professionals before influencing subsequent definitions in legislation. Polysemy is a significant problem in legal terminology, because we have the added complexity that legal terms can have significantly different meanings across jurisdictions, within contexts and over time. Thus, the meaning of a term is unavoidably related to the legislation it appears in and to its subsequent modifications: meaning and text are coupled together. Legislation can also be intentionally vague sometimes in order to allow for social and technological changes. A clear example from the IT Law sector is provided by Breaux [15] in HIPAA 164.512(e)(1)(iv) which "states that an entity must make 'reasonable' efforts to notify individuals of certain requests for their protected health information. The word "reasonable" is an intended ambiguity: exactly which mechanisms are considered reasonable, (e.g., postal mail, secure electronic mail or websites, etc.) varies depending on the type of communities served and the prevalence of relevant, existing technologies". Some problems of legal language derive from the imprecise nature of language. The Supreme Court[3] advises that in cases of attributive ambiguity, legislative intent may override literal interpretation:

> Ordinarily, as in everyday English, use of the conjunctive "and" in a list means that all of the listed requirements must be satisfied, while use of the disjunctive "or" means that only one of the listed requirements need be satisfied...however; if a "strict grammatical construction" will frustrate evident legislative intent, a court may read "and" as "or" , or "or" as "and."

Thus, the possibility to access to legislation is not sufficient, if also interpretation or interpretative sources are not available.

Finally, the ubiquitous use of cross-references in legislative text can also lead to problems, not only in readability, but also in determining which parts of a referenced article are relevant.

1.4. Research questions and methodology

These issues in accessibility and interpretation of the law are present in many legal orders. In summary, difficulties of accessibility arise because:

- the law is increasing in scope, volume and complexity;
- there are many specialist areas of laws and they are frequently not classified intuitively on official legislative portals. Some legislations contain norms on a range of different subjects;
- legal norms can come from different sources - regional, national or supra-national authorities, all of whom have their own official portals with different ways of presenting legislations;
- some legislation modify or override existing norms but do not explicitly say so. Where modifications are explicit, available legislations are often not consolidated with updates and modifications by subsequent legislations.

Difficulties of interpretation arise because of:

- legislations contain many legal "terms of art" whose meaning are not always made explicit in the legislation;
- many "terms of art" acquire different meanings in different contexts and over time;

[3] http://www.fas.org/sgp/crs/misc/97-589.pdf

- legislative text can vague and ambiguous, often intentionally, in order to allow for social and technological changes; problems of interpretation can derive out of the imprecise nature of language itself;
- legislation are full of cross-references, but the referenced articles are not quoted, and some legislation portals do not contain clickable links to other referenced legislation

These problems have significant consequences for society. They affect the freedom of citizens, the efficiencies of organisations and the compliance of business. The cost of clerical, research and professional legal work is high for law firms, financial institutions and public administrations. For regulatory compliance of enterprises, there is a real risk that legal experts might miss important information and misinterpret the law, resulting in significant costs in legal payments and reputation.

Lately, articles have begun to appear in specialist[4] and even mainstream[5] press about an increased interest in bespoke ITC solutions, and in particular, human language technologies, for legal domains. But how much is the demand in reality? And do these technologies actually address the challenges and problems of legal research? Yet, legal informatics, despite decades of research, is rarely applied in the commercial or legal world.

These difficulties are one of the reasons for the IT/Law alignment problem. There have been progresses to cope with this issue, but there are remaining challenges. Thus, to make a further step in the achievement of IT/Law alignment, the **research question** of this paper is:

> How to create a document and knowledge management system based on technologies from legal informatics to help address the above problems in accessing and interpreting the law?

The methodology we use is to take inspiration from technologies developed in the related fields of legislative drafting for parliaments (so called legislative XML) and legal ontologies extending the tool for building legal ontologies called Legal Taxonomy Syllabus by Ajani et al. [5]. We export these technologies in the context of applications for legal researchers and practitioners.

In the next section we provide as background a description of the growth of such technologies in legal informatics. In Section 3 we describe the main functionalities of the software and the workflow of users and knowledge engineers. In Section 4 we describe the technologies used and how we are starting to address the resource bottleneck using human language technologies, in this case, text similarity and a semi-automated classification mechanism. Section 5 describes the different uses of Eunomos for the financial sector, the legal profession, the public sector and citizens. Moreover, it describes our long term vision towards an Internet of Social Things, where laws can have an identity and be manipulated adding interpretation and can proactively inform interested users of their changes over time. Future and related work, and conclusions end the paper.

2. Legislative XML and legal ontologies

Legal informatics is the application of information technology to the legal domain, and includes technologies for storing and retrieving legislation, traversing legal terminology, representing norms in logical form as well as automated reasoning and argumentation. In this paper we focus on the technologies we adopt: legislative XML and legal ontologies.

2.1. Legislative XML

One of the greatest successes of this field of research is the growth of legislative XML, which has now been developed for several jurisdictions. XML is a hierarchical, rigorous, extensible, accurate and flexible language (or rather a meta-language) whose vocabulary of tags can be built for each community depending on the problem to be solved. At the same time, XML is rigorous in that it uses a lexicon, syntax and grammar which defines its rules. These rules define the behaviour of a tag (for example, that all paragraphs should be numbered), and this behaviour cannot be violated by the user. The NormaInRete standard is well-established XML standard used by many regional governments in Italy for the management and publication of legal documents online. The NormaInRete XML standard has been introduced in 2001 to provide wider electronic access to national and regional legislation and allow greater interoperability between government

[4] http://legalinformatics.wordpress.com/2009/08/07/susskind-on-the-end-of-lawyers/

[5] http://business.timesonline.co.uk/tol/business/law/ article7003373.ece

departments and institutions. It specifies a method for the description of legal sources, with a naming convention for their identification using the mechanism of Uniform Resource Names (URNs) (see Section 4.2).

Legislative XML formats have been developed in several jurisdictions. European examples include LexDania in Denmark, CHLexML in Switzerland, and eLaw in Austria. Although each legislative body has its own unique characteristics, they also have several characteristics in common such as actors, structures, procedures, documents and information. As a result, the Metalex interchange formats has been developed in Europe while the Akoma Ntoso Legislative XML standard [39] has been designed to be sufficiently flexible to be suitable for all African legislative bodies at national and regional levels. Akoma Ntoso was created in created in 2004 and was much influenced by the NIR standard. It has become popular beyond Africa and is the basis of LexML in Brazil. The Akoma Ntoso standard applies to all parliamentary documents produced by a legislative body, such as proposed legislation, registration of debates, drafts, reports, and agendas. It is extensible and customizable, adaptable to each local situation without sacrificing interoperability between systems.

2.2. Legal ontologies

Legislative XML provides a standard method for structuring legislation to aid the management and retrieval of norms. It does not help with semantic analysis of such information. Legal ontologies are a valuable resource in semantic analysis. Several anthropological and psycholinguistic studies support the intuitive design of ontologies as an excellent way for people to understand the relations between concepts. Top-down ontologies start from fundamental legal concepts defined in legal jurisprudence and proceed to narrower concepts. Bottom-up ontologies describe terms extracted from legislation or case law in specific domains. There are now several real-world projects that use ontologies.

Fernandez-Barrera and Casanovas's ONTOMEDIA project [23] adopts a bottom up approach, providing basic legal and judicial resources to citizens involved in consumer mediation processes. Users select their region and can query relevant norms on consumer law for their region. Citizens will be able to present their problem in natural language and be directed to relevant information available online. This functionality is based on mapping user representation of a problem to a regulative representation of the problem using information leaflets that explain regulations in normal language as an intermediary conceptual system. Their methodology is based on extraction of terms in everyday language from a corpus of consumer queries and enrichment of specialist ontologies on mediation and consumer law with the extracted terms from the consumer queries.

Cherubini and Tiscornia's Pubblica Ammistrazione e Stranieri Immigranti (P.A.eS.I.) [35] is a portal on immigration procedures. The ontology-based computable model of the normative framework helps immigration services as well as non-Italian citizens to find the information they need. Information is organised along 'life events' in which the individual or enterprise is involved e.g. gaining citizenship, employment and access to health services, with information sheets on each topic written in clear and plain language. About 230 procedures are mapped to related legislative norms, allowing citizens and organisations to query what they must do on the basis of which norms.

The ontology used in Eunomos is based on our Legal Taxonomy Syllabus [5,2]. The tool is based on a clear distinction between the notions of legal term and legal concept. The basic idea is that the basic conceptual backbone consists in a taxonomy of concepts (ontology) to which the terms can refer to express their meaning. One of the main points to keep in mind is that the Legal Taxonomy Syllabus does not assume the existence of a single taxonomy covering all languages. In fact, it has been convincingly argued that the different national systems may organize the concepts in different ways. For instance, the term contract corresponds to different concepts in common law and civil law, where it has the meaning of bargain and agreement respectively.

The traditional top-down approach to the development of ontologies as described by Visser and Bench-Capon [51] is not flexible enough in legal ontologies. Usually, ontologies are built starting from very general concepts which are then specialised in more detailed concepts. Moreover most ontologies are oriented to a single national tradition. In this process the knowledge engineers risk not to take into account the interpretation process of the legal specialists on the real multilingual data. These ontologies aim at modelling the legal code but not the legal doctrine, that is the work of interpretation and re-elaboration of the legal code which is fundamental for transposing EU directives into national laws. The philosophy of the Legal Taxonomy Syllabus is a two-step procedure pursued in the

UT project [4,44]. UT project (Uniform Terminology For European Private Law) is a Research Training Network (RTN) funded by European Commission[6]. The research network involves researchers from 7 universities spread across England, France, Germany, Italy, Netherlands, Poland, Spain. The results achieved by the Network can be divided between those relating to a better understanding of the historical divergences hampering uniform terminology, and those relating to the promotion of a common terminology in EU private law. As a first step, terms are collected in a database together with the legal sources where they appear, in order to identify the concepts. Then, for each different ontology (i.e., each specific language ontology and the general EU ontology), the set of concepts is organized in an ontology which can be different for different legal traditions. This reconstruction work is done by legal experts rather than knowledge engineers. In this phase the result is a lightweight ontology rather than an axiomatic one. Only relations among terms are identified without introducing restrictions and axioms. The function of these ontologies is to compare the taxonomic structure in the different legislations, to provide a form of intelligent indexing and to draw new legal conclusions. In a second phase, a knowledge engineer can reorganize the ontology and integrate it with a top-level well-founded ontology like DOLCE [24].

Another feature of the Legal Taxonomy Syllabus developed in Ajani et al. [2] is the ability to model the evolution of the meaning of concepts over time, depending on the amendment of the legislation defining them. When a new normative is approved and enacted it can define a number of new concepts; moreover it can happen that the same law can change a number of old concepts defined by old laws. In particular, these old concepts can become obsolete and no longer valid. We are aware of the difficulties concerning the modelling of time in artificial intelligence and in formal ontology creation. In the first version it was necessary to delete all old concepts, causing the loss of all historic information from the database, information that is quite valuable for a better understanding of the evolution of the normative. This problem was resolved by empowering Legal Taxonomy Syllabus with a new ontological relation called REPLACED BY. When the paragraph of a text defining a concept has been modified by a new legislation, the new one defines a new concept that will replace the old one in the ontology. There will be a relation of type REPLACED BY between the two concepts. Also in this case the new ontological relation has some peculiar characteristics that distinguish it from the usual ontological relations: First, a REPLACED BY relation brings with it a new data field not present in the other relations: the substitution date. Second, when the user performs a search in the concepts database the replaced ones will not be shown, unless the user asks for a certain past date, thus obtaining a snapshot of the legal ontology that was valid at that point. When a new concept replaces an old one, all the ontological relations in which the old concept participated in are automatically applied to the new concept. If some of them are no longer valid with the new concept, manual intervention from the user is required.

Many resources developed in the research field such as ontologies and automated reasoning systems are abandoned because they require prohibitively extensive manual annotation. Advances in natural language processing tools such as part-of-speech taggers and parsers, the growing usage of statistical algorithms for handling uncertainty and the availability of semantic resources such as WordNet and FrameNet, potentially provide opportunities for automated information extraction to help develop such resources. But legal language is not natural language, and the same issues that pose problems for human understanding also create difficulties for machine processing of legal text. Building user-friendly, sustainable and reliable applications for managing legal information is not easy. It requires real understanding of legal research and discrimination in the use of legal informatics technology to ensure that solutions are useful, reliable and cost-effective.

3. Eunomos - the core system

3.1. General Overview

The Eunomos online legal document and knowledge management system described in this paper, developed in the context of the ICT4LAW project[7] was created to help legal researchers and practitioners manage and monitor legislative information. The system is based on mature technologies in legal informatics - legislative XML and ontologies - combined in an intuitive

[6]Contract n. HPRN-CT-2002-00229

[7]ICT4LAW: "ICT Converging on Law: Next Generation Services for Citizens, Enterprises, Public Administration and Policymakers" funded by Regione Piemonte 2008-2013.

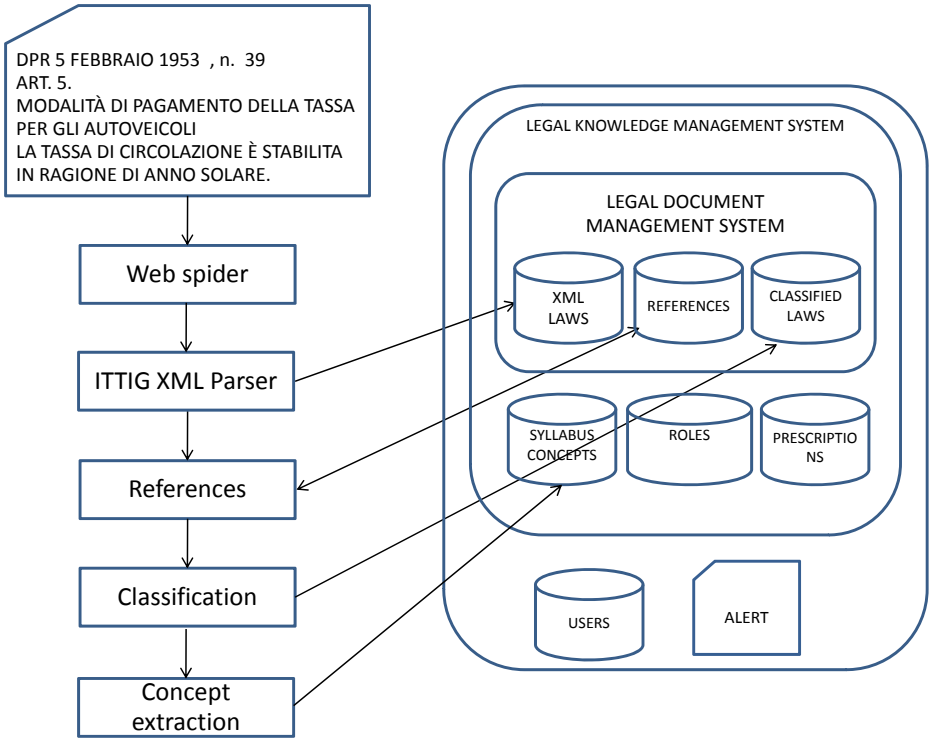

Figure 1. Key components of the Eunomos system

way that addresses requirements from the commercial sector to access and monitor legal information. Less developed technologies, such as logical representation of norms and information extraction of legislative text are not used now but may be in the future. Eunomos can be employed as an in-house software that enables expert users to search, classify, annotate and build legal knowledge and keep up to date with legislative changes. Alternatively, Eunomos can be offered as an online service so that legislation monitoring is effectively outsourced. The software and related services can be provided to several clients, which means that information and costs are shared. The Eunomos system is the basis of the Menslegis commercial service for compliance distributed by Nomotika s.r.l., a spinoff of University of Torino.

The system, being based on the Legal Taxonomy Syllabus, it is inherently multilingual and multilevel, so it can be used for different systems, using similar legislative XML standards, and even for the EU level, keeping separate ontologies for each system. In this paper, however, we will describe the current application for Italian legislation.

As stated in the Introduction, the basic idea of Eunomos is creating a stricter coupling between legal knowledge and its legislative sources, associating the concepts of its legal ontology with regulations structured using legislative XML.

The legal document management part of the system is composed of a legal inventory database of norms (about 70,000 Italian national laws in the current demo) converted into legislative XML format, with links between related legislation created by automated analysis of in-text references and each article semi-automatically classified into legal domains. Most laws are collected from portals by means of web spiders on a daily basis, but they can also be inserted into the database via a web interface. Currently the system harvests the Normattiva national portal[8], the portal Arianna of Regione Piemonte[9] and a portal of regulations from the Ministry of Economy. For each legislation, Eunomos stores and time-stamps the original and most up-to date versions, but nothing prevents includ-

[8]http://www.normattiva.it
[9]http://arianna.consiglioregionale.piemonte.it/

ing multiple versions of the coordinated text for users, like lawyers or judges, whose primary concern is not only to have up-to-date information on the law.

After they are converted into legislative XML, cross-references are extracted to build a network of links between norms citing one other. The semi-automated classification of norms is supported by classification and similarity tools described in Sections 4.3 and 4.4. Legal concepts can be extracted and modelled using the legal ontology tool Legal Taxonomy Syllabus, the specialist multilevel and multilegal ontology [3] for terminology management of European Directives and their national implementations described in Section 2.2. The ontology part of the database is saved as a table that is a repository of concepts, that are connected, but independent from, terms in a many-to-many relationship. The classical subject-predicate-object triple that defines the relationships between the concepts is stored in a separate table. Reconstructing transitive relations can be expensive in a relational database, so there is another cache table that stores the complete transitive closure of the ontology.

The ontology is well-integrated within the document management system, so that links can be made between concepts, the terms used to express the concepts, and items of the laws that feature the terms. Viceversa, terms in the text of legislations are annotated with references to the concepts. Figure 1 shows the components of the system and the flow of documents into the system. More technical details are discussed in Section 4.1.

In summary, the architecture of the system is composed of three levels:

- The proper legal document management system, composed of a database of norms in legislative XML, a database collecting the network of references between laws, using their unique processable identifier called URN (see Section 4.2), and a database classifying single articles or items of legislations in different domains. This is possible since the legislative XML provides a unique identifier not only to legislations, but also to its parts like articles and items.
- The legal knowledge management system composed of a database of concepts and of relations connecting them, together with the terms associated to concepts. This database is connected with the legal document management system to associate concepts and articles or items of legislations. Moreover, a database of prescriptions (obligations) and associated roles are present, as discussed in Boella et al. [12]. This component is outside the scope of the paper (see Section 5.1).
- The external tier is composed of a database of user profiles for login purposes and for keeping information about their domains of interest. It is also in charge of dispatching to users the alerts concerning updates in legislation of interest to them.

The population of the databases proceeds in the following way. Web spiders collect daily new legislation, identified by their URN identifier obtained by translating the human language title of the law. Then the text of the norm is automatically translated into legislative XML using a parser. References in the text of norms, already tagged in XML, are collected in a database. Then norms are classified semi-automatically, and the collection of concepts can start. In Section 3.2.2 we describe in detail the role of the knowledge engineer in this process.

3.2. Workflows

We will describe the features of Eunomos presenting two possible workflows: the one of the user and the one of the knowledge engineer. These two workflows are of particular importance: the former to ensure the acceptability of the system for legal researchers and practitioners, the latter to ensure that the cost of producing knowledge in the legal field is manageable.

Eunomos provides a web-based interface for users and Eunomos knowledge engineers to find information about laws and legal concepts in different sectors and different jurisdictions.

3.2.1. User Workflow

The Eunomos system is useful for surveying the law on a particular topic and read its interpretation. Alternatively, it can be used starting from the ontology to understand the basic concepts and navigate from the concepts to the legislation. Users can select their domain of interest, and then search for relevant legislation, since legislations are classified in a number of domains, at the level of article or even of item and paragraph in case of legislation containing articles belonging to several domains. They can refine their legislation search with keywords, index number, year, quoted text from legislation or from user comments associated with elements of legislation (see Figure 2). All versions of a law stored are retrieved, unless dates of va-

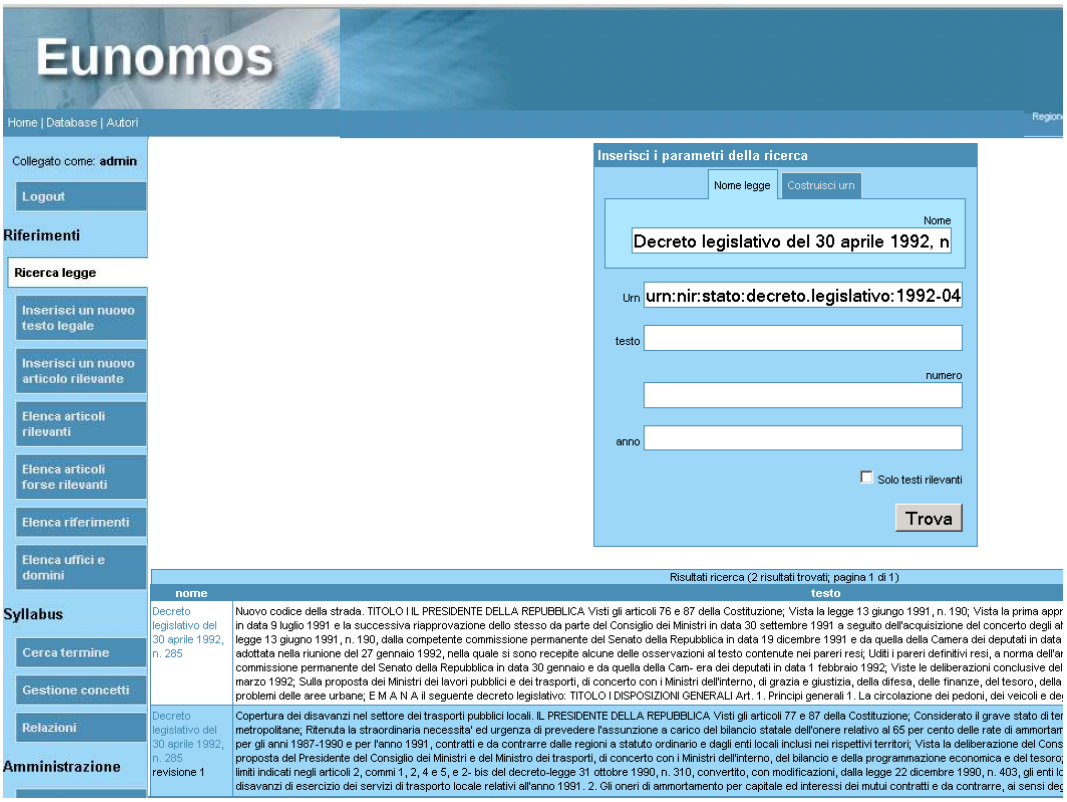

Figure 2. The search interface of Eunomos

lidity are restricted. Any relevant laws will then appear in a table in chronological order.

Clicking on an item in the table brings the selected legislation into view, remaining in the selected domain of interest, which can be changed at any moment. Users can click on different options to view useful information about legislation:

– The *Testo* (Text) option shows the full text of the legislation in HTML, PDF or XML as selected. Users can choose whether to view the legislation in its original form or as coordinated text (where these are available from institutional portals or inserted manually), i.e., modifications via subsequent legislation to norms in the legislation in question are inserted into the text of the modified legislation and a new version is uploaded[10]. References in the text to other articles or other legislation are automatically linked to the relevant articles or legislation using URNs that conform to the NormaInRete standard. Users can click on the link to view the referenced legislation in HTML. Alternatively, they can hover their mouse over the link, and the relevant article appears in a preview text box. To aid readability, cross-referenced text appear in pop-up text boxes as users hover the mouse of the cross-reference. Alternatively, users can go directly to the relevant article in the referenced legislation by clicking on the cross-reference hyperlink.
– In cases where legislation covers a number of norms for various domains, it is useful for users to be able to view only articles relevant to the domain in which they are interested. The *Leggi o articoli rilevanti* (Relevant laws or articles) option provides a list of articles in the selected legislation relevant to the domain selected by the user. Users can click on relevant articles to view the text or hover their mouse to see the article in a text box.
– the *Riferimenti importanti* (Important referencess) option provides a list of cross-references between

[10] Although consolidated text from state portals are not usually formally approved by Parliament and thus do not have legal status in themselves, they are the most authoritative consolidated text available.

a particular legislation and others in a separate page, with hyperlinks to relevant articles. This feature is useful for keeping track of legislative updates and modifications and to navigate to related legislations in the same domain.
- The *Leggi simili* (Similar laws) page is also useful for a legal researcher to obtain an overview of the context of the legislation. It is based on text similarity (see Section 4.3).
- the *Parole chiave* (Keywords) option brings a list of domain-specific concepts from the ontology whose associated terms appear in the visualized legislation. In the future, users will be able to click on the terms and go to the appropriate definition from the ontology, due to a sort of automated wikification, associating concepts to the text via links. Preliminary results on linking the text to ontology concepts are described in Boella et al. [9]. For a legal researcher who is seeking clarification on meaning and usage of terminology, a list providing all contexts in which the terms are used within the legislation under consideration can be most useful. In the future, users will be able to click on the terms and go to the appropriate definition from the ontology. For now, they can conduct a terminological search by clicking on Legal Taxonomy Syllabus from the same web interface.
- Registered users can add their comments on single articles or items of legislations.

The alternative use of Eunomos by users is by starting from the ontology and navigating it till reaching the desired legislation.

As described in Ajani et al. [3] each concept is associated with the terms expressing it, the language of the terms, jurisdiction, definitions and explanations in natural language, and links to the article or items of the laws that contribute to the definition of the concept. Users can view a previous definitions of the term that apply to older legislation as discussed in Section 2.2. The descriptions in natural language are made by legal knowledge engineers, taking into account the interpretation given by legal scholars. The notes field carries information about court decisions, scholarly interpretations or other information of interest. The specific difference of Eunomos is that the link to the XML version of the legislation is done via a URN identifiers, while in the original Legal Taxonomy Syllabus there was only an hyperlink to the textual version of the norm.

The concept search is an alternative way to do this. The user clicks on *Cerca termine* (Search terms), and then inputs a term. The results are all concepts expressed by that term. The user then clicks on the appropriate row, and sees which legislations are relevant for that concept. The user can also click on the *Mostra Ontologia* (Show ontology) to view the structure of the ontology involving the selected concept. Each domain-specific ontology within Legal Taxonomy Syllabus is hierarchical and the conceptual tree allows users to view hyperonomy/meronomy/synonomy relations. Figure 3 below shows a concept tree for vehicles with the hyponyms being trolley-buses, motorcycles etc.

Users also needs to keep up with the law. The Eunomos alert messaging system monitors legislative changes for them. When a law or concepts relevant to their domains of interest is inserted in the database by the knowledge engineer, users are notified. But we have also more just in time updates relying on the above mechanisms of reference analysis and text similarity: when a newly downloaded or inserted legislation refers to some article classified in the domain of interest of the user, or it is close to it according to text similarity, the user is alerted as well.

3.2.2. Knowledge Engineer Workflow

Given the challenges described in Sections 1.2 and 1.3, knowledge engineers are essential to maintain a reliable service and provide additional information where needed.

Eunomos knowledge engineers, in summary, are responsible for:

- checking the output of the legislative XML parser and correcting any errors arising out of irregular patterns in the text;
- inserting missing legislations in the database;
- classifying the domain of legislative norms selecting among the suggestions proposed by the automated classifiers;
- classifying the type of modificatory references;
- adding concepts and terms to the ontology;
- adding explanations in plain language of terms and legal obligations;
- adding relevant information from case law or scholarly interpretation.

To resolve the resource bottleneck, human language technologies are increasingly used at most of the above steps (see Section 4).

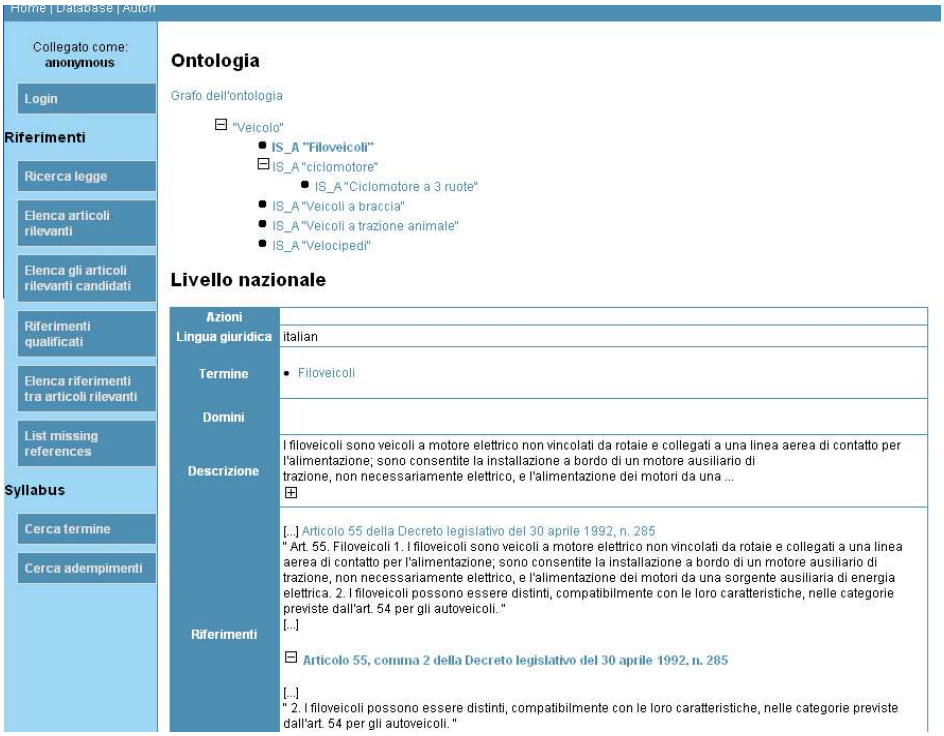

Figure 3. Legal Taxonomy Syllabus within the Eunomos system

Knowledge engineers have access to all the user interfaces as well as interfaces for adding annotations and populating ontologies.

It is of particular interest to describe how knowledge engineers can build knowledge about a particular legal domain starting with a well-known piece of legislation or searching for laws containing a particular domain-specific keyword from the database of laws. For legislation in new domains, each article has to be classified manually. As they work through the text, knowledge engineers also annotate cross-references and add terms to the ontology. The Eunomos system contains an automatic reference detection tool that automatically finds references to articles in other legislation and creates inline hyperlinks within the legislation text. Knowledge engineers then look at each explicit reference and denote its domain and its type: whether it is merely a simple reference or in fact modifies or overrides existing legislation (see Figure 4). They also check for cross-references missed by the parser due to irregular textual patterns by clicking on the *Riferimenti* (References) option which has a list of outgoing references created manually for the Normattiva website. Where legislation fails to mention which existing legislation it modifies or overrides, a knowledge engineer will need to find the connections and manually insert an implicit cross-reference. Moreover, Eunomos has an interface to make comments about legislation and all its paragraphs and articles. This feature is especially useful for annotating elements that have been implicitly modified or overridden by other legislation. The *Leggi simili* (Similar laws) list of the most similar legislation in the database, produced automatically by text similarity analysis, can be most useful for finding legislation implicitly modified by later legislation. Knowledge engineers can then also use this list to find other pieces of legislation belonging to the new domain, so that they can proceed to annotate these legislation as described above.

In Figure 4, we can see annotated articles from a piece of legislation. The knowledge engineer uses this interface to specify whether an article is relevant for the domain under consideration. The relevance for the domain has been preselected by the classification mechanism or by text similarity. Moreover, he can add a type (modification, suspension, etc.) for each reference to other legislation. Terms which are linked to concepts in the ontology of the relevant domain are highlighted to help the engineer understand the relevance of the article for the domain.

Figure 4. Annotating legislation

From the Eunomos interface, new terms and interpretations can be added to the ontology directly from the text of the law. In the Legal Taxonomy Syllabus project, to properly manage terminological and conceptual misalignment, a distinction was made between *legal terms* and *legal concepts*. The system consists of a taxonomy of unique concepts (ontology), to which any number of terms can refer to in order to express their meaning. Eunomos contains specific interfaces for managing and viewing terms and concepts. The *Crea concetto* (Create a concept) page enables a knowledge engineer to create a new concept starting from the description of it directly in the text of the legislation, so that such legislation is automatically linked to the concept. Then he can add metadata such as language, jurisdiction, date, description, notes and further references to legislation defining the concept. This process is illustrated in Figure 5.

Knowledge engineers are also active when new legislations are issued. When the system has a number of alrady classified legislations to learn from, a statistical classifier is used to determine the domain of each article. The knowledge engineer then checks the domain selected by the domain classifier. Usually, and particularly for well-populated domains, the classifier will select the correct domain for each article (see Section 4.4). Results are provided by Boella et al. [8,9]. Legislative articles are more difficult to classify than other text due to overlaps in vocabulary and articles which contain no real content except cross-references, so the knowledge engineer may need to resort to other supporting tools for this task: text similarity, prevalence of domain-specific terminology, and analysis of incoming and outgoing references. The *Leggi simili* list of similar legislation can give a good indication of the domains that are relevant for the new legislation. From the perspective of each relevant domain one by one, the *Candidati articoli rilevanti* (Candidate relevant articles) option provides a list of articles that could well be relevant to the domain on the basis of links to legislation classified as belonging to the domain in question. The rationale is that where paragraphs or articles contain references to classified paragraphs or articles in existing legislation, it is more than likely that the new paragraph or article belongs to the same domain. If the reference is to a particular article from the same domain, the evidence is labelled as strong. If the reference is to a piece of legislation which contains articles from the same domain as well as other domains, the evidence is labelled as weaker. The *Parole chiave* (Key-

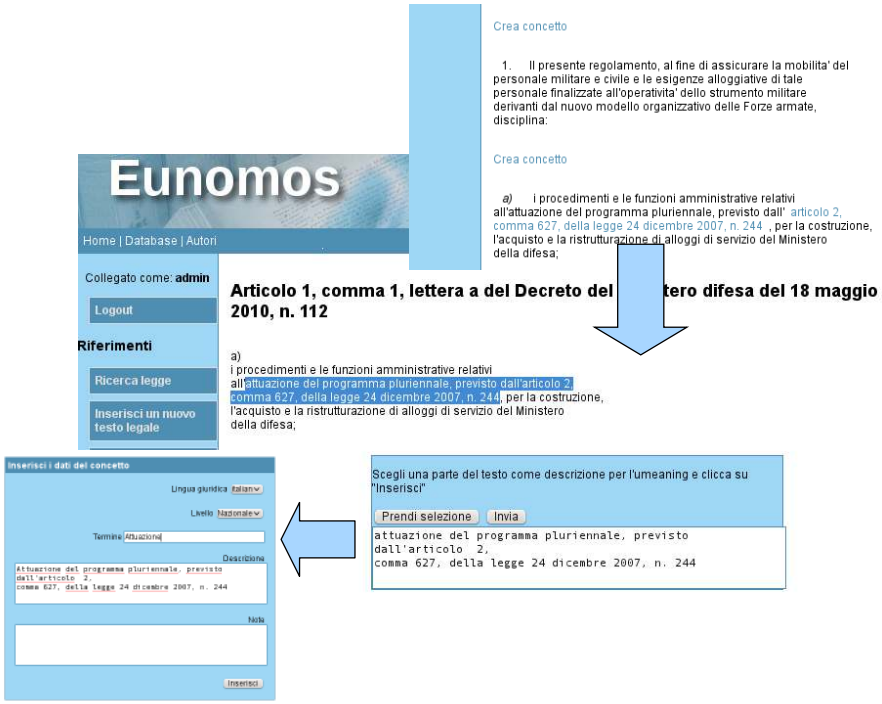

Figure 5. Creating concepts.

words) can also be useful for identifying relevant domains. The Eunomos ontologies are populated and updated manually, but once a term has been manually associated with a particular domain, the system ensures that all instances of domain-specific terms are highlighted in yellow when legislation is viewed from the perspective of the relevant domain suggesting that the encompassing article belongs to that domain. The *Parole chiave* (Keywords) list of all articles containing each term in the ontology found in the legislation can be useful for finding any new definitions or usage that needs to be recorded in the ontology. The knowledge engineer can also add new terms to the ontology, check explicit references and add implicit references as described above.

Once this work is complete, an alert message can be sent to all users who are noted in the system as being interested in the domain in question notifying them of new legislation and any modifications made to existing legislation.

4. Technologies

4.1. System Architecture

The system is implemented in PHP for the web application, Javascript and Ajax, for the front end, XML and XSLT for the documents, and C++ for the web spiders retrieving legislations. All the data, XML and ontologies included, are stored in the PostgreSQL relational database, which supports also XML. The database architecture is divided into two independent parts, managing the Legal Taxonomy Syllabus ontology (see Ajani et al. [3] for details) and the legal text repository. The Eunomos database of norms and legal concepts is accessible to any number of users via a web-based interface with secure login. Knowledge engineers also edit the data via the web interface. The web application to the system is divided into three parts:

– the pure presentation, using the Smarty[11] template engine;

[11] http://www.smarty.net

– a level, implemented in a set of PHP classes, that manages the input and the output to and from the templates; and
– the core business logic, involving another set of PHP classes that manages the input and output to the underlying database, supporting operations such as inserting a concept in the ontology or searching the legal text repository for a particular phrase in the laws of a given year. Triggers of PostgreSQL are used to enforce consistency of ontology relations.

4.2. Legislative XML

Laws are converted into NormaInRete (NIR) XML format using the Institute of Legal Information Theory and Techniques (ITTIG)'s XML parser[12] if they are in pure textual format.[13] Maintaining laws in NIR XML format makes it easier for Eunomos to extract elements such as paragraphs, articles and references so that knowledge engineers can categorise and annotate the elements, and lawyers can view specific relevant information. Within the Eunomos database, the unique identifier for each legislation and elements within legislation is the URN. URNs facilitate the construction of a global hypertext among the legal documents in a network environment with computer resources distributed among several publishers. It also allows the construction of knowledge bases containing the relationships between these documents.

An URN for a document constructed according to the NIR standard will have the following components:

– An ID for the original document, comprising the authority responsible for publishing the law (e.g., Ministry, Region, City, Court), the type of measure (e.g., law, decree, order, decision, etc.), the date and number and IDs for any annexes.
– A version identifier, including the date of issue.
– The ID of the press publishing the law.
– An identifier of the fragment of the resource itself the URN refers to (e.g., article, paragraph, etc.).

The URN for a particular document can be used in an XML or HTML file, e.g.:

```
<urn valore=
    "urn:nir:stato:legge:1996-12-31;675"/>
```

[12] www.xmleges.org

[13] The Arianna portal already exports documents to NIR XML format.

The segment of Figure 6 shows an article which modifies existing legislation. The URN address of the modified legislation is provided in the header section denoted by the <inlinemeta> tag. We have included a small part of the article to show the references to the URN addresses being used within the article text.

Eunomos uses the XML Leges Linker tool developed by ITTIG to find cross-references, an URN name resolver to obtain actual addresses of legislative articles, and XSLT to find and display outgoing and incoming hypertext links.

4.3. Text similarity

Eunomos uses a text similarity algorithm, the Cosine Similarity, to find the most similar pieces of legislation in the whole database. Since each piece of legislation contains a lot of text, they are indexed with the PostgreSQL internal inverted index facility in order to enable fast full text searches and ranking for document similarity. The Cosine Similarity metric uses the *Term Frequency-Inverse Document Frequency* (TF-IDF) measure to gauge the relative weight to be apportioned to various key words in the respective documents. The Cosine Similarity metric is particularly useful for finding similar single-domain legislation. However, legislation that contain norms on different topics can introduce noise into the comparative process. We are now adapting the software to include similarity searches on an article level.

As part of this work, we have looked at the issue of how to determine how many articles to select and return by identifying where the similarity values suffer a *significant fall*. This separates real similar articles from the rest. Results are discussed in Boella et al. [8].

4.4. Text classification

Even if the technicalities of the classification process we use is outside the scope of this paper, we summarize here our methodology, described in details in Boella et al. [8,9].

For each new piece of legislation, the classification task is: 1) to find which domains are relevant to the legislation, and 2) to identify which domain each article belongs to. The first task enables targeted email notification messages to be sent to all users interested in the particular domains covered by new legislation. The second task enables users to view, in each piece of new legislation, only articles relevant to a particular domain. Although there are plenty of algorithms for text

```
<articolo id="art1" xml:lang="it">
 <inlinemeta>
  <disposizioni>
   <modificheattive>
    <dsp:sostituzione implicita="no">
     <dsp:pos xlink:href="#art1-com1" xlink:type="simple" />
     <dsp:norma
       xlink:href="urn:nir:stato:regio.decreto:1942-03-16;267:legge.fallimentare">
       <dsp:pos xlink:href="#rif8"/>
     </dsp:norma>
     <dsp:novella>
      <dsp:pos xlink:href="#mod185-vir1"/>
     </dsp:novella>
    </dsp:sostituzione>
   </modificheattive>
  </disposizioni>
 </inlinemeta>
 <num>Art. 1.</num>
 <rubrica xml:lang="it"> Sostituzione dell'
  <rif id="rif7"
   xlink:href="urn:nir:stato:regio.decreto:1942-03-16;267:legge.fallimentare#art1">
   articolo 1 del regio decreto 16 marzo 1942, n. 267 </rif>
 </rubrica>
```

Figure 6. An example of NIR XML annotation.

classification, we used the well-known Support Vector Machines (SVM) for this task, since it frequently achieves state-of-the-art accuracy levels [20,29]. The association between legislation articles and a category label are fed to an external application based on the WEKA toolkit [27] and incorporated in Eunomos, creating a model that can be used to classify new laws inserted on a daily basis into the database by web spiders or users. The process of transforming text into vectors requires selection of suitable terms, and use of a weighting function as part of the frequency calculations. Again, we use the TF-IDF weighting function as proposed in [45], as it takes into account both the frequency of a term in a text and how characteristic it is of text belonging to a particular class. We use a dependency parser for Italian called TULE [32] that performs a deep analysis over the syntactic structure of the sentences and allows a direct selection of the informative units, i.e., lemmatized nouns. This is a better solution than the more common practice of using WordNet [36] or other top-domain ontologies to eliminate stopwords and lemmatize informative as they are unable to recognise and lemmatize many legal domain-specific terms. In Boella et al. [9] we have shown how to use the specific terminology of the ontology to identify relevant phrases in the text of norms so to add further features to the classifier, showing an increase of the performance, in particular in case of classes with few documents.

5. Applications

The Eunomos system is envisaged as being useful to a wide range of user groups. We have extended the core system for compliance officers in the first instance, because they have the greatest need and enthusiasm for a system of this kind, leading to the development of the Menslegis professional service mentioned above. To ensure we prioritise development according to business opportunities: compliance officers, the legal profession, public administration, the voluntary sector and citizens. In each case user scenario, the knowledge can be shared with several clients, lowering the cost of legislation monitoring and knowledge building overall. Another advantage of having several clients using the model is that with more people using the system, errors are more likely to be quickly detected and corrected. Putative links are verified by domain experts as a matter of course.

Eunomos can be employed as an in-house software that enables expert users to search, classify, anno-

tate and build legal knowledge and keep up to date with legislative changes. Alternatively, Eunomos can be used as a software to create services package, so that legislation monitoring is effectively outsourced.

5.1. The Financial Sector

Banks and insurance companies are required by law to ensure compliance with strict regulations. In Italy, all the compliance regulations are stipulated in Legislative Decree 231/01, a radical piece of legislation that changed the nature of legal obligations for banks and insurance companies. Now, such organisations can even be held responsible for criminal activities by employees they did not prescribe or authorise. Compliance with financial regulations is an extremely complex area of law, and there are not many legal experts in the field. The great complexities of Legislative Decree 231/01 is largely caused by a very chaotic and heterogeneous law. For example, the regulation of so-called 'Reati Presupposti' (presumed crimes) (Articles 24 et seq.), has always been characterized by continuous references (explicit and implicit) to articles in the Penal Code, Civil Code and the Code of Criminal Procedure, as well as articles in other legislation. Every year several clauses and sub-clauses are added to the legislation.

The stricter duty of care to ensure compliance with regulations means that financial institutions must adapt to continuous legislative changes to their legal obligations, and demonstrate that they have systems and procedures for searching for legal changes, and monitoring employee activities. The basic Eunomos system described above has been extended so that the ontology includes prescriptions on what the financial institution must do or not do to comply with the law, containing the following fields: deontic clause, active role, passive role, crime, sanction. A macro-prescription can also be stored which specifies a general principle and contains links to specific prescriptions that come under this principle [12,11]. The structuring of prescriptions in terms of concepts enables the user to make fine-tuned searches such as: list the prescriptions for which the concept of responsible has the active role. This will return prescriptions for all agents that can play the active role of responsible, like director, but also CEO or other ones. The relevant fields for active role (e.g. director), passive role (e.g. consumer), punishment (e.g. 1 year of jail) are all populated by concepts within the ontology and are linked to from the prescriptions.

5.2. The Legal Profession

The legal profession is a difficult market and yet is arguably in dire need of a system like Eunomos. To operate efficiently, a law firm needs to regularly create and update legal documents, access reliable information on the state of the law and keep track of changes in legislation and contracts. Currently much of this work is done by hand, even though the market shows clear signs that clients request IT solutions. Lawyers are reluctant to adopt Information technologies and use less IT than businessmen. Within law firms, IT is used mostly for accounting. Research in law firms and legal offices is conducted mainly by search engine keyword search, which is time consuming and achieves partial results. Law firms typically use folder trees as repositories and folder names as classification tags. The complex tree of folders and sub-folders is technically unrestrained, but law firms have guidelines on how to name folders and files. They use master contracts to help formulate actual contracts for clients, but no links are made between elements in master contracts and derived contract instances. Different versions of contracts are maintained using Microsoft Word's basic versioning features. But suggested amendments, and the motivation behind such amendments, frequently get lost in a trail of emails between those responsible for negotiating contracts, and their counterparts in the other legal firm. Since so much information is not recorded or maintained in a systematic way, knowledge and business can be lost as associates move to other firms and clients move with the main associate who handled their case. Legal document management also fails to address the need to continually review documents in the light of regulatory changes. This requirement means that various parts of documents need to keep track of the laws they refer to.

To address these requirements, the Eunomos system could be extended with a contract management repository that links to relevant legislation, using again legislative XML to structure the document and the ontology to represent the meaning of contracts.

5.3. The Public Sector

The infrastructure provided by Eunomos is also suitable for officers working for a wide range of public sector organisations. They may want to add a functionality to obtain laws and regulations that are not available from the main legislative portals, and new web spiders would need to be developed accordingly.

Since public sector organisations are not in competition for business and work together in certain domains, this presents opportunities for building knowledge in a different way. Organizations may wish to share they knowledge, as they are already doing using specialized forums, newsletters and mailing lists. But also they may wish to integrate their own taxonomies, or add interpretations of norms or concepts in the ontology based on their experience in a collaborative way. Given its online nature and its user management facilities, a Web 2.0 development of Eunomos is possible, making it a collaborative instrument for creating knowledge.

Tools would need to be developed to integrating other ontologies, with mechanisms for removing duplicates, resolving conflicts, and mapping terms to concepts. Integrating external knowledge carries risks for quality control. This problem is well known in the IATE terminological database, which is plagued by duplicates and inconsistency. Their solution is giving users the ability to posting comments to terminology experts concerning possible inconsistencies. It should be noted that while legal ontologies have been developed in the research community, they are usually too complicated for non-technical users and public organisations prefer to use taxonomies or thesauri, which require less training but are inadequate to deal with the complexity of different usages for terms. Eunomos's intuitive lightweight ontology would make it easy for non-technical expert users to add data. Users may also wish to raise questions or comment on new legislation as it comes in, assessing the impact on public organisations and changes necessary to their own procedures. The Eunomos system could be adapted to accommodate these requirements, by including threads or wikis to enable online discussions by various specialist user groups, and an expanded editorial service to manage user contributions.

5.4. Citizens

It is intended to provide a version of Eunomos for citizens in the future. Citizens will benefit from accessing not only the laws themselves but also explanations and definitions provided by Eunomos knowledge engineers and domain experts. This would overcome the delusion that making available to citizens the text of legislations could be useful to them.

The Eunomos citizen service could, for instance, help small voluntary sector organisations ensure that they understand and comply with health and safety regulations. With public funding, Eunomos could be extended to enable citizen participation on legislative proposals. The requirement to evaluate the "popularity" of laws among citizens and gauge the impact of laws on society is a stated objective of the ICT4LAW project and is already enshrined in Italian law (article 5 of law n. 50/1999). That law states that a Regulatory Impact Assessment has to be performed before enacting laws and consolidating provisions. Relying on explicit surveys is costly and often collects biased information. A better solution would be to obtain parliamentary debates or draft legislation, and attach threads to each proposal. Comments would be linked to the legislation (even relevant articles) to which they refer. Opinion monitoring software might in future be used to help provide first analysis of the comments, although Conrad and Shilder [19]'s survey into the scope and utility of opinion mining in legal blogs show the limitations of current opinion monitoring technology for complex topics such as the law.

5.5. Towards an Internet of Social Things

Besides the above applications, we see Eunomos as a step towards an innovative view of law, which is beneficiary of the recent technological developments.

This trend is witnessed also by other innovative initiatives. The UK legislation.gov.uk website publishes UK legislation, regional government statutory instruments county council and church acts in Metalex legislative XML format. The vision behind this portal is that opening their data allows other people to create products and services using the information they hold. Moreover, opening the website to collaboration by other institutions and people can help in improving the quality of the data [50]. Eunomos goes in this direction by leveraging on the data made available by Italian institutional portals, and where these portals are more advanced, like the case of the Arianna portal of Regione Piemonte offering data in NIR legislative XML format, the benefit for service provided by Eunomos is greater.

But legal informatics is not limited to legislation portals, but has a long tradition of tools for document management and knowledge representation. These tools allow users to find associations and meanings within related documents. Web 2.0 technologies enable such knowledge to be constructed by collaborative efforts. Web 3.0 - the so called Semantic Web - provides instruments to represent this knowledge in a way that can be processed using Artificial Intelligence techniques, like we do using the Legal Taxon-

omy Syllabus ontology. The latest development in Information Technology is the Internet of Things: evolving from a network of interconnected computers to a network of interconnected objects, from books to cars, from electrical appliances to food, and thus creating an Internet of things (IoT)' ([37], page 3). According to Sundmaeker et al. [49], page 43, '[p]hysical and virtual "things" have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network. In the IoT, "things" are expected to become active participants in business, information and social processes where they are enabled to interact and communicate among themselves and with the environment by exchanging data and information "sensed" about the environment...Interfaces in the form of services facilitate interactions with these "smart things" over the Internet, query and change their state and any information associated with them.'

If everything can be labelled and modelled, the possibilities for improving our everyday lives are endless. Norms are not part of physical reality but they are real as well, in the sense of object of a socially constructed reality, as envisaged by Searle [47]. Can norms too become living objects on the web? From the labelling perspective, they certainly can: as already mentioned, many norms already have their own IP addresses (or URNs). They can also be proactive, utilising web spiders to look for new legislation affecting them by explicit or implicit amendments. An Internet of Social Things would allow norms to be abstracted and interpreted while linked to their sources, and enable such knowledge to be built in a distributed way, as a result of distributed effort. This requires a combination of Web 3.0 and the Internet of Things. However, the key properties of norms are the concepts expressed in the norms and their inter-relationships. This enterprise requires not only machine readable legislation, but also the development of state-of-the-art legal knowledge management systems, like Eunomos, with linked data, multilevel ontologies, and advanced search facilities, to enable users to cut through the information overload and get the information they need in an organized and structured way.

Even if all these things can be modelled with certainty at a particular point in time, the nature of our legal system requires that norms must be allowed to evolve to reflect social changes. Updating norms as a consequence of amendments in subsequent legislation and interpretation is a continuous process that requires more complicated models than those that exist for physical things. By and large, current technology stops at the point of reporting to users that they may need to update norms or concepts as legislative changes are reported automatically. The Internet of Social Things presupposes a process of connecting parts (primary and secondary legislation) in a coherent whole which must be made coherent with other wholes in a systems of norms. The nature of the Internet allows us do this process in a distributed and reflective way, collecting and linking data automatically from different sources and allowing multiple actors to transform information into knowledge within a semantic framework, which can be distributed to a wider audience.

6. Future work

Eunomos is a stable piece of work subject to new developments. Our priority for future work is to use robust human language technologies that can help the work of the knowledge engineer, so to resolve the resource bottleneck problem. One development on the immediate horizon is automated analysis of explicit references. Currently Eunomos can find most explicit references using the XML Leges Linker, but an expert user needs to specify whether the reference is a simple reference or whether it modifies or overrides other legislation. We are looking to incorporate the human language technologies which enable the type of modification to be discerned automatically [33,10]. The tool for recognising types of modifications could also be used in a new module for automatically generating different versions of consolidated text, as done by Palmirani e Brighi [40]. Currently the system stores the original and most recent versions of legislation, and this is sufficient for the needs of prospective users. Nevertheless, the Eunomos system contains a functionality for adding any number of intermediate versions, so a consolidation module could be added in the future if required. Another high-priority work on the agenda is to improve the accuracy of the norm classifier by adding information to legislative articles sent to the classifier module. One problem we noticed is the difficulty of classifying articles where an amendment is made to a law without indicating its content: e.g., "article 101 of law 231/01 is abrogated". Adding the content of article 101 to the amending article could help improve the classification of the less informative article. For classifications assigned with less certainty, we can add as additional information the classification assigned to adjacent articles, where the classification assigned to those

articles have a higher probability. Another piece of additional information we could include is the descriptions of key concepts, as provided in the ontology.

Another area for future development is to exploit Eunomos's potential to cater for multilingual and multilevel legal research, since some clients may be interested in specialist databases for foreign legal systems. Some clients may find it useful to have a similar functionality to Lau [31]'s U.S. "50 state survey of the law" within Eunomos to help business undertake a survey of European, national and regional laws governing a particular topic area. While Eunomos uses the NormaInRete standard internally, as standards are developed for interchange between different legislative XML formats [14], it should be possible to use Eunomos in other jurisdictions. This would require suitable parsers to structure laws in XML in different languages. It is already possible, however, to model EU directives and their national implementations, and the Legal Taxonomy Syllabus ontology is already multilingual. The question then arises whether legislation from different jurisdictions can be compared for similarity or classification purposes. To extend the ontologies, we may investigate ways to extract terminology and map terms from various jurisdictions using similarity measures (as in Cheng et al. [18,17]).

There is good research on semantic technologies that are not being taken forward because of the bottleneck of building knowledge representation systems. The use of automated Information extraction techniques could significantly reduce this bottleneck. Future research on Eunomos will include populating fields such as deontic clause, passive role, active role, crime and sanction in the extended ontology for prescriptions using information extraction (IE) techniques. Information extraction research and evaluation has usually been performed on text taken from news articles or medical reports written in clearly identifiable sentences. Legislative text is an under-researched area in IE not least because legislative text is difficult to process.

For instance, semantic technologies could be used to map prescriptions to Business Process Management (BPM) activities (in-house banking processes). Banks manage thousands of BPM activities and a module that maps them to norms would be a valuable resource in ensuring that these banking processes are compliant.

In Boella et al. [11] we are also developing the conceptual model of roles in prescriptions using the model of Boella and van der Torre [13].

7. Related work

The proposal closest to Eunomos is the "Fill the gap" project by Palmirani et al. [41]. This project proposes a platform where legal documents are modelled using XML standards and the ontology layer is used as the interconnection technique between the pure text of the document and the embedded legal knowledge, including rules representing the norms expressed by the textual document. The ontology is used for modelling the legal concepts and to represent the properties and the T-Box axioms of the main legal values (e.g., copyright, work, etc.), including geo-spatial (e.g., jurisdiction) and legal temporal dimensions (e.g., enforceability, efficacy, applicability of the norms). The text, annotated in XML using Akoma Ntoso standard, and the metadata are connected manually to the ontology framework and finally, the rules, formalized in defeasible logic, are connected to the textual provisions and to general and abstract legal concepts modelled in the legal ontology. Eunomos does not cover rule modelling, since rules are considered too complicated for available knowledge engineers, and has a simpler treatment of the temporal dimension. Moreover, it does not foresee the construction of editors and visualization tools for rules. In contrast, Eunomos has been tailored carefully on the needs of users and on the capabilities of knowledge engineers, leading to a commercial product, resulting in a lightweight ontology acceptable by lawyers and introducing productivity tools like semi-automated classification and automated harvesting of laws.

The Eunomos system has also some similarities with that of Bianchi et al. [7] in that it is designed to help users view laws and classify terms. But the scope of the Eunomos project is wider, designed as it was to address real problems in accessing and managing information by legal researchers. While Bianchi et al. [7] take XML files as input, Eunomos can download text-based laws made available in official portals and convert them into XML, where XML files are not available. Eunomos has a number of useful features for viewing and updating information, and an automatic alert messaging system on legislative updates. The downside is that Eunomos requires considerable maintenance work, as web spiders need to keep up to date with any modifications made to online legal portals, and expert users are required to verify classification and find implicit references. The use of ontology in the two systems are also quite different. Bianchi et al. [7] use the Semantic Turkey [26] ontology, where

definitions can be taken from any source and arranged in any order. The Eunomos approach is more cautious, taking into account the strict demand for accuracy from the legal sector, encouraging the expert user to create links to definitions in legislation, judgement and official journals, and to track the evolution of terms in a systematic manner. Both Eunomos and Bianchi et al. [7] make use of statistical and reference data to help users find related norms though Bianchi combines these elements by factoring incoming and outgoing references into its statistical model.

Concerning text similarity, Bianchi et al. [7] used similarity techniques as well as incoming and outgoing references to find related paragraphs in different Italian legislation. They submitted the full text of the input paragraph as input query to the Terrier [38] open-source search engine in order to retrieve a list of related paragraphs. Four domain experts determined stated that 90% of the five top-ranked paragraphs were related, and 55% of the first 40 paragraphs were related. Lau [31] used Cosine Similarity and pattern rules for dates and measurements, references and neighbouring provisions, to identify related provisions in different legislation. Tagging was used for key phrase extraction. The vector model was used as the basis of different feature comparisons. The results showed that this mixture of features outperformed traditional bag-of-word model Latent Semantic Indexing where the average root mean square error were 22.9 and 27.4 respectively.

Concerning text classification for legal text, it is instructive to refer to de Maat et al. [21]'s comparison of machine learning versus knowledge engineering in classification of legal sentences, since Eunomos uses similar techniques. The conclusion of de Maat et al. [21]'s research (ibid, page 16) was that "a pattern based classifier is considered to be more robust in the categorization of legal documents at a sentence level". However, their classification task was quite different since that research was concerned with classifying the type of norms as delegations, penalizations, etc., while we categorize norms as belonging to different topic areas. The author (ibid. page 14) noted that SVMs were better than patterns at categorisation where word order was less restricted. Biagioli et al. [6] classified paragraphs from Italian law using Multiclass Support Vector Machines. However, they were also concerned with classification into types rather than topics, in their case high-level meta-classes such as 'Prohibition Action', 'Obligation Addressee', 'Substitution', and so on.

Concerning the idea of developing collaborative tools for building knowledge in the public sector, it is relevant to refer to Ghidini et al. [25]'s MoKi system, in which a wiki page, containing both unstructured and structured information, is associated with entities within the ontology and process model. The unstructured information is in natural language and may contain diagrams or pictures. The structured part has the same information encoded in the BPMN modelling language. The MoKi system has been developed for the public sector, but a version has also been developed for modelling business process management activities

There is a number of works that consider the theoretical issues related to the construction of legal ontologies [34,48,16]. In particular the framework presented in Kralingen [30] is a frame-based system that classifies the legal facts. A basic component of this system is the legal concept description, i.e., Kralingen proposes a distinction between a legal term and a legal concept similar to the distinction that we have adopted in the Legal Taxonomy Syllabus. From a practical point of view, there are two projects that are related in someway to the Legal Taxonomy Syllabus ontology of Eunomos. The "EURLex" system[14] is a web portal that interfaces a number of databases in order to access a wide collection of legal documents produced by the EU. However, in order to obtain a full coverage, EURLex limits the complete accessibility to legal documents, particularly for the needs of lawyers. Each query, even when using boolean search, reports too large instances without comprehensible classifications for the expectations of national jurists and practitioners, and thus hinders the applicability of EURLex for most legal uses in the Member States' legal. "Eurovoc"[15] is a web application that accesses a number a multilingual thesauri. The main point of this project is the splitting of the legal terms into two sets: the descriptor and non-descriptor. A non-descriptor legal term can be always be mapped into a descriptor legal term that has the same meaning. Moreover, the basic hypothesis is that each descriptor can be translated straightforwardly into the official languages of the EU. In contrast to the Legal Taxonomy Syllabus, the main purpose of Eurovoc is the information extraction. Indeed, the sparsness problems related to the bags of word techniques can be reduced by replacing the non-descriptor with the corresponding descriptor. However

[14] http://europa.eu.int/eur-lex/

[15] http://europa.eu.int/celex/eurovoc/

Eurovoc does not distinct between a legal terms and a legal concepts, and cannot resolve easily the problems related to the polysemy.

Related work on legal ontologies include also Peters et al.'s [42] LOIS database of legal terms, which adopted the structure of WordNet [22] and EuroWordNet [43]. It can be particularly suitable for information retrival for which the LOIS database was developed, as the collapse of terms into synsets aids the recall if not always the precision of document retrieval. Whilst the final goal of LOIS is to support applications concerning information extraction, the Legal Taxonomy Syllabus ontology of Eunomos is concerned with the access of human experts to the EU documents.

Agnoloni et al. [1]'s FrameNet ontology departs from the WordNet structure, emphasising that meaning depends on "under which *Circumstances*, which *State of affairs* is sanctioned under which *Principle*". Like the Legal Taxonomy Syllabus ontology, Agnoloni et al. [1] separate concepts from terms. However, unlike Legal Taxonomy Syllabus they assume that translated terms are exact and that equivalent multilingual terms map onto the same concept.

8. Conclusions

In this paper we have illustrated the Eunomos software, a legal document and knowledge management system to help law researchers and practitioners manage complex information, which incorporates state-of-the art research from legal informatics.

The Eunomos system addresses the retrieval and interpretation problems mentioned in Section 1.4 with the following functionalities:

- the problem of increase in scope, volume and complexity of the law is addressed by creating a large database of laws converted into legislative XML and downloaded automatically from legislative portals, which are annotated and updated regularly;
- the problem of specialisation is addressed by the semi-automated classification of articles, enabling users to view only those sections of legislation that are relevant to their domain of interest;
- the problem of fragmentation of laws is handled by enabling users to view legislation at European, national and regional level from the same web interface;
- the problem of keeping up with changes in the law is addressed by alert messages sent to users notifying them that a newly downloaded legislation is relevant to their domain of interest. Where legislation is updated, users can view consolidated text where available from state portals, as well as the original version. Where previous laws are modified or abrogated implicitly, Eunomos provides a mechanism to annotate the legislation with implicit cross-references (and hyperlinks) to the amending piece of legislation.
- the issue of legal "terms of art" that can vary in meaning in different contexts and over time is addressed with multi-level updatable domain-specific ontologies where terms can be aligned with various concepts and definitions; concepts are associated with the specific textual sources by links.
- the issue of vague and imprecise language is addressed with additional information, clarifications and interpretation supplied by knowledge engineers based on thorough legal research;
- the issue of cross-references is addressed by a facility whereby the user can either hover over a cross-reference, and the referenced article appears in a pop-up text box, or click on a hyperlink to the referenced article to see the text in context.

The system has been developed with clearly-defined aims and objectives to support the work of law firms, law scholars, and in-house legal offices in financial institutions and public sector organisations.

Eunomos is being developed as a commercial software part of a wider suite distributed by Nomotika s.r.l., a spinoff of Università di Torino.

References

[1] T. Agnoloni, M. Fernandez Barrera, M.T. Sagri, D. Tiscornia, and G. Venturi. When a framenet-style knowledge description meets an ontological characterization of fundamental legal concepts. In *AICOL 2009*, pages 93–112, 2009.

[2] G. Ajani, G. Boella, L. Lesmo, M. Martin, A. Mazzei, D. Radicioni, and P. Rossi. Multilevel legal ontologies. In E. Francesconi, S. Montemagni, W. Peters, and D. Tiscornia, editors, *Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language*, volume 6036 of *Lecture Notes in Computer Science*, pages 136–154. Springer, 2010.

[3] G. Ajani, G. Boella, L. Lesmo, G. Mazzei, and P. Rossi. Multilingual Ontological Analysis of European Directives. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 21–24. ACL, 2007.

[4] G. Ajani and M. Ebers, editors. *Uniform Terminology for European Contract Law*. Nomos, Baden Baden, 2005.

[5] G. Ajani, L. Lesmo, G. Boella, A. Mazzei, and P. Rossi. Terminological and ontological analysis of european directives: multilinguism in law. In *The 11th International Conference on Artificial Intelligence and Law, Proceedings of the Conference (ICAIL)*, pages 43–48. ACM, 2007.

[6] C. Biagioli, R. Francesconi, A. Passerini, S. Montemagni, and C. Soria. Automatic semantics extraction in law documents. In *The Tenth International Conference on Artificial Intelligence and Law, Proceedings of the Conference*, pages 133–140. ACM, 2005.

[7] M. Bianchi, M. Draoli, G. Gambosi, M.T. Pazienza, N. Scarpato, and A. Stellato. ICT tools for the discovery of semantic relations in legal documents. In *Proceedings of the 2nd International Conference on ICT Solutions for Justice (ICT4Justice)*, 2009.

[8] G. Boella, L. di Caro, and L. Humphreys. Using classification to support legal knowledge engineers in the eunomos legal document management system. In *Fifth International Workshop on Juris- informatics(JURISIN)*, 2011.

[9] G. Boella, L. di Caro, L. Humphreys, and L. Robaldo. Using legal ontology to improve classification in the eunomos legal document and knowledge management system. In *Semantic Processing of Legal Texts (SPLeT) at LREC12*, 2012.

[10] G. Boella, L. di Caro, L. Humphreys, and L. van der Torre. NLP challenges for eunomos a tool to build and manage legal knowledge. In *Proceedings of the eighth international conference on Language Resources and Evaluation (LREC)*, 2012.

[11] G. Boella, L. Humphreys, and L. van der Torre. The role of roles in eunomos, a legal document and knowledge management system for regulatory compliance. In *Proceedings of Information Systems: a crossroads for Organization, Management, Accounting and Engineering (ITAIS) Conference*, 2012.

[12] G. Boella, M. Martin, P. Rossi, L. van der Torre, and A. Violato. Eunomos, a legal document and knowledge management system for regulatory compliance. In *Proceedings of Information Systems: a crossroads for Organization, Management, Accounting and Engineering (ITAIS) Conference*, Berlin, 2012. Springer.

[13] Guido Boella and Leon van der Torre. The ontological properties of social roles in multi-agent systems: definitional dependence, powers and roles playing roles. *Artif. Intell. Law*, 15(3):201–221, 2007.

[14] A. Boer and R. Winkels. What's in an interchange standard for legislative XML? *I Quaderni*, 18:32–41, 2005.

[15] T. D. Breaux:. *Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems*. PhD thesis, North Carolina State University, Raleigh (NC), 2009.

[16] J. Breuker, A. Valente, and R. Winkels. Legal ontologies: A functional view. In *Procs. of 1st LegOut Workshop on Legal Ontologies*, pages 23–36, 1997.

[17] Chin Pang Cheng, Gloria T. Lau, Kincho H. Law, Jiayi Pan, and Albert Jones. Regulation retrieval using industry specific taxonomies. *Artif. Intell. Law*, 16:277–303, September 2008.

[18] Chin Pang Cheng, Jiayi Pan, Gloria T. Lau, Kincho H. Law, and Albert Jones. Relating taxonomies with regulations. In *Proceedings of the 2008 international conference on Digital government research*, pages 34–43, 2008.

[19] J.G. Conrad and F. Schilder. Opinion mining in legal blogs. In *Proceedings of the 11th international conference on Artificial intelligence and law*, ICAIL '07, pages 231–236, New York, NY, USA, 2007. ACM.

[20] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[21] E. de Maat, K. Krabben, and R. Winkels. Machine learning versus knowledge based classification of legal texts. In *Proceeding of the 2010 conference on Legal Knowledge and Information Systems: JURIX 2010: The Twenty-Third Annual Conference*, pages 87–96. IOS Press, 2010.

[22] C. Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.

[23] M. Fernandez-Barrera and P. Casanovas. Towards the intelligent processing of non-expert generated content: Mapping web 2.0 data with ontologies in the domain of consumer mediation. In *Proceedings of International Conference on Artificial Intelligence and Law Workshop, Applying Human Language Technology to the Law*, pages 18–27, 2011.

[24] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening ontologies with dolce. In *Proc. EKAW 2002*, Siguenza (SP), 2002.

[25] Chiara Ghidini, Marco Rospocher, and Luciano Serafini. Moki: a wiki-based conceptual modeling tool. In *Proceedings of the EKAW2010 Poster and Demo Track*, volume 674 of *CEUR Workshop Proceedings (CEUR-WS.org)*, Lisbon, Portugal, 11th October-15th October 2010 2010.

[26] D. Griesi, M. T. Pazienza, and A. Stellato. Semantic turkey - a semantic bookmarking tool (system description). In *4th European Semantic Web Conference (ESWC 2007)*, volume 4519 of *Lecture Notes in Computer Science*, pages 779–788. Springer, 2007.

[27] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November 2009.

[28] N. Holmes. Accessible Law. http://blog.law.cornell.edu/voxpop/2011/02/15/accessible-law/, 2011.

[29] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. *Machine Learning: ECML-98*, pages 137–142, 1998.

[30] V. Kralingen. A conceptual frame–based ontology for the law. In *Procs. of 1st LegOut Workshop on Legal Ontologies*, pages 15–22, 1997.

[31] Gloria Lau. *A Comparative Analysis Framework for Semi-Structured Documents, with Applications to Government Regulations*. PhD thesis, University of Stanford, 2004.

[32] L. Lesmo. The Turin University Parser at Evalita 2009. *Proceedings of EVALITA*, 9, 2009.

[33] A. Mazzei, D. Radicioni, and R. Brighi. NLP-based extraction of modificatory provisions semantics. In *The 12th International Conference on Artificial Intelligence and Law, Proceedings of the Conference (ICAIL)*, pages 50–57. ACM, 2009.

[34] L.T. McCarty. A language for legal discourse: Basic features. In *Proc. of Second International Conference on Artificial Intelligence and Law*, 1989.

[35] M.Cherubini and D.Tiscornia. An Ontology-based model of procedural norms and regulated procedures. Technical Report 1/2010, ITTIG-CNR, 2010.

[36] G.A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[37] Commission of the European Communities. Internet of Things, An action plan for Europe.

http://ec.europa.eu/information_society/policy/rfid/documents/commiot2009.pdf, 2009.

[38] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and C. Lioma. Terrier: A High Performance and Scalable Information Retrieval Platform. In *Proceedings of ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR 2006)*, 2006.

[39] M. Palmirani. Legislative change management with akomantoso. In G. Sartor, M. Palmirani, E. Francesconi, and M.A. Biasiotti, editors, *Legislative XML for the Semantic Web*, volume 4 of *Law, Governance and Technology Series*, pages 101–130. Springer, 2011.

[40] M. Palmirani and R. Brighi. Norma-system: A legal document system for managing consolidated acts. In *Database and Expert Systems Applications, 13th International Conference, DEXA 2002, Proceedings*, volume 2453 of *LNCS*, pages 310–320. Springer, 2002.

[41] M. Palmirani, T. Ognibene, and L. Cervone. Legal rules, text, and ontologies over time. In *Proceedings of RuleML 2012*, 2012.

[42] W. Peters, M.T. Sagri, and D. Tiscornia. The structuring of legal knowledge in lois. *Artif. Intell. Law*, 15(2):117–135, 2007.

[43] W. Peters, P. Vossen, P. Díez-Orzas, and G. Andriaens. Cross-linguistic alignment of wordnets with an inter-lingual-index. *Computers and the Humanities*, 32(2-3):221–251, 1998.

[44] P. Rossi and C. Vogel. Terms and concepts; towards a syllabus for european private law. *European Review of Private Law (ERPL)*, 12(2):293–300, 2004.

[45] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[46] G. Sartor. Access to legislation in the semantic web. In M.A. Biasiotti and S. Faro, editors, *From Information to Knowledge - Online Access to Legal Information: Methodologies, Trends and Perspectives*. IOS, 2011.

[47] J.R. Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.

[48] R.K. Stamper. The role of semantics in legal expert systems and legal reasoning. *Ratio Juris*, 4(2):219–244, 1991.

[49] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelflee. Vision and Challenges for Realising the Internet of Things. http://www.internet-of-things-research.eu/pdf/IoT_Clusterbook_March_2010.pdf, 2010.

[50] C. Tullo. Access to legislation in the semantic web. In M.A. Biasiotti and S. Faro, editors, *Online Access to UK Legislation: Strategy and Structure*. IOS, 2011.

[51] 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*, 6:27–57, 1998.