Artif Intell Law (2016) 24:245–283
DOI 10.1007/s10506-016-9184-3

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

Guido Boella[1,4] · Luigi Di Caro[1,4] · Llio Humphreys[2] · Livio Robaldo[2,4] · Piercarlo Rossi[3,4] · Leendert van der Torre[2]

Published online: 28 June 2016

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. Using NLP tools to semi-automate the lower-skill tasks makes this ambitious

✉ Livio Robaldo
livio.robaldo@uni.lu;
http://www.nomotika.it

Guido Boella
boella@di.unito.it;
http://www.nomotika.it

Luigi Di Caro
dicaro@di.unito.it;
http://www.nomotika.it

Llio Humphreys
llio.humphreys@uni.lu

Piercarlo Rossi
piercarlo.rossi@unipmn.it;
http://www.nomotika.it

Leendert van der Torre
leon.vandertorre@uni.lu

[1] Department of Computer Science, University of Turin, Turin, Italy

[2] Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[3] Università del Piemonte Orientale, Vercelli, Italy

[4] Nomotika s.r.l., Turin, Italy

project a realistic commercial prospect as it helps keep costs down while at the same time allowing greater coverage. We describe the core system from workflow and technical perspectives, and discuss applications of the system for various user groups.

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 recognizes 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 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.

The Eunomos software described in this paper was originally developed to support regulatory compliance in the context of the ICT4Law[1] project, further extended in the subsequent years in the context of the projects ITxLaw[2] and EUCases,[3] and still nowadays in the context of the ongoing projects ProLeMAS,[4] BO-ECLI,[5] and MIREL.[6]

Eunomos is the basis of the Menslegis commercial service for compliance distributed by Nomotika s.r.l., a spinoff of University of Torino, Italy, in which four of the authors are partners.

Currently, Eunomos deals with Italian legislation only; nevertheless, in the context of the ongoing EU projects, its applicability is going to be extended to the whole EU legislation.

[1] http://www.ict4law.org.

[2] http://www.itxlaw.eu.

[3] http://www.eucases.eu.

[4] http://www.liviorobaldo.com/ProLeMAS.htm.

[5] http://www.bo-ecli.eu.

[6] http://www.mirelproject.eu.

In what follows, we present 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.

1.2.1 Problem 1: Laws are not clearly classified

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.

1.2.2 Problem 2: Multiple jurisdictions

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." (Lau 2004). In Europe, due to subsidiarity, laws are applicable from European, national, regional and municipal levels.

1.2.3 Problem 3: Volume of law

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 6000 pages of internal regulations.[7]

1.2.4 Problem 4: Accessibility

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 (2011) 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

[7] Source: ABILab.

issues at stake and the different roles played by its users (legislators, judges, administrators, parties in economic or other transactions)" (Sartor 2011, p. 7). 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 2011) 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."

1.2.5 Problem 5: Updates and consolidated text

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:[8] "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 continuous, 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

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

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 2004).

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.

1.3.1 Problem 1: "Terms of art": different to ordinary meaning

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.

1.3.2 Problem 2: "Terms of art": can vary in different contexts and jurisdictions

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.

1.3.3 Problem 3: Intentional vagueness

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 (2009) 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".

1.3.4 Problem 4: General problems of language

Some problems of legal language derive from the imprecise nature of language. The Supreme Court[9] advises that in cases of attributive ambiguity, legislative intent may override literal interpretation: "Ordinarily, as in everyday English, use of the

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

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.

1.3.5 Problem 5: Cross-references

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;
- Legislative text can be 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[10] and even mainstream[11] 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 (Ajani et al. 2007). 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 Sect. 3 we describe the main functionalities of the software and the workflow of users and knowledge engineers. In Sect. 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. 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.

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

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

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 allows greater interoperability between government 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 Sect. 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 (Palmirani 2011) 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.

The ONTOMEDIA project (Fernandez-Barrera and Casanovas 2011) 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.) (Cherubini and Tiscornia 2010) 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 (Ajani et al. 2007). 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 (1998) 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 (Ajani and Ebers 2005; Rossi and Vogel 2004) project. The UT (Uniform Terminology For European Private Law) project is a Research Training Network (RTN) funded by European Commission.

The research network involves researchers from seven 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 (Gangemi et al. 2002).

Another feature of the Legal Taxonomy Syllabus developed in Ajani et al. (2007) 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 (Fellbaum 1998) and FrameNet (Fillmore and Collin 2000), 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 was developed in the context of the ICT4Law[12] project, further extended in the subsequent years in the context of the projects ITxLaw[13] and EUCases,[14] and still nowadays in the context of the ongoing projects ProLeMAS,[15] BO-ECLI,[16] and MIREL.[17]

It 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 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, Italy, in which four of the authors are partners.

The system, being based on the Legal Taxonomy Syllabus ontology, 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 version) converted into legislative XML format, with links between related

[12] http://www.ict4law.org.

[13] http://www.itxlaw.eu.

[14] http://www.eucases.eu.

[15] http://www.liviorobaldo.com/ProLeMAS.htm.

[16] http://www.bo-ecli.eu.

[17] http://www.mirelproject.eu.

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,[18] the portal Arianna of Regione Piemonte[19] 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 including 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 Sects. 4.4 and 4.5. Legal concepts can be extracted and modelled using the legal ontology tool called Legal Taxonomy Syllabus, the specialist multilevel and multilegal ontology (Ajani et al. 2007) for terminology management of European Directives and their national implementations described in Sect. 2.2. The ontology is part of the database and it 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 RDF subject-predicate-object triple[20] 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 Sect. 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 Sect. 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

[18] http://www.normattiva.it.

[19] http://arianna.consiglioregionale.piemonte.it.

[20] http://www.w3.org/RDF.

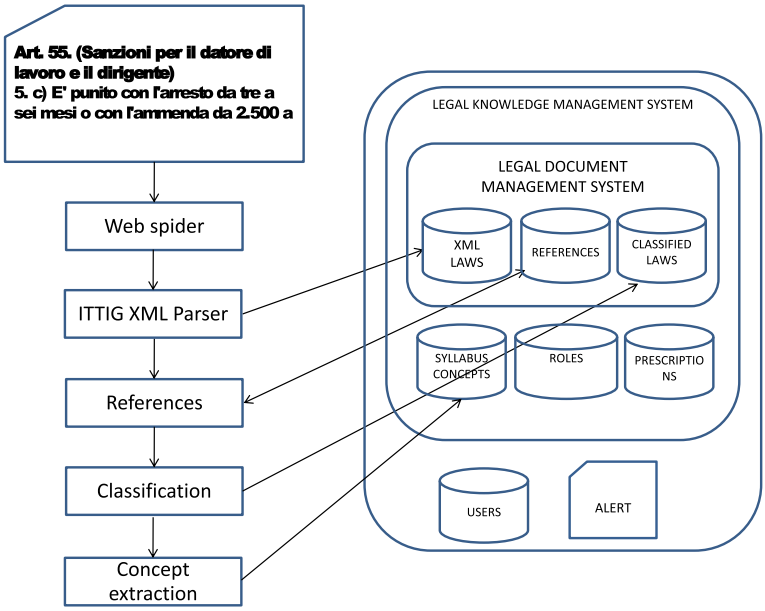

Fig. 1 Key components of the Eunomos system

Boella et al. (2012c). This component is outside the scope of the paper (see Sect. 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 Sect. 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 Fig. 2). All versions of a law stored are retrieved, unless dates of validity 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:

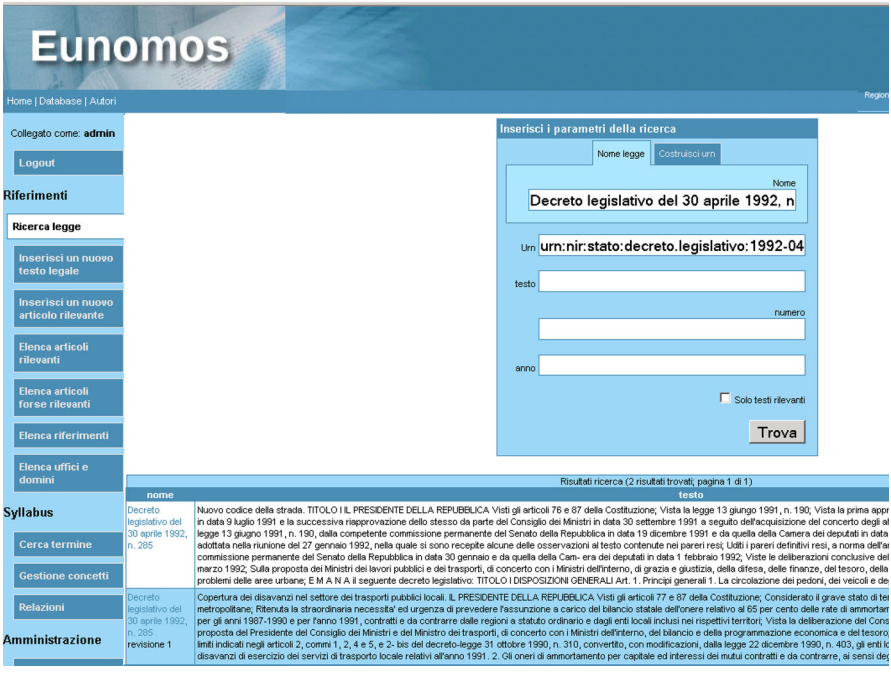

Fig. 2 The search interface of the Eunomos system

- 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.[21]
  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 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 Sect. 4.4).
- 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. 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 the Legal Taxonomy Syllabus ontology from the same web interface.
- Registered users can add their comments on single articles or items of legislations.

[21] 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.

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. (2007) 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 Sect. 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 Eunomos system create links to the XML versions of the legislation via URN identifiers.

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 ontology 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.

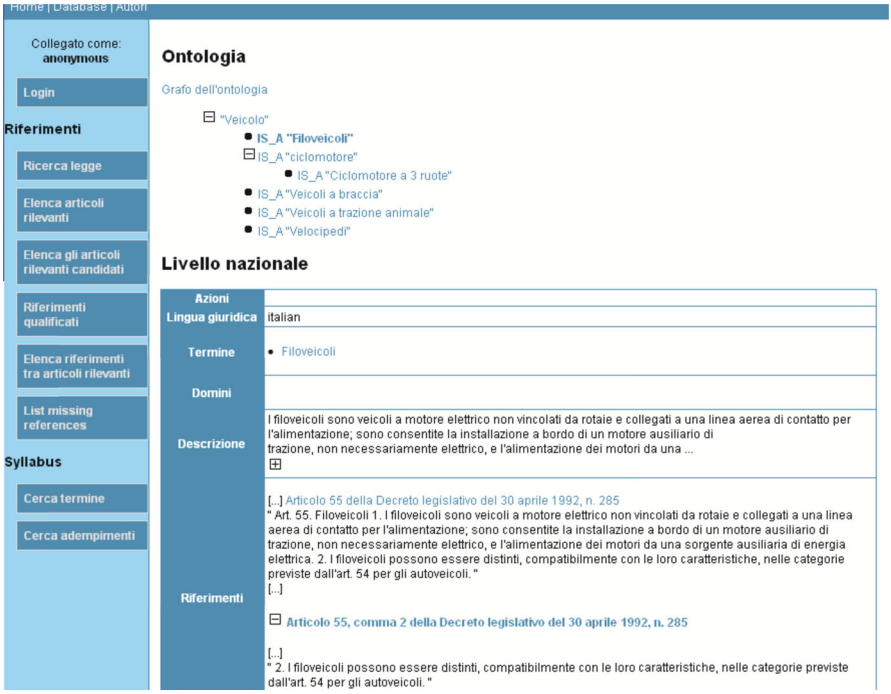

Fig. 3 Legal taxonomy syllabus ontology within the Eunomos system

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 Sects. 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:

- Inserting missing legislations in the database;
- Checking the output of the legislative XML parser and correcting any errors arising out of irregular patterns in the text;
- Adding cross-references between legal documents or validating the ones suggested by the automatic reference detection tool;
- Classifying the type of modificatory references or validating the ones suggested by the automatic reference classification tool;
- Classifying the domain of legislative norms selecting among the suggestions proposed by the automated classifiers;

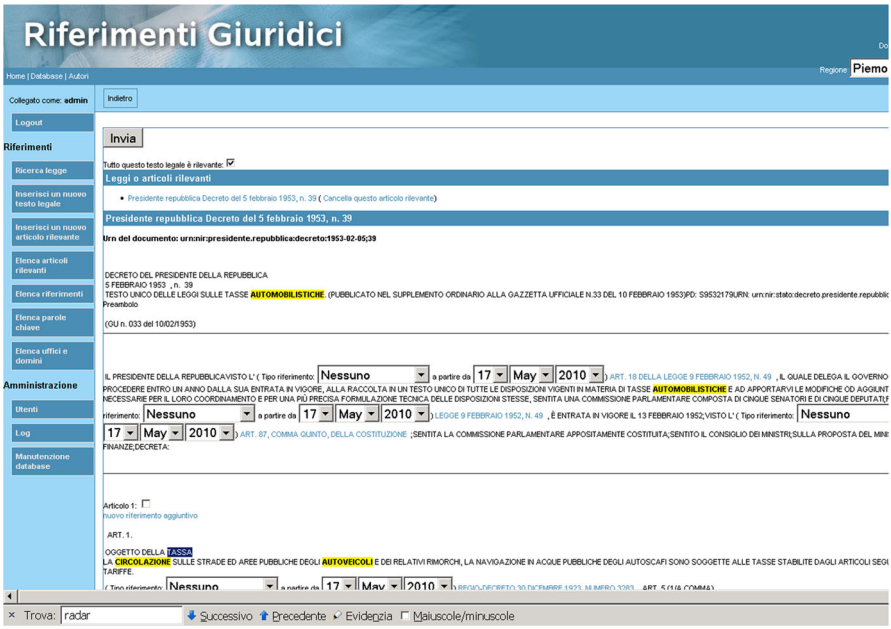

Fig. 4 Annotating legislation

- Adding concepts and terms to the ontology, and link them to the legislative text;
- 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 Sect. 4).

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. As they work through the text, knowledge engineers also annotate cross-references and add terms to the ontology as well as links between the concepts in the ontology and the text. The Eunomos system contains an automatic reference detection tool that automatically finds and classifies references to articles in other legislation and creates inline hyperlinks within the legislation text. Knowledge engineers then look at each explicit reference and possibly correct its domain and its type: whether it is merely a simple reference or in fact modifies or overrides existing legislation (see Fig. 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 Fig. 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, the system suggests him a type (modification, suspension, etc.) for each reference to other legislation, which he can possibly modify. 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.

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 ontology 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. Once the concept is created, automatic rule-based pattern-matching procedures look for occurrences of the new concept in the documents. This process is illustrated in Fig. 5.

Knowledge engineers are also active when new legislations are issued. When the system has a number of already 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 Sect. 4.5).

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 an

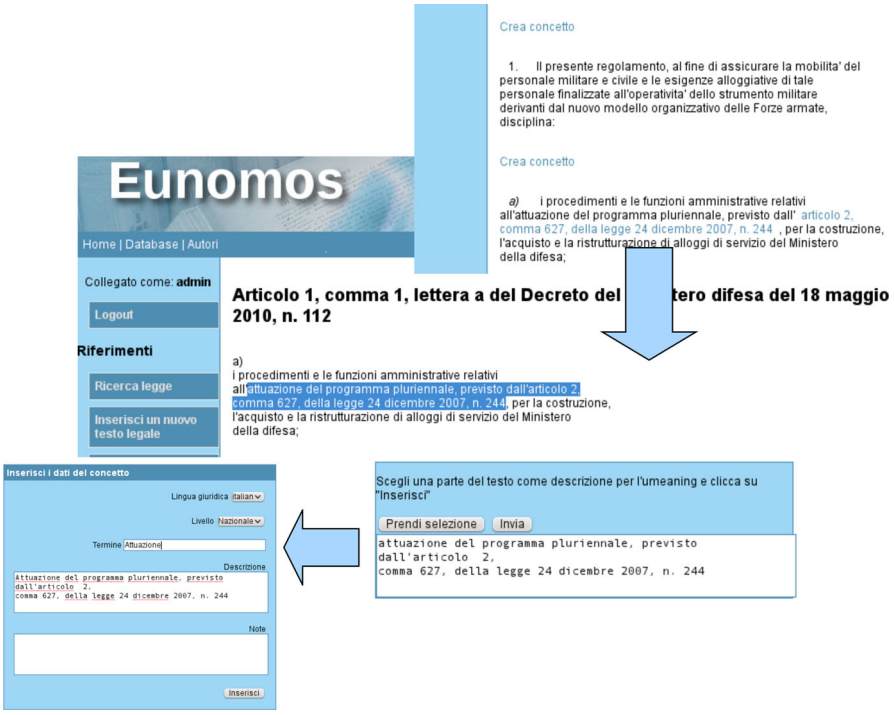

Fig. 5 Creating concepts

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* (Keywords) can also be useful for identifying relevant domains.

The Eunomos ontologies are populated and updated semi-automatically: 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.

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 Eunomos legal document and knowledge management 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, including XML files and ontologies as well as a cache table that stores the complete transitive closure of the ontological (transitive) relations in order to enhance the performance of the queries, 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 and the legal text repository.

We chose to store the RDF subject-predicate-object triples defining the ontology as well as the connections between the concepts and the terms used to express them, into a relational database, rather than a NoSQL database, in order to "take advantage of 35+ years of research on efficient storage and querying, industrial-strength transaction support, locking, security, etc." (Bornea et al. 2013). On the other hand, NoSQL DBMSs are characterized by a schema-less data model, which facilitates operations on RDF data models via object oriented programming. However, it is still a new technology in a constant improvement. We address the interested reader to Neumann and Weikum (2010), Nayak et al. (2013) and Aluç

et al. (2014) among others, as well as to specialized forums on the Web, where expert programmers and database administrators discuss advantages and disadvantages of the two kinds of DBMSs.

As argued above in Sects. 1.2 and 1.3, the Eunomos system indexes and classifies Italian legislation with respect to the non-ambiguous definitions of the Legal Taxonomy Syllabus ontology. In other words, it grounds concepts of the Legal Taxonomy Syllabus ontology to their legislative sources, structured into a uniform XML format, in order to facilitate searches and the updating of consolidated text. The update of the knowledge bases is *semi-automatically* carried out by legal experts.

More advanced reasoning tasks, able to *automatically* infer new knowledge from the existing one, and which will possibly require a massive use of object programming software, are the object of our future works (cf. Sect. 6 below). In our future solutions, we will possibly consider NoSQL implementations to enhance the overall efficiency.

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. Specifically, the Web application to the system is divided into three parts:

- The pure presentation, using the Smarty[22] template engine;
- 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[23] if they are in pure textual format.[24] 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

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

[23] www.xmleges.org.

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

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"/>
```

The segment of Fig. 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.

Other rule-based procedures based on the TULE parser (Lesmo 2009) have been implemented in order to find and create links to the ontology whenever new concepts are added to it (*Crea concetto* page). A pattern-matching rule is automatically generated from the description of the concept, e.g., "direttore di banca" (bank director). Then, it is executed on the legal documents in order to find other occurrences of the new concept. TULE allows for a certain degree of flexibility against morphological inflections; with respect to the last example, for instance, the pattern-matching rule is also able to link "direttrice di banca" to the new concept, where "direttrice" is the feminine inflection of "direttore".

Both the XML Leges Linker tool and the additional rule-based pattern-matching procedure that links concepts to their occurrences in text report very good performance, in that the linguistic variation of the text they recognize is rather low.

4.3 Rule-based classification of modificatory provisions

Eunomos uses a rule-based pattern-matching module to automatically determine whether a reference is a simple reference or it modifies or overrides other legislation. In case of errors, the interface of Eunomos enables knowledge engineers to manually override the result of the pattern-matching procedure.

```
<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>
```

Fig. 6 An example of NIR XML annotation

Contrary to the identification of references and ontological concepts, classifying modificatory provisions features an higher linguistic variation, and rules must deal with ambiguities. For instance, the verb "sopprimere" (to suppress) may be used in legislation to specify either an "abrogazione" (abrogation) or a "sostituzione" (substitution). When the verb is followed by the preposition "da" (by), it usually specifies a substitution, e.g. "Articolo X è soppresso da Articolo Y' ("Article X is suppressed by Article Y"). Otherwise, it usually specifies an abrogation.

To deal with this ambiguity, the rule-based module includes two rules: a default rule that classifies the modificatory provision as abrogation and an higher-priority rule that checks whether the verb is used in a linguistic pattern that denotes a substitution. For ease of understanding, we provide only conceptual representations

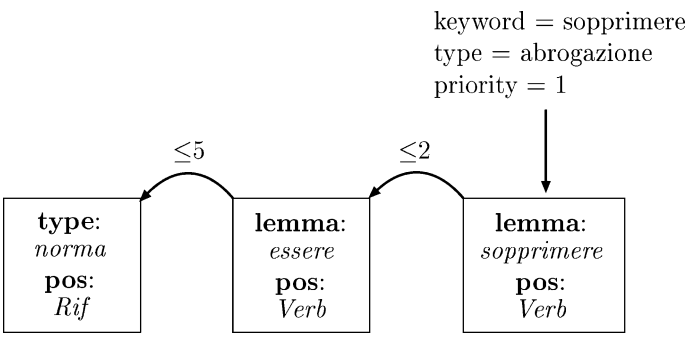

Fig. 7 A rule for some kinds of 'abrogazioni' (*abrogations*)

of the rules in the figures below. Figure 7 shows the conceptual representation of the default rule that classifies the modificatory provision as abrogation. The rule is triggered when the system finds in the input text a verb with the lemma 'sopprimere'.

Then, it checks whether there is a verb with lemma 'essere' (to be) in the two[25] preceding words, and whether there is a normative reference in the five preceding words of the verb with lemma 'essere'. The normative references, found by the automatic reference detection tool, are substituted with the strings `rif1`, `rif2`, etc. and considered as proper nouns by the TULE parser.

When the rule in Fig. 7 is satisfied, the provision is annotated as 'abrogazione', with the normative reference occurring therein identified as 'norma'.

On the other hand, we add in the system the rule in Fig. 8 and assign to it a higher priority than the rule in Fig. 7, so that it is executed before the latter.

In Fig. 8, the checks carried out on the words preceding the keyword 'sopprimere' are the same as for those in Fig. 7. Furthermore, the rule in Fig. 8 requires the occurrence of the preposition 'da' immediately after the keyword and a normative reference (that will be annotated as 'novella') among the five words following the preposition.

To evaluate the Eunomos module for extracting legal modifications, we used a dataset composed of 180 files, containing 2306 modificatory provisions manually annotated by the legal experts of the CIRSFID research center[26] of the University of Bologna.

Our system obtains 86.60 % recall and 98.56 % precision. The match between a provision automatically calculated by the module and the corresponding one stored in the corpus is considered valid only if it matches both the type of the provision (abrogation, substitution, insertion, etc.) and *all* its arguments, such as "norma" and "novella" in Fig. 8. A similar system has been proposed in Lesmo et al. (2009). That system also uses the TULE parser and it has been evaluated on the same corpus of 2306 modificatory provisions from CIRSFID. Lesmo et al. (2009) reports 71.7 % recall and 83.0 % precision.

It is worth noticing that the system presented here achieves an very high level of precision, close to 100 %, because the rules behave as a kind of "filter". In other words, the system uses *ad-hoc* rules, each of which describes a specific valid pattern. As a consequence, (almost) any provision matching with this pattern is precisely classified by the pattern itself. Recall is lower in that rules are added one by one, which turns out to be an highly time-consuming task.

4.4 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

[25] We specified a maximum distance of 2 words in order to encompass both sentences of the form 'Il rif1 è soppresso' (The rif1 is suppressed) and sentences of the form 'Il rif1 è stato soppresso' (The rif1 has been suppressed). In Italian, the lemma of both words 'è' and 'stato' is 'essere'.

[26] http://www.cirsfid.unibo.it.

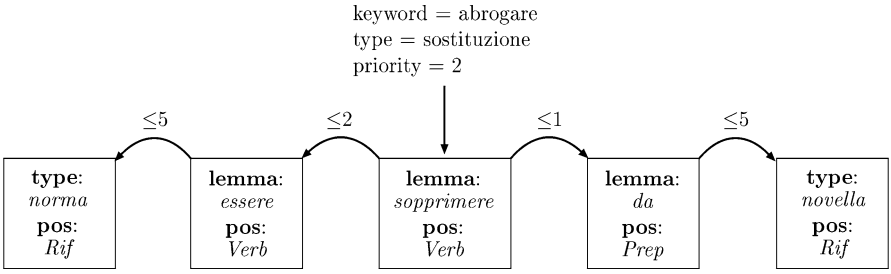

Fig. 8 A rule for certain kind of 'sostituzioni' (*substitutions*)

contains a large amount 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 contains 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.

For each piece of text, Eunomos may generate a list of the most similar texts in the whole database using Cosine Similarity. Where labelled data is not available, Cosine Similarity can be also used to build a training set for a supervised classification module.

Applying Cosine Similarity to search for relevant text is a common practice in general-purpose Information Retrieval tasks (Salton and Buckley 1988). In these cases, the only issue is to determine how many texts to select and return. This means choosing an appropriate threshold (or cutoff) to apply to the ordered list of relevant

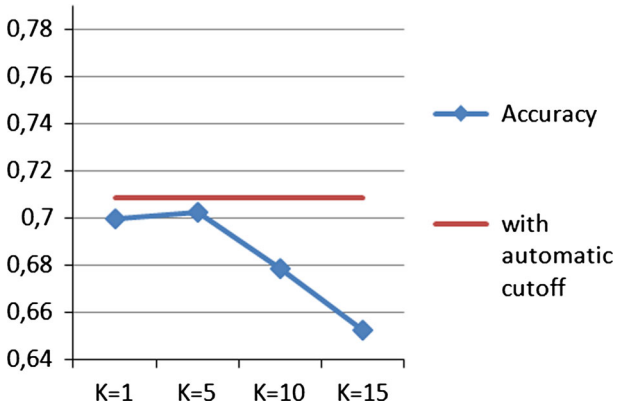

Fig. 9 Evaluation of the accuracy of the cosine similarity-based approach for finding relevant articles, using the class labels associated to the articles. Note that the accuracy levels reached by the automatic technique is higher than with the use of fixed cutoffs

articles created with the Cosine Similarity measure. A naive solution for truncating the list of texts that are ordered by its similarity with the input one is to use a fixed cutoff k. This way, only the first k articles are considered as relevant. However, this approach does not take into account the distribution of the ordered similarity values. An alternative approach is to find where the similarity values suffer a significant fall. This separates the actual similar texts from the rest. A practical way to implement this idea is to analyze the distribution of the ordered values looking at the highest difference (or highest "jump") between adjacent values in the list (Boella et al. 2011)

In our experiments, we made use of the categories associated with already-labeled documents (see Sect. 4.5) as part of the similarity evaluation process. More in detail, given one document d and a set of similar ones Sd, the evaluation task looks at whether the documents contained in Sd have the same categories as the input document d. Figure 9 shows the result of the accuracy when fixing the cutoff k, and when using our document-level automatic estimation of k. This shows that, notwithstanding the benefit of using a variable and data-dependent approach for estimating the cutoff, the accuracy level reached by this technique is noticeably higher than with the use of fixed cutoffs.

4.5 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. (2011, 2012a).

Table 1 The data used for training the classifier

Class	Description	Cardinality	S_1	S_2	S_3
C_1	Risks evaluation	11	x	x	x
C_2	Contracts	6	x	x	
C_3	Management of emergencies	9	x	x	x
C_4	Controls	5	x	x	
C_5	Information	7	x	x	
C_6	Formation and updating	7	x	x	
C_7	Public health surveillance	4	x		
C_8	Periodic meetings	4	x		
C_9	Communications	5	x	x	
C_{10}	Proscriptions	1	x		
C_{11}	Work environments	38	x	x	x
C_{12}	Work equipments and devices for personal protection	43	x	x	x
C_{13}	Signals for security and health on work	1	x		
C_{14}	Devices with video display terminals	9	x	x	x
C_{15}	General obligations	6			
–	–	156	150	140	110

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.

We use the TULE parser (Lesmo 2009) 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 (Fellbaum 1998) 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 the proposed system, we used a set of documents that have been manually annotated with categories, allowing the training of a supervised classification module. Such training set is composed by 156 legal texts and 15 categories (or classes). Since statistical methods requires sufficiently large textual information for each category, we filtered out those with few associated documents, building three datasets with different degrees of filtering (namely, S_1, S_2, and S_3), see Table 1. Note that dataset S_3 preserves more than 70 % of the original data (i.e. 110 documents out of 156), although it contains only 5 out of 15 total categories. One category (C_{15} in Table 1) has not been used since the associated texts do not contain any specific topic.

Although there are plenty of algorithms for text classification, we used the well-known Support Vector Machines (SVM) for this task, since it frequently achieves state-of-the-art accuracy levels (Joachims 1998; Cortes and Vapnik 1995). This algorithm makes use of vectorial representations of text documents and works by calculating the hyperplane having the maximum distance with respect to the nearest data examples. More in detail, we used the Sequential Minimal Optimization algorithm (SMO) (Platt 1999) with a polynomial kernel. The association between text and a category label has been fed into an external application based on the WEKA toolkit (Hall et al. 2009) and incorporated in Eunomos, creating a model that can be used to classify new laws inserted on a daily basis into the database by

Table 2 Separability Index (SI) and accuracy values computed on the three datasets S_1, S_2, and S_3, using the tenfold cross validation scheme

Dataset	Separability Index (SI) (%)	Accuracy (%)
S_1	66.66	71.33
S_1 + TULE	72.31	78.00
S_2	69.28	74.28
S_2 + TULE	74.07	82.96
S_3	83.63	89.09
S_3 + TULE	91.09	92.72

The accuracy is calculated as the percentage of correctly classified documents on the total

web spiders or users. The WEKA toolkit was used as a framework for the experiments because it supports several algorithms and validation schemes allowing an efficient and centralized way to conduct experiments and evaluate the results of the system.

In addition to the standard bag-of-words approach (where each text is represented as an unordered collection of words), we also wanted to test whether the system TULE and its additional features increased the accuracy of the classification module with respect to the standard use of WordNet lemmas. This is marked with the label "+TULE" in the results of Table 2. As it can be noticed, the use of the additional features improved the accuracy of the classifier.

The evaluation of a classification task can range from very poor to excellent depending on the data. A simple way to estimate the complexity of the input is to compute the separation and compactness of the classes. The Separability Index (SI) (Greene 2001) measures the average number of documents in a dataset that have a nearest neighbor within the same class, where the nearest neighbor is calculated using Cosine Similarity. Tests on the whole dataset revealed a SI of 66.66 %, which indicates a high overlap among the labelled classes. Table 2 shows the SI values for all the three datasets. The SVM classifier achieves an accuracy of 92.72 % when trained with the n-folds cross validation scheme (Greene 2001) on dataset S3 + TULE (using $n = 10$, which is a common practice in the literature).

As shown in Table 2, the classifier achieves lower accuracy levels with datasets S_1 and S_2, though it was already expected from their low SI values. Nevertheless, it is interesting to see that classification on dataset S_1 is still acceptable in terms of accuracy despite its very low SI. This is due to the fact that, although there is a large overlap between the dictionaries used in different classes, there are some terms that characterize them properly.

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, distributed by the spin-off Nomotika s.r.l..

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.

In the rest of the section, we briefly outline some of the use cases where Eunomos is or could be employed.

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. 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 (Boella et al. 2012b, c).

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. On the other hand, most law firms 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.

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.

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.

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. And, opinion monitoring software might in future be used to help provide first analysis of the comments.

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 (NLP) technologies that can help the work of the knowledge engineer, so to resolve the resource bottleneck problem.

The future of the Eunomos system rests in the ongoing projects ProLeMAS,[27] BO-ECLI,[28] and MIREL.[29]

ProLeMAS (PROcessing LEgal language in normative Multi-Agent Systems) is a Marie Sklodowska-Curie Individual Fellowship (IF) research project aiming at filling the gap between current logical frameworks designed to represent legal knowledge, mostly propositional, and the richness of NL semantics, for which first-order logical frameworks are needed. The project proposes to use Jerry R. Hobbs's logic (Hobbs 1998; Robaldo and Miltsakaki 2014) for the semantic representation of legal text and to integrate it in Input/Output logic (Makinson and van der Torre 2000), which appears as one of the new achievements in deontic logic in recent years (Gabbay et al. 2013). Finally, ProLeMAS aims at implementing a concrete NLP pipeline for automatically building formulae from existing legal documents. The pipeline will be integrated in Eunomos; specifically, we are defining rule-based methods to extract norms from the XML stored in Eunomos and associate them with logical formulae on which it will be possible to perform reasoning.

BO-ECLI (Building On the European Case Law Identifier) is an e-Justice project (JUST/2014) aiming at developing a (backwards compatible) 2.0 version of the ECLI-standard,[30] and at implementing an open-source software toolkit for computer-based extraction of legal links, to be connected with ECLI search engine of the European e-Justice portal. The mentioned open-source toolkit will be employed in Eunomos for creating references with the case law mentioned in legal documents.

MIREL (MIning and REasoning with Legal text) is a Marie Sklodowska-Curie Research and Innovation Staff Exchange (RISE) research project, i.e. it funds short-term exchanges for staff belonging to the partners in order to promote networking opportunities, sharing of knowledge and the skills development of staff members. MIREL involves sixteen academic and industrial partners, at least one for each continent, among which the University of Torino and the spin-off Nomotika s.r.l..

[27] http://www.liviorobaldo.com/ProLeMAS.htm.

[28] http://www.bo-ecli.eu.

[29] http://www.mirelproject.eu.

[30] http://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-en.do.

The project will create an international and inter-sectorial network to define a formal framework and to develop tools for mining and reasoning with Legal texts, with the aim of translating these legal texts into formal representations that can be used for querying norms, compliance checking, and decision support.

6.1 Planned activities

In what follows, we list the concrete activities we are going to implement in Eunomos, in the context of the ongoing research projects mentioned above.

6.1.1 Generating consolidated legal text

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 and Brighi (2002). 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.

6.1.2 Implementing multilingual search engines

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 (2004)'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 (Boer and Winkels 2005), 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.

6.1.3 Extending the coverage of the ontology

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. 2008a, b) and syntax-based Machine Learning algorithms (Boella et al. 2013, 2014; Boella and Di Caro 2013). In our long-term research plans we aim at associating norms with (extended) Input/Output logic formulae whose predicates are 1:1 connected with the classes of the reference ontology, thus enabling automatic inferences on the addresses of the norms. Robaldo et al. (2016) presents an initial research study in that direction.

6.1.4 Information extraction and business process management

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.

Boella et al. (2012b) we are also developing the conceptual model of roles in prescriptions using the model of Boella and van der Torre (2007).

7 Related work

The proposal closest to the Eunomos system is the "Fill the gap" project by Palmirani et al. (2012). 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. (2009) in that it is designed to help users view laws and classify terms. While Bianchi et al. (2009) 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. Furthermore, 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. (2009) use the Semantic Turkey (Griesi et al. 2007) 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. (2009) make use of statistical and reference data to help users find related norms though (Bianchi et al. 2009) combines these elements by factoring incoming and outgoing references into its statistical model.

Concerning text similarity, Bianchi et al. (2009) 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 (Ounis et al. 2006) 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 (2004) 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. (2010)'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. (2010)'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. (2005) 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', etc.

Concerning the idea of developing collaborative tools for building knowledge in the public sector, it is relevant to refer to Ghidini et al. (2010)'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 (McCarty 1989; Stamper 1991; Breuker et al. 1997). In particular the framework presented in Kralingen (1997) is a frame-based system that classifies the legal facts. A basic component of this system is the legal concept description, i.e., Kralingen (1997) proposes a distinction between a legal term and a legal concept similar to the distinction that we have adopted in the Legal Taxonomy Syllabus ontology.

From a practical point of view, there are two projects that are related in someway to the Legal Taxonomy Syllabus ontology of the Eunomos system. The "EURLex" system[31] 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[32] 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 ontology, 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 Eurovoc does not distinguish between legal terms and legal concepts, and cannot resolve easily the problems related to the polysemy.

Related work on legal ontologies include also Peters et al.'s (2007) LOIS database of legal terms, which adopted the structure of WordNet (Fellbaum 1998) and EuroWordNet (Peters et al. 1998). 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 the Eunomos system is concerned with the access of human experts to the EU documents.

Agnoloni et al. (2009)'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. (2009) separate concepts from terms. However, unlike the

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

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

Legal Taxonomy Syllabus ontology, 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 Sect. 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 Eunomos system resolves the resource bottleneck by decoupling all competences needed to build a large reliable legal knowledge base for regulatory compliance.

We need to overcome, on the one hand, the limitation of the manual updating of the knowledge bases—this would be highly time-consuming and error-prone—and,

on the other hand, to support current NLP technologies that, even at the best of their performances are however unable to fully-automatically carry out the work.

The Eunomos employs a semi-automatic approach to build and update the knowledge bases, where user-friendly interfaces allow knowledge engineers to enter a massive amount of data without the intervention of experts in the underlying technologies, who are required to modify them in rare and exceptional occasions only. In other words, knowledge engineers do not need to have any competence in machine-readable formalisms, NLP, or the other technologies used in the system. During their daily work, knowledge engineers enter new data into the database by correcting the reference links and the document domains automatically *suggested* them by the system, and possibly adding further explanations in plain text. The domain classifier is periodically re-trained on the new (enlarged) training sets. Concerning the rule-based procedures, an expert in NLP, by periodically looking at the missing or incorrect linguistic patterns found by these procedures, decides if and how modifying the rules. Nevertheless, a revision of the rule is indeed rarely required in that legal texts are usually plenty of recurring linguistic patterns and have a limited lexicon, thus the current set of rules is already able to find the correct links in the majority of cases.

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 the University of Torino.

Acknowledgments Guido Boella, Luigi Di Caro, and Leendert van der Torre have received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie Grant Agreement No. 690974 for the project "MIREL: MIning and REasoning with Legal texts". Livio Robaldo has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 661007 for the project "ProLeMAS: PROcessing LEgal language in normative Multi-Agent Systems".

References

Agnoloni T, Barrera MF, Sagri M, Tiscornia D, Venturi G (2009) When a framenet-style knowledge description meets an ontological characterization of fundamental legal concepts. In: AICOL 2009, pp 93–112

Ajani G, Ebers M (eds) (2005) Uniform terminology for European contract law. Nomos, Baden Baden

Ajani G, Lesmo L, Boella G, Mazzei A, Rossi P (2007) 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). ACM, pp 43–48

Aluç G, Özsu MT, Daudjee K (2014) Workload matters: why rdf databases need a new design. Proc VLDB Endow 7(10):837–840. doi:10.14778/2732951.2732957

Biagioli C, Francesconi E, Passerini A, Montemagni S, Soria C (2005) Automatic semantics extraction in law documents. In: Proceedings of the tenth international conference on artificial intelligence and law (ICAIL). ACM, pp 133–140

Bianchi M, Draoli M, Gambosi G, Pazienza M, Scarpato N, Stellato A (2009) ICT tools for the discovery of semantic relations in legal documents. In: Proceedings of the 2nd international conference on ICT solutions for justice (ICT4Justice)

Boella G, Di Caro L (2013) Extracting definitions and hypernym relations relying on syntactic dependencies and support vector machines. In: ACL (2), pp 532–537

Boella G, van der Torre L (2007) The ontological properties of social roles in multi-agent systems: definitional dependence, powers and roles playing roles. Artif Intell Law 15(3):201–221
Boella G, di Caro L, Humphreys L (2011) Using classification to support legal knowledge engineers in the Eunomos legal document management system. In: Fifth international workshop on Juris-informatics (JURISIN)
Boella G, di Caro L, Humphreys L, Robaldo L (2012a) Using legal ontology to improve classification in the Eunomos legal document and knowledge management system. In: Semantic processing of legal texts (SPLET) at lrec12
Boella G, Humphreys L, van der Torre L (2012b) The role of roles in Eunomos, a legal document and knowledge management system for regulatory compliance. In: Proceedings of the information systems: a crossroads for organization, management, accounting and engineering (ITAIS) conference
Boella G, Martin M, Rossi P, van der Torre L, Violato A (2012c) 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. Springer, Berlin
Boella G, Di Caro L, Robaldo L (2013) Semantic relation extraction from legislative text using generalized syntactic dependencies and support vector machines. In: Theory, practice, and applications of rules on the Web. Springer, Berlin, pp 218–225
Boella G, Di Caro L, Ruggeri A, Robaldo L (2014) Learning from syntax generalizations for automatic semantic annotation. J Intell Inf Syst 43(2):231–246
Boer A, Winkels R (2005) What's in an interchange standard for legislative XML? I Quad 18:32–41
Bornea MA, Dolby J, Kementsietsidis A, Srinivas K, Dantressangle P, Udrea O, Bhattacharjee B (2013) Building an efficient RDF store over a relational database. In: Proceedings of the 2013 ACM SIGMOD international conference on management of data. ACM, New York, NY, USA, pp 121–132. Retrieved from doi:10.1145/2463676.2463718
Breaux TD (2009) Legal requirements acquisition for the specification of legally compliant information systems (unpublished doctoral dissertation). North Carolina State University, Raleigh
Breuker J, Valente A, Winkels R (1997) Legal ontologies: a functional view. In: Proceedings of the 1st legout workshop on legal ontologies, pp 23–36
Cheng CP, Lau GT, Law KH, Pan J, Jones A (2008a) Regulation retrieval using industry specific taxonomies. Artif Intell Law 16:277–303
Cheng CP, Pan J, Lau GT, Law KH, Jones A (2008b) Relating taxonomies with regulations. In: Proceedings of the 2008 international conference on digital government research, pp 34–43
Cherubini M, Tiscornia D (2010) An ontology-based model of procedural norms and regulated procedures (Tech. Rep. No. 1/2010). ITTIG-CNR, Florence, Italy
Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
de Maat E, Krabben K, Winkels R (2010) Machine learning versus knowledge based classification of legal texts. In: Proceedings of the legal knowledge and information systems conference: Jurix 2010. IOS Press, pp 87–96. Retrieved from http://portal.acm.org/citation.cfm?id=1940559.1940573
Fellbaum C (1998) WordNet: an electronic lexical database. MIT Press, Cambridge
Fernandez-Barrera M, Casanovas P (2011) 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 the international conference on artificial intelligence and law workshop, applying human language technology to the law, pp 18–27
Fillmore C, Collin F (2000) FrameNet: frame semantics meets the corpus (U. manuscript, Ed.)
Gabbay D, Horty J, Parent X, van der Meyden R, van der Torre L (2013) Handbook of deontic logic and normative systems. College Publications, London
Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L(2002) Sweetening ontologies with dolce. In: Proceedings of the EKAW 2002. Siguenza, SP
Ghidini C, Rospocher M, Serafini L(2010) Moki: a wiki-based conceptual modeling tool. In: Proceedings of the EKAW2010 poster and demo track, vol 674, Lisbon, Portugal
Greene J (2001) Feature subset selection using Thornton's separability index and its applicability to a number of sparse proximity-based classifiers. In: Proceedings of the annual symposium of the pattern recognition association of South Africa
Griesi D, Pazienza MT, Stellato A (2007) Semantic turkey—a semantic book-marking tool (system description). In: 4th European semantic Web conference (ESWC 2007), vol 4519. Springer, Berlin, pp 779–788

Hall M, Eibe F, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The Weka data mining software: an update. In: SIGKDD exploration newsletter, vol 11, pp 10–18
Hobbs J (1998) The logical notation: ontological promiscuity. In: Chapter 2 of discourse and inference. http://www.isi.edu/∼hobbs/disinf-tc.html
Holmes N (2011) Accessible law. http://blog.law.cornell.edu/voxpop/2011/02/15/accessible-law/
Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. Mach Learn ECML–98:137–142
Kralingen V (1997) A conceptual frame-based ontology for the law. In: Proceedings of the 1st legout workshop on legal ontologies, pp 15–22
Lau G (2004) A comparative analysis framework for semi-structured documents, with applications to government regulations (unpublished doctoral dissertation). University of Stanford
Lesmo L (2009) The Turin University Parser at Evalita 2009. In: Proceedings of the EVALITA, vol 9
Lesmo L, Mazzei A, Radicioni DP (2009) Extracting semantic annotations from legal texts. In: Proceedings of the 20th ACM conference on hypertext and hypermedia. ACM, New York, NY, USA, pp 167–172
Makinson D, van der Torre LWN (2000) Input/output logics. J Philos Logic 29(4):383–408
McCarty L (1989) A language for legal discourse: Basic features. In: Proceedings of the second international conference on artificial intelligence and law
Nayak A, Poriya A, Poojary D(2013 March). Article: type of NoSQL databases and its comparison with relational databases. Int J Appl Inf Syst 5(4):16–19. (Published by Foundation of Computer Science, New York, USA)
Neumann T, Weikum G (2010) The RDF-3X engine for scalable management of RDF data. VLDB J 19(1):91–113. doi:10.1007/s00778-009-0165-y
Ounis I, Amati G, Plachouras V, He B, Macdonald C, Lioma C (2006) Terrier: a high performance and scalable information retrieval platform. In: Proceedings of the ACM SIGIR'06 workshop on open source information retrieval (OSIR 2006)
Palmirani M (2011) Legislative change management with akoma-ntoso. In: Sartor G, Palmirani M, Francesconi E, Biasiotti M (eds) Legislative XML for the semantic Web, vol 4, pp 101–130. Springer, Berlin
Palmirani M, Brighi R (2002) Norma-system: a legal document system for managing consolidated acts. In: Proceedings of the database and expert systems applications conference, dexa, vol 2453. Springer, Berlin, pp 310–320
Palmirani M, Ognibene T, Cervone L (2012) Legal rules, text, and ontologies over time. In: Proceedings of the RuleML 2012
Peters W, Vossen P, Díez-Orzas P, Andriaens G (1998) Cross-linguistic alignment of wordnets with an inter-lingual-index. Comput Hum 32(2–3):221–251
Peters W, Sagri M, Tiscornia D (2007) The structuring of legal knowledge in lois. Artif Intell Law 15(2):117–135
Platt J (1999) Sequential minimal optimization: a fast algorithm for training support vector machines. Adv Kernel Methods Support Vector Learn 208:98112
Robaldo L, Miltsakaki E (2014) Corpus-driven semantics of concession: Where do expectations come from?. Dialogue & Discourse 5(1):1–36
Robaldo L, Humphreys L, Sun L, Cupi L, Santos C, Muthuri R (2016) Combining input/output logic and reification for representing real-world obligations. In: Post-proceedings of 9th international workshop on Juris-informatics (JURISIN 2015), lecture notes in artificial intelligence
Rossi P, Vogel C (2004) Terms and concepts; towards a syllabus for European Private Law. Eur Rev Priv Law 12(2):293–300
Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manag 24(5):513–523
Sartor G (2011) Access to legislation in the semantic Web. In: Biasiotti M, Faro S (eds) From information to knowledge—online access to legal information: methodologies, trends and perspectives. IOS
Stamper R (1991) The role of semantics in legal expert systems and legal reasoning. Ratio Juris 4(2):219–244
Visser P, Bench-Capon T (1998) A comparison of four ontologies for the design of legal knowledge systems. Artif Intell Law 6:27–57