Populating legal ontologies using semantic role labeling

This is the author's manuscript

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/1762571 since 2025-02-28T10:32:47Z

Published version:
DOI:10.1007/s10506-020-09271-3

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
Populating Legal Ontologies using Semantic Role Labeling

Llio Humphreys, Guido Boella, Leendert van der Torre, Livio Robaldo, Luigi Di Caro, Sepideh Ghanavati, Robert Muthuri

Abstract This article seeks to address the problem of the ‘resource consumption bottleneck’ of creating legal semantic technologies manually. It describes a Semantic Role Labeling based information extraction system to extract definitions and norms from legislation and represent them as structured norms in legal ontologies. The output is intended to help make laws more accessible, understandable, and searchable in a legal document management system.

Keywords classification · information extraction · ontology · normative reasoning · semantic role labelling · artificial intelligence · law

1 Introduction

Paper-based laws and regulations are not accessible to most of the population. 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 the Open Government Data and Linked Open Data initiatives. However, publishing laws online does not make them truly accessible. Sartor (2011), page 7, envisages a future legal semantic web where legal contents on the web will be enriched with machine processable information: “This information will then be automatically presented in many different ways, according to the different issues at stake and the different roles played by its users (legislators, judges, administrators, parties in economic or other transactions).” This article endorses the vision of providing machine processable information about laws suited to the needs of different users. However, the research that is the subject of this article is informed by the terminological and frame-based legal ontologies of Eunomos legal document management system (Boella et al. 2012, Boella et al. 2016) and Legal-URN compliance management system (Ghanavati 2013) because these systems show the kind of information
required by individuals and organisations from legislation, and provide suitable representations for structuring norms. These systems are described in section 4 below.

This article is concerned with a major obstacle to the goal of maintaining legal information and compliance systems: the 'resource consumption bottleneck' (Hepp, 2007) of creating semantic technologies manually. The use of automated information extraction techniques could significantly reduce this bottleneck. However, Lenci et al. (2009) argue:

Technologies in the area of knowledge management and information access are confronted with a typical acquisition paradox. As knowledge is mostly conveyed through text, content access requires understanding the linguistic structures representing content in text at a level of considerable detail. In turn, processing linguistic structures at the depth needed for content understanding presupposes that a considerable amount of domain knowledge is already in place.

The research question of this article is: How to address the resource bottleneck problem of creating specialist knowledge management systems? In particular, how to semi-automate the extraction of definitions, norms and their elements to populate legal ontologies?

This article shows that the acquisition paradox can be addressed by combining state-of-the-art general-purpose natural language processing modules with pre- and post-processing using rules based on domain knowledge. The output is intended to help make laws more accessible, understandable and searchable in a legal document management system. This research is not concerned with automated legal interpretation (e.g. Walton et al. (2016), Malerba (2017)) or automated legal reasoning (e.g. Robaldo et al. (2019), Robaldo and Sun (2017), Governatori et al. (2013)). Another limitation is that this article is concerned with legislation, and not with case law and legal scholarship, although the approaches proposed could be adopted for such sources in the future.

The article proceeds as follows. Section 2 provides a literature review of legal ontology learning and information extraction. Section 3 provides a summary of Semantic Role Labeling. Section 4 describes relevant features of the systems on which our ontology population methodology is based. Section 5 describes our methodology for extracting definitions using Semantic Role Labeling, while section 6 shows how using the same methodology can be used to classify and extract different types of norms and their elements. Section 7 provides an evaluation of our approach, while section 8 ends the article with conclusions and ideas for future work.

2 Legal Ontology Learning and Information Extraction

Research on ontology learning (creating new ontologies) and ontology population (populating existing ones) is an important field of ontology engineering,
albeit not without limitations. Biemann (2005) states that “none of the methods used today are good enough for creating semantic resources of any kind in a completely unsupervised fashion, albeit automatic methods can facilitate manual construction to a large extent”. While ontologies can be learned from structured and unstructured data, most research on ontology population is based on extracting data from unstructured text.

Research on extracting simple concepts and their inter-relationships (rather than definitions) can involve rules or machine learning. Many concepts and ontological relations can be extracted based on simple patterns. Hearst (1992) used the patterns “X, Ys and other Zs” and “Ws such as X, Y and Z” to extract IS-A relations. Berland and Charniak (1999) used similar patterns to find PART-OF relations. A pattern-matching approach was also employed by Walter (2009) and de Maat et al. (2009) to extract legal definitions from court decisions and legislation respectively. An alternative approach is semantic clustering. Following from the distributional hypothesis of Harris (1968) that “the meaning of entities, and the meaning of grammatical relations among them, is related to the restriction of combinations of these entities relative to other entities”, Hindle’s system classifies words according to their predicate-argument structures, i.e. nouns can be classified according to the verbs they occur with, and verbs according to the nouns they occur with.

For frame-based ontologies, we need to look at information extraction, in particular template filling. Information extraction from natural text is challenging because of language variability: the fact that the same information can be expressed with different words and syntactic constructs. Examples of language variability are lexical synonymy, syntactic word order and world knowledge inferences.

Traditionally, information extraction is approached in a supervised manner based on a set of examples expressing the relations or entities and constructed manually. The Simple Information Extraction (SIE) of Giuliano et al. (2006) is a modular information extraction system designed to be easily and quickly portable across tasks and domains. SIE is composed of a general purpose machine learning algorithm, the Support Vector Machine (SVM), combined with several customisable modules. A crucial role in the architecture is played by an instance filtering module, which is based on the assumption that entities to be recognised are unlikely to have low information content (Gliozzo et al. 2005).

The core of many unsupervised information extraction systems, such as those of Yangarber et al. (2000) and Stevenson and Greenwood (2005), are ‘paraphrasing’ modules to generate semantically equivalent components with lexical or syntactic variation. The synonym sets in the WordNet (Miller et al. 1990) general-purpose lightweight ontology are also very useful for this purpose, and have been used by Moldovan and Rus (2001) and Mihalcea and Moldovan (2000). Syntactic word order patterns, such as active/passive formulations can be generated according to standard template rules and grouped together in equivalence classes. The TEASE system of Szpektor et al. (2004) is a generic paraphrasing extraction system that extracts relations between a
pivot (lexical entry) and a template (dependency parse fragment). The surrounding words of the lexical entries are used as anchor sets to extract templates.

An alternative approach to handling language variability is the transformation of text into logical form. The system of [Rus (2002)] transformed WordNet concept definitions into logical forms designed to be as close as possible to natural language. The notation module was developed for a question-answering system. The answer extraction procedure consisted of four steps: transforming questions and answers into logic forms, forming WordNet-based lexical chains between pairs of concepts, unifying lexical chains, and extracting inferences. The advantage of using logic form transformation as part of an information extraction system is that the knowledge could be used by various applications that require different information to be structured in different ways. However, this approach has already been tried and tested on legal text. [Wyner (2012)] explored the use of C&C/Boxer (Curran et al. (2007)) to extract norms from UK citizenship legislation, concluding that such systems perform better on Controlled English than natural language constructions typically found in legislative text.

[Wyner and Peters (2012) 2011] have researched information extraction for legislation and case reports. Their approach is semantic annotation of text, creation of a gold standard, and development of automated annotation tools. The gold standard requires annotations from domain experts, and they use the Teamware unified environment for these tasks. The Teamware system uses the open source General Architecture for Text Engineering (GATE) tool for information extraction (Cunningham et al. (2002)) to pre-annotate the text, thereby removing some aspects of the annotation task for domain experts. The GATE platform enables template-based extraction on the basis of lookup lists (gazetteers), heuristic pattern-based grammars as well as a pipeline of standard natural language processing components such as tokenization, sentence splitting, part-of-speech tagging, morphological analysis (lemmatisation), verb and noun phrase chunking, and a parser. A simpler approach is used in the SALEM system (Bartolini et al. (2004)), which is tasked with assigning each legislative paragraph a provision type and automatically tagging the parts of the paragraph with domain-specific semantic roles identifying the legal entities (i.e. actors, actions and properties). The text is POS-tagged and shallow parsed into chunks. The recognition of provision types and the subsequent role identification are based on structural patterns which are combined with lexical conditions (using regular expressions) and dependency relations.

There is much information extraction research involving machine learning in the legal domain. Of particular relevance to this article is the extraction of active roles, passive roles and involved objects in norms by [Boella et al. (2013b)]. Based on the idea that a semantic tag may be characterised by limited sets of syntactic contexts, their supervised Machine Learning approach involves the use of syntactic dependencies extracted with the parser TULE as factors in a Support Vector Machine classifier (Cortes and Vapnik (1995)). [Gao and Singh (2014)] use pattern-matching and machine learning to extract
commitments, authorisations, powers, prohibitions and sanctions from business contracts. The objective is similar to section 6 of this article, in that they first classify the norm type, and then extract the elements of the norm. They identify norms based on use of modal verbs. They use a classifier for identifying norms using verb and clause conjunctions. Elements of norms are extracted based on heuristics such as ‘If a norm sentence has a subordinate clause led by conjunction words such as “if” and “unless”, the subordinate clause expresses the antecedent.’ Grabmair et al. (2011) and Biagioli et al. (2005) also use machine learning approaches for classifying norms and extracting elements of legislation. Their methodology relies on the costly labour-intensive task of annotating legal corpora, and is less sensitive to specific linguistic expressions that are not commonly used in the corpus. Moreover, this makes it more difficult to fine-tune the extraction process.

Semantic role labeling (see section 3) has emerged as a suitable intermediary for unsupervised information extraction (Surdeanu et al. 2003; Kaisser and Webber 2007). To be clear, while the identification of general thematic roles is based on a general-purpose system trained on a general corpora, the information gained can then be used in a simpler system for domain-specific template filling. However, there is little work on SRL for information extraction in legal informatics. Venturi et al. 2009 use a FrameNet based methodology for extracting norms. Their methodology relies on a laborious methodology of building a legal domain extension to the general FrameNet, and producing a legal corpus annotated with frame information. Palmirani et al. 2011 use a FrameNet approach to extract suspensions of norms in a three-step process: identification of relevant sentences, syntactic analysis, and semantic annotation using tree matching. Bertoldi and de Oliveira Chishman 2011 investigated the use of FrameNet for building a legal ontology for Brazilian law, but decided against it, as they found that the legal frames were neither language nor jurisdiction independent as intended. While PropBank Semantic Role Labeling has been used in legal informatics for Abstract Meaning Representation (AMR) Viet et al. 2017 and event extraction for legal case retrieval Maxwell et al. 2009, the work presented in this article is, to our knowledge, the first attempt to extract definitions and norms from legislation using a PropBank Semantic Role Labeler Kingsbury and Palmer 2002. It is submitted that this approach could achieve wider coverage than FrameNet Fillmore et al. 2004, as has been the observation in other information extraction research e.g. Kaisser and Webber 2007. Moreover, the use of shallow rather than deep roles allows for greater flexibility in the selection and classification of data extracted, tailored to the requirements of the relevant application.

3 Semantic Role Labeling

Semantic Role Labeling (SRL) is the task of detecting basic event structures in a sentence such as “who” did “what” to “whom”, “when” and “where” Marquez 2009. A semantic role (also known as semantic case, thematic
role, theta role or case role) is the underlying relationship that a participant has with the main verb in a clause (Loos et al. (2004); Payne (1997)).

Many verbs allow a variable number of semantic roles to be realized in various syntactic positions (diathesis alternations), as can be seen in the following example from Martin and Jurafsky (2016):

[John/AGENT] broke [the window/THEME].
[John/AGENT] broke [the window/THEME] with [a rock/INSTRUMENT].
[The rock/INSTRUMENT] broke [the window/THEME].
[The window/THEME] broke.
[The window/THEME] was broken by [John/AGENT].

While an agent or instrument can both be the grammatical subjects of a sentence, a sentence such as the following is grammatically unacceptable:

**John and a hammer broke the window**

Fillmore (1968) explained why: only noun phrases representing the same case may be conjoined. Fillmore used the term ‘case’ with reference to the case system that exist in many languages. For instance, in ancient Greek or Russian, the phrase ‘with a rock’ in the sentence “John broke the window with a rock” would be expressed with a single noun with the instrumental case marker, which is different from the nominative case used for “John”. In English, the instrumental case is ‘flagged’ by the preposition ‘with’. However, there is no rigid one-to-one mapping between flags and cases -‘with’ can also flag the cases ‘Manner’ (‘with glee’) and ‘Accompanier’ (‘with Nadia’) (Hirst (1992)). Levin (1993) noted that syntactic constraints on verbs and the arguments they may take are semantically determined, and created verb classes whose members pattern together with respect to diathesis alternations.

There is no consensus on a definitive list of semantic roles (Márquez et al. (2008)) that should be used for semantic role labeling. In FrameNet (Fillmore et al. (2004)), which is based on frame semantics (Fillmore (1976)), arguments are related to deep roles related to specific scenarios or frames, such as Suspect, Authorities and Offense. PropBank (Kingsbury and Palmer (2002)), on the other hand, uses general roles and verb-specific roles based on the verb classes of Levin (1993). The roles are numbered, rather than given semantic names, although in general, Arg-0 corresponds to Agent while Arg-1 corresponds to Patient. Both FrameNet and PropBank use extra-thematic elements such as Time, Manner and Place. Márquez et al. (2008) assert that most research on SRL is now conducted on PropBank, mainly because of its greater coverage. While the FrameNet corpus is a selection of illustrative sentences, PropBank has annotated the semantic roles of all verbs in the Penn Treebank II Wall Street Journal Corpus (Marcus et al. (1994)).

SRL systems rely on automated part-of-speech tagging and parsing. Most SRL systems use constituency parsers, probably because they have traditionally been better resourced for the English language. However, Johansson and Nugues (2008b,a) argue that dependency structures offer a more transparent encoding of predicate argument relations (e.g. grammatical function such
as subject and object is an integral concept in dependency syntax) and thus dependency structures are more suitable for explaining the syntax-semantics interface (Mel’cuk et al. (1988); Hudson (1984)). Moreover, in their comparison of constituent-based and dependency-based SRL systems for FrameNet, they found that their performance was roughly the same, except that dependency-based systems outperformed constituent-based systems when using out-of-domain test sets, due to their lesser reliance on lexical features (Johansson and Nugues (2008b)). Another reason for choosing SRL systems based on dependency parsers is that they can be more efficient (Ciaramita et al. (2008)), rendering them more suitable for real-life applications (Surdeanu et al. (2008)).

Most automated SRL systems follow this three step architecture:

- filtering (or pruning) the set of argument candidates for a given predicate;
- local scoring of argument candidates for possible role labels, including a ‘no-argument’ label;
- joint scoring to combine the predictions of local scorers and ensure that the eventual labeling satisfies structural and SRL-dependent constraints.

The Mate Tools Semantic Role Labeler (Björkelund et al. (2009)) used for this article follows this architecture. The first stage consists of a pipeline of independent classifiers. It carries out the predicate disambiguation with a set of greedy classifiers, where one classifier is applied per predicate lemma. It then uses a beam search to identify the arguments of each predicate and to label them, yielding a pool of candidate propositions. The second stage consists of a reranker that is applied to the candidates using local models and proposition features. It combines the score of the greedy classifiers and the reranker in a third stage to select the best candidate proposition. Mate Tools came second in the CoNLL-2009 shared SRL task on the joint parsing of syntactic and semantic dependencies in the closed challenge (Hajič et al. (2009)), and is freely available to download or use via an online demo.

The definition and norm extraction system is based on abstract semantic representations from a general-purpose SRL module, as well as dependency parse trees from the same system. This simplifies the sets of rules required which in cascade identify possible norms and definitions, classify their types, and then on the basis of their types, map arguments in a semantic role tree to domain-specific slots in a legal ontology. The use of an abstract semantic representation also allows flexibility in the determination of how norms are structured so that the system can easily be modified for different applications. To evaluate the definitions and norm extraction system, the system was tested on both seen and unseen data from a corpus of EU Directives.

4 Systems on which Ontology Population Methodology is Based

The research that is the subject of this article is highly influenced by the European Legal Taxonomy Syllabus (Ajani et al. (2016)) lightweight ontology framework, the terminological and frame-based legal ontologies of Eunomos
legal document management system (Boella et al. (2012), Boella et al. (2016)) and LEGAL-URN compliance management system (Ghanavati (2013)) because these systems show the kind of information required by individuals and organisations from legislation, and provide suitable representations for structuring norms. This section describes the systems that inspired this research.

4.1 European Legal Taxonomy Syllabus

European Legal Taxonomy Syllabus (Ajani et al. (2016)) was developed in the context of the UT (Uniform Terminology For European Private Law) project (Ajani and Ebers (2005), Rossi and Vogel (2004)). The research network involved researchers from 7 universities spread across England, France, Germany, Italy, Netherlands, Poland, and Spain. The aims were better understanding of the historical divergences hampering uniform terminology, and the promotion of a common terminology in EU private law. The motivation for the development of a multilingual legal ontology framework was that anyone involved in trans-national activities would benefit from access to user-friendly specialist ontologies to manage the deep, complex and interconnected terminology required for understanding laws in different jurisdictions. A legal ontology framework that handles these issues could be useful for many different types of users e.g. lawyers who deal with cross-border issues, international financial institutions, or EU translators.

The main assumptions of the European Legal Taxonomy Syllabus ontology framework come from studies in comparative law (Rossi and Vogel (2004)) and ontologies engineering (Klein (2001)). European Legal Taxonomy Syllabus
Postprint version of Populating Legal Ontologies ... 9

(ELTS) is based on a clear distinction between a legal term and a legal concept. The concepts are arranged into ontologies. Lexical terms are then linked to all concepts that apply. Making a clear distinction among terms and their interlingual acceptions (or axes) is a standard way to properly manage large multilingual lexical databases ([Sérasset 1995], [Lyding et al. 2006]). This approach is also well suited to complex interlinking domains. The system does not assume the existence of a single taxonomy covering all languages since different national systems may organize legal concepts in different ways. It is a framework of interlinked but separate ontologies for different languages and jurisdictions in Europe.

The challenge in developing European Legal Taxonomy Syllabus was to design an ontology framework that captures the complex factors that influence the meaning of legal terms, but at the same time structures that knowledge in a way that leads to clarity rather than confusion. ELTS is not a formal ontology in the standard sense, i.e. an axiomatic ontology formalized, for instance, in description logic ([Palmirani et al. 2018]). Rather, it is a lightweight ontology, i.e. a knowledge base storing low-level legal concepts, connected via low-level semantic relations, and related to linguistic patterns that denote legal concepts.

For the end-user, each term in each ontology is presented in an associated information table with the following information:

- language
- jurisdiction
- domain
- description in natural language
- references to legislative definitions
- links from national terms (and their definitions) to European ones and vice versa
- links to equivalent terms in other languages
- a list of other terms having the same meaning
- an ontology graph showing parent and child concepts
- notes

The descriptions in natural language are made by legal experts who seek to explain but do not over-simplify legal issues. The notes field carries information about court decisions, scholarly interpretations or other information of interest.

Each ontology within European Legal Taxonomy Syllabus ontology is hierarchical and the conceptual tree allows users to view hyperonomy/meronomy/synonymy relations. Figure 2 shows a concept tree for vehicles with the hyponyms being trolley-buses, motorcycles etc.

The ELTS bottom-up approach is suitable for dealing with legal interpretation and the evolution of terms. As more legislation-specific definitions about terms are collected, it becomes possible to find common attributes and derive a more general definition. The general definition is then linked to all the definitions that informed it via a group_by relation. Legislation-specific and generic
definitions can co-exist, with generic definitions grouping legislative definitions together with doctrinal interpretation.

Another important issue is that when new legislation is approved and enacted, it can introduce a number of new definitions which render old definitions obsolete. ELTS handles this issue by introducing a temporal dimension into the ontology, allowing new concepts to replace old ones while still retaining the old concepts in the system for reference purposes. This is achieved with the replaced_by relation.

For the UT project, terms were collected in a database together with the legal sources where they appeared, 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 were organized by legal experts into ontologies which can be different for different legal traditions. The result is lightweight rather than axiomatic ontologies. The function of these ontologies is to compare the taxonomic structure in the different legislation, and to provide a form of intelligent indexing.

The ELTS ontology framework is well-integrated within the Eunomos legal document management system described below, so that links can be made between concepts, the terms used to express the concepts, and items of the laws that feature the terms. Vice versa, terms in the text of legislation are annotated with references to the concepts.

4.2 Eunomos

The second system of relevance is Eunomos - a legal knowledge document management system to help users keep track of the law and ever-changing legal obligations (Boella et al. (2016)). The Eunomos system is envisaged as being useful to a wide range of user groups but is targeted towards compliance officers in the first instance, because they have the greatest need and enthusiasm for a system of this kind. The basic idea of Eunomos is to create a stricter coupling between legal knowledge and its legislative sources, associating the concepts of its legal ontology with regulations structured using legislative XML. The system takes inspiration from technologies developed in the related fields of legal ontologies, extending the European Legal Taxonomy Syllabus (Ajani et al. (2016)) tool for building legal ontologies, and legislative drafting for parliaments (so-called legislative XML). Maintaining laws in 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. Most laws are collected from portals by means of web spiders, but they can also be inserted into the database via a web interface. Eunomos uses the XML Leges Linker tool developed by ITTIG to find cross-references, a rule-based system (Robaldo et al. (2011)) for mining concepts in legal documents, and a

\[\text{www.xmleges.org}\]
In Eunomos, the European Legal Taxonomy Syllabus ontology framework is extended to include prescriptive norms (as opposed to merely terminological definitions found in constitutive norms). A prescription is treated as a composite type of concept, which is subject to all the relations of simple concepts (is_a, part_of, replaced_by etc.), but with additional characteristics. Each prescription is necessarily connected to relevant concepts defined in the terminological ontology via the following relations:

- Deontic clause: the type of prescription: obligation, prohibition, permission, exception.
- Active role: the addressee of the norm (e.g. director, employee).
- Passive role: the beneficiary of the norm (e.g. customer).
- Description: the prescription reworded as necessary to aid comprehension.
- Norm Identifier: hyperlink to relevant provision in the source document.
- Violation: the crime or tort resulting from violation (often defined in other legislation such as a Penal Code).
- Sanction: the sanction resulting from violation (e.g., a fine of 1 ‘quote’, where ‘quote’ is defined in other legislation).

\[1\] Boella et al. (2013a)
\[2\] Boella et al. (2015)
\[3\] http://www.eucases.eu
\[4\] https://op.europa.eu/en/web/eu-vocabularies
Notes: information about court decisions, scholarly interpretations or other information of interest.

Prescriptions are a powerful way to structure information about norms that are usually scattered across different sections, not necessarily from the same piece of legislation. It makes it easy for users to quickly find norms according to various search criteria indicated by the above relations. For instance, a semantic search for prescriptions that relate to sellers will also bring up prescriptions for traders (the grandparent of seller in the terminological ontology).

The ontology of prescriptions can also show different interpretations of the same norm. Although Eunomos is a lightweight ontology, and lacks formal semantics, there is an alternative relation available for prescriptions, inspired from specifications hierarchies where specifications of a concept can be labelled as disjoint. Alternative interpretations should be specified as being candidate or non-candidate, where in the latter case the company has determined that the interpretation is unlikely, undesirable or irrelevant.

4.3 Legal-URN

The final system of interest is Legal-URN (Ghanavati (2013)), a sophisticated system for systematic modelling and analysis of compliance issues that enables business users to factor in legal requirements as part of their strategic planning. The system is based on a Requirements Engineering (RE) approach to legal compliance, and models legal norms in the same notation as goal and business process management, albeit with deontic extensions. Legal-URN supports business process compliance by extending the model-based compliance framework based on the User Requirements Notation (URN) Language (ITU-T (2012)).

The Legal-URN framework, has four layers for legal and organisational models:

1. **Official Source Documents** that define the legislation on one side and organisational structures, policies and processes on the other side.
2. **A Hohfeldian Model** (as expressed by Siena (2010)) which consists of a set of Hohfeldian statements (Wenar (2011)) of norms including structured elements of legal statements.
3. **Goal Models** based on URN’s Goal-oriented Requirement Language (GRL), which capture the objectives and requirements of both organisation and legislation.
4. **Business Process Models** based on URN’s Use Case Maps (UCM), which define the business processes that implement organisational policies on the one hand and represents steps mandated by legislation on the other hand.

To build the Hohfeldian model, the following rules need to be considered:

1. Each legal statement shall be atomic. This means that each legal statement contains one \langle actor \rangle (the subject), one \langle modalverb \rangle, one to many
Then, the Hohfeldian classifications (i.e. duty-claim, privilege-noclaim, etc.) are transformed into Permission and Obligations and the Legal-GRL model of regulation and the GRL model of organization are developed. GRL’s main concepts come from management and from socio-technical systems and include actors, which have intentional elements (goals, softgoals, tasks, and resources) and indicators, linked through various relationships (AND/OR decompositions, dependencies, and weighted contributions (Cai and Yu (2002)). The compliance analysis is done in this step.

While the Goal-oriented Requirement Language (GRL) models the “why” aspect of requirements, Use Case Maps (UCM) model “what” aspects with scenarios (use cases).

LEGAL-URN has some limitations. The framework does not include a regulations repository. Furthermore, developing the Hohfeldian model is currently manual. It also lacks legal interpretations (Bobbitt (2006)) to help identify sets of business process patterns which can be legally compliant. Integration with the Eunomos system described above would solve these problems, as described by Boella et al. (2014).

The Eunomos repository of laws - with legislative XML for clickable cross-references, definitions of terms and their inter-relations in specialist ontologies - replaces the LEGAL-URN “Law and Regulation Documents” level. At the legal provisions level, there is a new representation that integrates Eunomos prescriptions and LEGAL-URN Hohfeldian models. Table 1 shows the mapping of fields and relations between the two representations. Many fields can be mapped directly, some require adaptation, and others are taken from

---

**Fig. 3** Hohfeldian Model Structure and Mapping with Legal-GRL in LEGAL-URN

\[(\text{Clause}) \ (\langle \text{verb} \rangle \ \& \ \langle \text{action} \rangle), \ 0 \text{ to many optional } \langle \text{crossreference} \rangle, \ 0 \text{ to many optional } \langle \text{precondition} \rangle \text{ and } 0 \text{ to many optional } \langle \text{exception} \rangle.\]
### Table 1 Integration of Prescriptions and Hohfeldian Models

| Prescriptions | Hohfeldian Model | Integrated Representation |
|---------------|-----------------|---------------------------|
| Deontic Clause | Hohfeldian Modality | Hohfeldian Modality |
| Active role | Subject | Responsible Actor |
| Passive role | - | Beneficiary |
| Violation | - | Violation |
| Sanction | - | Sanction |
| - Modal Verb | - | - |
| Description | Clause | Clause |
| - Precondition | Precondition | Precondition |
| - | - | - |
| IsA Relation | - | IsA Relation |
| PartOf Relation | - | PartOf Relation |
| Exception Relation | Exception | Exception Relation |
| Norm Identifier | Section + Article | Norm Identifier |
| - Cross-reference | Cross-reference | Cross-reference |
| - | - | Stakeholder |

The ontology structures described above were used as a baseline for developing the ontology population methodology. However, when developing the system, it was found that there were some differences in the types of definitions and norms found in the corpus, and the extraction tool was modified accordingly.

### 5 Extracting Definitions

Populating a legal ontology based on a bottom-up approach necessarily involves the laborious task of storing definitions from all relevant legislation. This section describes how this work can be facilitated with NLP. While most definitions are in the Definitions section of the legislation and follow the regular ‘definiendum equals definiens’ form, there are other less obvious definitions, often found in the normative provisions, which are highly influential. Subsection [5.1] provides the general methodology. Subsection [5.2] provides descriptions of each type of definition and how they are extracted.

#### 5.1 Methodology

To extract definitions, sets of rules were devised which in cascade identify possible norms and definitions, classify their types, and then on the basis of their types, use further rules to map arguments in a Mate Tools Semantic Role Labeler (SRL) (Björcklund et al. [2009]) semantic role tree to domain-specific slots in a legal ontology. The idea behind using SRL is that it enriches one representation.
Table 2  Description of the fields (columns) of a Mate Tools SRL output table

| Field # | Name   | Description                                           |
|---------|--------|-------------------------------------------------------|
| 1       | ID     | Token counter, starting at 1 for each new sentence   |
| 2       | FORM   | Form or punctuation symbol (the token; “split” for English) |
| 3       | LEMMA  | Gold-standard lemma of FORM                           |
| 4       | PLEMMAS| Automatically predicted lemma of FORM                  |
| 5       | POS    | Gold-standard POS (major POS only)                    |
| 6       | PPPOS  | Automatically predicted major POS by a language-specific tagger |
| 7       | FEAT   | Gold-standard morphological features (if applicable)  |
| 8       | PFTEATS| Automatically predicted morphological features (if applicable) |
| 9       | HEAD   | Gold-standard syntactic head of the current token (ID or 0 if root) |
| 10      | PHEADS | Automatically predicted syntactic head                |
| 11      | DEPREL | Gold-standard syntactic dependency relation to HEAD    |
| 12      | PDEPREL| Automatically predicted dependency relation to PHEAD    |
| 13      | FILLPRED| Contains ‘Y’ for argument-bearing tokens               |
| 14      | PRED   | (sense) identifier of a semantic “predicate” coming from a current token (in the ID order) |
| 15...   | APREDn | Columns with argument labels for each semantic predicate |

The input to the SRL is normalised text of legislation. The normalisation module is limited to making legislative text akin to standard written text to facilitate information extraction.

The normalisation module uses pattern-matching and the Brill (1992) part-of-speech tagger in the Python Natural Language Toolkit (NLTK) by Bird (2006) to perform the following tasks:

- sentence and word segmentation (with identification of abbreviations, Unique Resource Locations (URLs), reference numbers, and dates)
- identify titles (uppercase or lacking verb)
- identify and transform references into processable units
- transform certain words to enable SRL handling
- identify lists (and their nestedness) and transform the lists into proper grammatical sentences, by adding introductory clauses and endings to each list item
- concatenate elements of citations e.g. ‘S_Article_2’ to avoid parsing errors

The normalised text is then transformed into an input file with a word index and word surface on each line. The output is a table of semantic role dependencies, in accordance with the specification of The CoNLL-2009 Shared Task (Hajič et al., 2009).

The information extraction system relies on three sets of rules for extracting definitions:
– the first set identifies sentences that contain definition by looking for dependency paths to relevant predicates;
– given the definition sentences identified in the first set, the second set of rules the type of definition;
– given the definitions and definition type, the third set map roles in the SRL tree to ontology slots relevant to the type of definition.

Below we discuss the different types of definitions and the rules for extracting them.

5.2 Definition Types

5.2.1 Regular Definitions

The vast majority of definitions constitute a ‘definiendum equals definiens’ formula, with the ‘equals’ part expressed in a limited number of set patterns.

Here is one example from Directive 98/5/EC

For the purposes of this Directive: ‘host Member State’ means the Member State in which a lawyer practises pursuant to this Directive;

This sentence is identified as a definition whenever the word ‘means’ appears with the part-of-speech VBZ.

Figure 4 shows the arguments available for extraction from the above example sentence. The definiendum and definiens are extracted in accordance with SRL roles. For regular definitions, the Definition is the A1 (SBJ) of the relevant predicate (usually a verb), the Definiens is the A1 (OBJ) and Reason is AM-ADV (ADV). Note that in addition to the traditional ‘Definition’ and ‘Definiens’ elements found in all ontologies, we also have the element ‘Scope’.

The XML output for the purpose of populating the ontology is then:

```xml
<Norm>
  <NormType>Definition</NormType>
  <Definiendum>host Member State</Definiendum>
  <Definiens>
    the Member State in which a lawyer practises pursuant to this Directive
  </Definiens>
  <Scope>for the purposes of this Directive</Scope>
</Norm>
```

4 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

5 In a pre-processing step, each ‘is’ has been converted into ‘become’ to obtain the best SRL results
Fig. 4 SRL extraction of a regular definition
Based on analysis of the training corpus, the following patterns of parser dependencies were collated for the predicates used for regular definitions (where -> indicates sequence and bracketed items are parts of speech):

- Head -> means (VBZ)
- Head -> any sequence of words -> means (VBZ)
- Head -> shall (MD) -> mean
- Head -> constitutes (any POS)
- Head -> any sequence of words -> defined (any POS)

In general, the part of speech is added as a criteria only when there is potential for error (for instance, ‘means’ as a noun, as in ‘means of transport’, has an altogether different meaning to ‘means’ as a verb). Otherwise, parser errors could unnecessarily degrade the performance of this module.

5.2.2 Include/Exclude Definitions

Another kind of definition found in the corpus is one that states which items are included or excluded under a particular category. Include/example definitions are often used to emphasise the inclusion or exclusion of certain items where this would otherwise be uncertain or even surprising. Include/exclude definitions are not explanatory like regular definitions, although often appended to regular explanatory definitions. Moreover, include/exclude definitions are incomplete. There may (or may not) be other items that are included and/or not included. Include definitions can easily be represented in an ontology as a ‘kind-of’ relation. More difficult are exclude definitions. An absence of a kind-of relation is not enough: in the example below, ‘the provision of services’ is not a type of ‘practice of the profession of lawyer’, but then neither is the sun or the moon.

Below is an example of an exclude definition from Directive 98/5/EC:

6 Practice of the profession of lawyer within the meaning of this Directive shall not include the provision of services, which is covered by Directive 77/249/EEC.

The system extracts include/exclude definitions that have the following patterns of parser dependencies:

- Head -> includes (any POS)
- Head -> do (any POS) -> include (VB)
- Head -> does (any POS) -> include (VB)
- Head -> can (MD) -> include (VB)
- Head -> may (MD) -> include (VB)
- Head -> excludes (any POS)

6 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
For this type of definition (unlike regular definitions), the A2 (SBJ) and A1 (OBJ) arguments of the word ‘include’ are extracted and assigned to the fields ‘Definiendum’ and ‘Excludes’ (or ‘Includes’) respectively. All the roles for ‘include’ are displayed below in XML:

```xml
<SRL>
  <PREDICATE>include</PREDICATE>
  <A2>
    Practice of the profession of lawyer
  </A2>
  <AM-MOS>shall</AM-MOD>
  <AM-NEG>not</AM-NEG>
  <A1>
    the provision of services, which becomes covered by S_E77_F249_FEEC.
  </A1>
</SRL>
```

Manipulation of the first argument allows us to extract the ‘Scope’ argument. Here is the structured norm in XML after mapping SRL roles to ontology fields:

```xml
<Norm>
  <NormType>Definition</NormType>
  <Definiendum>
    practice of the profession of lawyer
  </Definiendum>
  <Scope>within the meaning of this Directive</Scope>
  <Excludes>
    the provision of services, which is covered by Directive 77/249/EEC.
  </Excludes>
</Norm>
```

Include/exclude definitions often occur within sections other than that entitled ‘Definitions’. While these may be not have the status of ‘official’ definition, they do provide an important indication of the intentions of the legislature, and therefore are likely to be influential. As such, they are important to include in the ontology. Here is one such example from Directive 2004/108/EC.

---

7 Directive 2004/108/EC of the European Parliament and of the Council of 15 December 2004 on the approximation of the laws of the Member States relating to electromagnetic compatibility and repealing Directive 89/336/EEC Text with EEA relevance.
Member States are responsible for ensuring that radiocommunications, including radio broadcast reception and the amateur radio service operating in accordance with International Telecommunication Union (ITU) radio regulations, electrical supply networks and telecommunications networks, as well as equipment connected thereto, are protected against electromagnetic disturbance.

As with the exclude definition, the A2 (SBJ) and A1 (OBJ) arguments of the word ‘include’ are extracted and assigned to the fields ‘Definiendum’ and ‘Includes’ respectively. The roles for the predicate ‘include’ are:

```xml
<PREDICATE>include</PREDICATE>
<A2>radiocommunications</A2>
<A1>
  radio broadcast reception and the amateur radio service operating in accordance with International Telecommunication Union (ITU) radio regulations, electrical supply networks and telecommunications networks, as well as equipment connected thereto
</A1>
</SRL>

Further work is needed to separate the elements mentioned for the structured norm:

```xml
<NormType>Definition</NormType>
<Definiendum>
  radiocommunications
</Definiendum>
<Includes>
  radio broadcast reception and the amateur radio service operating in accordance with International Telecommunication Union (ITU) radio regulations
</Includes>
<Includes>
  electrical supply networks
</Includes>
<Includes>
  telecommunications networks
</Includes>
<Includes>
  equipment connected thereto
</Includes>
</Norm>
```
5.2.3 **Definition by Example**

Legislation also sometimes contain definitions by example, such as the example below from Directive 98/5/EC.

For the purpose of this Directive: ‘‘signature-creation data’’ means unique data, such as codes or private cryptographic keys, which are used by the signatory to create an electronic signature;

Such definitions are also not explanatory. They are somewhat similar to include definitions except that in this case the instances are typical, and therefore invite extension by analogy. Unlike include/exclude definitions, there is a sense of completeness, that the instances must belong either to the examples or something similar. What is implied in the example above is the following:

```xml
<Norm>
  <NormType>Definition</NormType>
  <Definiendum>signature-creation data</Definiendum>
  <Scope>this Directive</Scope>
  <Definiens>unique data which are used by the signatory to create an electronic signature</Definiens>
  <Includes>codes</Includes>
  <Includes>private cryptographic keys</Includes>
  <Includes>anything similar to codes</Includes>
  <Includes>anything similar to private cryptographic keys</Includes>
</Norm>
```

For brevity and ease of readability, an Example tag may be preferred for a lightweight ontology with an XML output as follows:

```xml
<Norm>
  <NormType>Definition</NormType>
  <Definiendum>signature-creation data</Definiendum>
  <Scope>this Directive</Scope>
  <Definiens>
    unique data which are used by the signatory to create an electronic signature
  </Definiens>
  <Includes>codes</Includes>
  <Includes>private cryptographic keys</Includes>
  <Includes>anything similar to codes</Includes>
  <Includes>anything similar to private cryptographic keys</Includes>
</Norm>
```

---

8 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
unique data which are used by the signatory to create an electronic signature

<Example>codes</Example>
<Example>private cryptographic keys</Example>

Due to time constraints, definitions by example were not extracted for the purpose of this article, and remains the subject of future work.

5.2.4 Definition by Reference

Not every piece of legislation contains every concept, and some legislation explicitly refer to other legislation for definitions of certain concepts. This is an example from Directive 98/44/EC:

The concept of ‘plant variety’ is defined by Article 5 of Regulation (EC) No 2100/94.

It is important to store these definitions, otherwise some other definition might be assumed instead. The references can be resolved within the context of the ontology - by linking to the definition entry or copying its content, depending on the functionalities of the system. For this research, the IE extractor outputs the elements in XML format as before. Each reference is converted into a one-word format, which makes it not only less error-prone for the semantic role labeler, but also easier to convert into legislative XML URNs. They are not converted directly into URNs but rather into a different one-word-format that retains topographic information so that they can be converted back into the original format if necessary.

Here are the SRL roles for ‘define’:

<SRL>
<PREDICATE>define</PREDICATE>
<A1>The concept of plant variety</A1>
<AM-MNR>by S_E_LEC_RNo2100_F94</AM-MNR>
</SRL>

The XML output for ontology population is as follows.

<Norm>
<Definiendum>plant variety</Definiendum>
<Definiens>S_Article_5_of_ Regulation_(EC)_No_2100/94</Definiens>
</Norm>

\(^9\) Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the legal protection of biotechnological inventions
6 Extracting Norms

Section 5 discussed extracting definitions to populate a legal terminological ontology. This section describes the extraction and structuring of norms to populate a frame-based ontology of norms (although strictly speaking, definitions are also norms and are found in the normative provisions section of EU legislation). The Eunomos ontology of legal norms and LEGAL-URN Hohfeldian model present norms in a structured, searchable and intuitive format for representing who must (or may) do what and when. The focus on actions in norms renders this work particularly suitable for a dependency parsing and Semantic Role Labeling approach. Subsection 6.1 provides the general methodology. Subsection 6.2 provides descriptions of each type of definition and how they are extracted.

6.1 Methodology

The methodology for extracting norms is essentially the same as described in section 5. In addition to the three sets of rules for identifying sentences containing norms, the norm type, and the roles in the SRL tree to map to slots in a legal ontology, further rules are needed for extracting nested norms, conditions, exceptions, etc.

Section 6 describes the different types of norms. Here are presented the rules for extracting them. In this section, we will mostly show the relevant SRL output in XML format as the sentences are typically too long to display as per the visualisation tool.

6.2 Types of Norms

6.2.1 Obligation

Consider for example the passive sentence below from 2007/60/EC, which represents an obligation:

Hence, objectives regarding the management of flood risks should be determined by the Member States themselves and should be based on local and regional circumstances.

Figure 5 shows that the SRL tool understands that the agents (role A0) are the Member States, relegating the objectives to the object (A1) role of the verb ‘determine’. Moreover, it also abstracts from the fact that the root of the parse tree is the modal verb followed by an auxiliary. Thus it becomes simpler to write rules on the SRL output than on the parse tree.

---

10 Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks (Text with EEA relevance).
Fig. 5 SRL extraction of a passive obligation
| determine.01 | AM-DIS | AM-MOD | A0 | C-AM-MOD | A0 | A1 | A1 | A1 | A1 | A2 |
|-------------|--------|--------|----|----------|----|----|----|----|----|----|
| objective.01|        |        |    |          |    |    |    |    |    |    |
| management.01|        |        |    |          |    |    |    |    |    |    |
| risk.01     |        |        |    |          |    |    |    |    |    |    |
| base.02     |        |        |    |          |    |    |    |    |    |    |
| circumstance.01|        |        |    |          |    |    |    |    |    |    |

Parsing sentence required 61ms.

**Fig. 6** SRL extraction of an active obligation
The XML output is:

```xml
<Norm>
  <NormType>Obligation</NormType>
  <ActiveRole>Member States themselves</ActiveRole>
  <Action>objectives regarding the management of flood risk should be determined and should be based on local and regional circumstances</Action>
</Norm>
```

If we convert this sentence to an active sentence, the elements extracted are practically the same (figure 6).

Hence, Member States themselves should determine objectives regarding the management of flood risks based on local and regional circumstances.

6.2.2 Right

Here is an example from Directive 98/5/EC:

Any lawyer shall be entitled to pursue on a permanent basis, in any other Member State under his home-country professional title, the activities specified in Article 5.

The path to root is [[u'shall', u'MD', 2], [u'become', u'VB', 3], [u'entitled', u'VBN', 4]]. The word ‘be’ has been transformed in a pre-processing step to ‘become’ to facilitate semantic role labeling. Note that the modal verb ‘shall’ is usually associated with an Obligation norm type. However, there are specific patterns such as ‘shall be entitled’ or ‘shall be granted’ that indicate a Right rather than Obligation. The module looks for such exceptional patterns, based on the corpus used to develop the system.

The relevant SRL roles for the predicate ‘entitled’ are as follows:

```xml
<SRL>
  <PREDICATE>Right</PREDICATE>
  <A2:SBJ>Any lawyer</A2:SBJ>
  <A1:OPRD>
    to pursue on a permanent basis, in any other Member State under his home-country professional title, the activities specified in S_A5
  </A1:OPRD>
</SRL>
```

11 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
<Norm>
  <NormType>Right</NormType>
  <ActiveRole>Any lawyer</ActiveRole>
  <Action>
    to pursue on a permanent basis, in any other Member State under his home-country professional title, the activities specified in Article 5
  </Action>
</Norm>

6.2.3 Permission

Here is an example from Directive 98/5/EC:

A lawyer registered in a host Member State under his home-country professional title may practise as a salaried lawyer in the employ of another lawyer, an association or firm of lawyers, or a public or private enterprise to the extent that the host Member State so permits for lawyers registered under the professional title used in that State.

The modal verb ‘may’ as head of the sentence indicates that the type of norm is a Permission. The verb ‘practice’ is dependent on the head, and used as the ‘predicate’ from which to extract arguments for the ontology:

<SRL>
  <PREDICATE>practice</PREDICATE>
  <A0:SBJ>
    A lawyer registered in a host Member State under his home-country professional title
  </A0:SBJ>
  <A2:ADV>
    practise as a salaried lawyer in the employ of another lawyer, an association or firm of lawyers, or a public or private enterprise, to the extent that the host Member State so permits for lawyers registered under the professional title used in that State
  </A2:ADV>
</SRL>

The keywords ‘to the extent’ in the A2 argument triggers a rule to extract the condition. The rule states that the condition includes the trigger words and all the words that follow up to a comma, semi-colon or full stop.

12 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
<Norm>
  <NormType>Permission</NormType>
  <ActiveRole>
    A lawyer registered in a host Member State
    under his home-country professional title
  </ActiveRole>
  <Action>
    practise as a salaried lawyer in the employ of
    another lawyer, an association or firm of lawyers,
    or a public or private enterprise
  </Action>
  <Condition>
    to the extent that the host Member State so permits
    for lawyers registered under the professional title
    used in that State
  </Condition>
</Norm>

In the example below, on the other hand, arguments from the SRL output
are joined together. Here is the original norm from Directive 98/5/EC:

One or more lawyers who belong to the same grouping in their
home Member State and who practise under their home-country
professional title in a host Member State may pursue their
professional activities in a branch or agency of their
grouping in the host Member State.

The SRL arguments for the predicate ‘pursue’ are as follows:

<SRL>
  <A0:SBJ>
    One or more lawyers who belong to the same grouping in
    their home Member State and who practise under their
    home-country professional title in a host Member State
  </A0:SBJ>
  <A1:OBJ>pursue their professional activities</A1:OBJ>
  <AM-LOC:LOC>
    in a branch or agency of their grouping in the
    host Member State
  </AM-LOC:LOC>
</SRL>

Joining the A1 and AM-LOC arguments together we get the following
structured norm:

13 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998
to facilitate practice of the profession of lawyer on a permanent basis in a Member State
other than that in which the qualification was obtained.
<NormType><Permission/></NormType>

<ActiveRole>
One or more lawyers who belong to the same grouping in their home Member State and who practise under their home-country professional title in a host Member State
</ActiveRole>

<Action>
pursue their professional activities in a branch or agency of their grouping in the host Member State
</Action>

6.2.4 Power

Here is an example from Directive 98/5/EC:

It may require that, when presented by the competent authority of the some Member State, the certificate be not more than three months old.

The path to the tree, ['may', 'MD', 1], ['require', 'VB', 2], is an instance of a rule that identifies a path ['may', MD], ['require', ANYPOS] as a candidate for argument extraction, and identifies this as a pattern of a Power. Note that the modal verb 'may' is not in itself an indication of a Power norm type, since it is also used in Permissions. The module deals with this by identifying 'Power' verbs (such as 'require', 'grant' etc.) that follow the modal, based on the corpus used to develop the system. The indexes on the instance path refer to the position of the word in the sentence. The path in the rule engine usually specifies either a keyword or part-of-speech tag. This allows flexibility in the level of generalisation required to avoid listing all variations while maintaining accuracy.

The relevant SRL output is presented below in an intuitive XML notation for the sake of readability.

<SRL>
  <PREDICATE>require</PREDICATE>
  <A0>It</A0>
  <A1>that, when presented by the competent authority of the home Member State, the certificate be not more than three months old</A1>
</SRL>

The roles of the tree can then be mapped onto slots in a frame-based ontology (such as that of Eunomos and Legal-URN).

In this example, the A0 role becomes the active role of a concept representing a power.

---

14 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
The action element could be further analysed by additional sets of rules, taking advantage of the analysis of the SRL, to understand a condition (AM-TMP):

\[
\text{be when presented by the competent authority of the home Member State not more than three months old}
\]

Note that there is no active role for the obligation. This slot can only be filled by resolving the implicit anaphora. In the sentence, we have a candidate for this slot: the A0 role in the “presented by the competent authority” subordinate sentence. However, domain and contextual understanding is required to verify the anaphora resolution.

Alternatively, a simpler approach, and that which taken in the system developed for this article, is to view the sentence as containing two separate norms, a Power and a (conditional) Obligation.
6.3 Types of Meta-Norms

6.3.1 Legal Effect

Here is an example from Directive 98/5/EC:\footnote{15} Although it is not a prerequisite for the decision of the competent authority in the host Member State, the temporary or permanent withdrawal by the competent authority in the home Member State of the authorisation to practise the profession shall automatically lead to the lawyer concerned being temporarily or permanently prohibited from practising under his home-country professional title in the host Member State.

The roles of the predicate ‘lead’ in the SRL output (shown below in XML) can be conveniently mapped to slots in the ontology.

\[
\text{<SRL>}
\begin{align*}
\text{<PREDICATE>lead</PREDICATE>}
\text{<A0:SBJ>}
\text{the temporary or permanent withdrawal by the competent authority in the home Member State of the authorisation to practise the profession}
\text{</A0:SBJ>}
\text{<A2:ADV>}
\text{to the lawyer concerned being temporarily or permanently prohibited from practising under his home-country professional title in the host Member State}
\text{</A2:ADV>}
\text{<AM-ADV:ADV>}
\text{Although it becomes not a prerequisite for the decision of the competent authority in the host Member State}
\text{</AM-ADV:ADV>}
\text{</SRL>}
\]

After normalisation, the output for the ontology is as follows:

\[
\text{<Norm>}
\begin{align*}
\text{<NormType>Legal Effect</NormType>}
\text{<Situation>}
\text{the temporary or permanent withdrawal by the competent authority in the home Member State of the authorisation to practise the profession}
\text{</Situation>}
\end{align*}
\]

\footnote{15} Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
the lawyer concerned is temporarily or permanently prohibited from practising under his home-country professional title in the host Member State.

it is not a prerequisite for the decision of the competent authority in the host Member State.

Note that the condition in the above example is in truth a non-condition, it is a statement that a possible condition does not in fact apply. The handling of alternative condition-like clauses remains a subject for future work.

6.3.2 Scope

Statements about scope are about the subject or object of a rule. Here is an example from Directive 98/5/EC:

Irrespective of the rules of professional conduct to which he is subjected in his home Member State, a lawyer practising under his home-country professional title shall be subjected to the same rules of professional conduct as lawyers practising under the relevant professional title of the host Member State in respect of all the activities he pursues in its territory.

PathToRoot is [[u'shall', u'MD', 26], [u'become', u'VB', 27], [u'subject', u'VBN', 28]]. The SRL output is:

Irrespective of the rules of professional conduct to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
which he becomes subjected in his home Member State

The output to the ontology is:

<Norm>
  <NormType>Scope</NormType>
  <ActiveRole>
    a lawyer practising under his home-country professional title
  </ActiveRole>
  <Rule>
    to the same rules of professional conduct as lawyers practising under the relevant professional title of the host Member State in respect of all the activities he pursues in its territory
  </Rule>
  <Condition>
    Irrespective of the rules of professional conduct to which he is subjected in his home Member State
  </Condition>
</Norm>

Note again the non-condition in the Condition field.

6.3.3 Exception

Exceptions to norms can take place within the same sentence as a norm or outside (in which case, it usually pertains to the sentence norm before). Exceptions can be represented in different ways: as a separate entity with an ‘Exception’ relation to a norm, or as as an ‘Exception’ field within the norm itself (this is the format used in Eunomos and LEGAL-URN). In the XML output of this research, it is represented as a separate entity. Here is an example of an Exception sentence from Directive 98/5/EC:

Nevertheless, a lawyer practising under his home-country professional title shall become exempted from that requirement if he can prove that he is covered by insurance taken out or a guarantee provided in accordance with the rules of his home Member State, insofar as such insurance or guarantee is equivalent in terms of the conditions and extent of cover.

The arguments from the predicate ‘exempted’ are as follows:

17 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
<SRL>
  <PREDICATE>exempted</PREDICATE>
  <A1:SBJ>
    a lawyer practising under his home-country professional title
  </A1:SBJ>
  <A2:ADV>from that requirement</A2:ADV>
  <AM-ADV:ADV>
    if he can prove that he becomes covered by insurance taken out or a guarantee provided in accordance with the rules of his home Member State, insofar as such insurance or guarantee becomes equivalent in terms of the conditions and extent of cover
  </AM-ADV:ADV>
</SRL>

These are mapped into the following slots in the XML output:

<Norm>
  <NormType>Exception</NormType>
  <WhatIsExcepted>
    a lawyer practising under his home-country professional title
  </WhatIsExcepted>
  <ExceptedFrom>from that requirement</ExceptedFrom>
  <Condition>
    if he can prove that he is covered by insurance taken out or a guarantee provided in accordance with the rules of his home Member State, insofar as such insurance or guarantee is equivalent in terms of the conditions and extent of cover
  </Condition>
</Norm>

6.3.4 Hierarchy of Norms

Finally, the corpus contains several statements expressing the relative hierarchy of one norm with respect to another. We therefore need to have relations between norms that express relative hierarchy. As an intermediate step, we can extract a Hierarchy MetaNormType. This is an example from Directive 98/5/EC:

Integration into the profession of lawyer in the host Member State shall be subject to Article 10.

\[18\] Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
The word ‘subject’ has been transformed by the normalising module into ‘subjected’ to allow the SRL module to extract arguments from the predicate.

\[
\text{<SRL>}
\text{<PREDICATE>subjected<PREDICATE/>}
\text{<A1:SBJ>}
\text{Integration into the profession of lawyer in the host Member State}
\text{</A1:SBJ>}
\text{<A2:ADV>Article 10</A2:ADV>}
\text{</SRL>}
\]

This is then transformed into the following output:

\[
\text{<Norm>}
\text{<NormType>Hierarchy</NormType>}
\text{<LowerPriority>}
\text{Integration into the profession of lawyer in the host Member State}
\text{</LowerPriority>}
\text{<HigherPriority>Article 10</HigherPriority>}
\text{</Norm>}
\]

6.3.5 Rationale

Directives provide the rationale for the existence of the legislation by stating its general purpose and referring to supporting preceding legislation.

Here is an example from Directive 98/5/EC\(^{19}\) of the first kind:

The purpose of this Directive is to facilitate practice of the profession of lawyer on a permanent basis in a self-employed or salaried capacity in a Member State other than that in which the professional qualification is obtained.

Having transformed each ‘is’ and ‘are’ to ‘become’ and ‘becomes’, the path-ToRoot here is [[u‘becomes’, u‘VBZ’, 5]].

The SRL output is:

\[
\text{<SRL>}
\text{<PREDICATE>become<PREDICATE/>}
\text{<A1:SBJ>The purpose of this Directive</A1:SBJ>}
\text{<A2:PRD>}
\text{to facilitate practice of the profession of lawyer on a permanent basis in a self-employed or salaried capacity in a Member State other than that in which the professional qualification was obtained.}
\text{</A2:PRD>}
\text{</SRL>}
\]

\(^{19}\) Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.
The output to the ontology is:

<Norm>
  <NormType>RationalePurpose</NormType>
  <Rule>this Directive</Rule>
  <Purpose>
    to facilitate practice of the profession of lawyer on a
    permanent basis in a self-employed or salaried capacity in
    a Member State other than that in which the professional
    qualification is obtained
  </Purpose>
</Norm>

Other trigger words and phrases that indicate conditions are: ‘so long as’, ‘in case’, ‘on condition’, ‘provided that’ and ‘insofar’, while ‘unless’ and ‘except’ usually indicate an exception.

7 Evaluation

Before evaluating the performance of the system for extracting definitions and norms, it is worth taking into account the performance of the Semantic Role Labeler on legislative text. The system was tested on 224 sentences from Directive 95/46/EC20, 58 definitions and 166 norms. For each sentence, the arguments for all the verb predicates in the sentences were evaluated, and the overall sentence was evaluated as accurate if all the arguments for all the verbs were correct. 78.5% of definitions had correct arguments for all verbs. However, only 52% of norms had the same. Generally, norms are more complex, and therefore more errors are introduced. Nevertheless, not all errors have consequences for the definition and norm extraction system, since only certain predicates are used by the system.

The approach described in Sections 5 and 6 was initially tested on Directive 98/5/EC21, excluding the preamble and annexes. Three sentences were discarded with corrupted output from the normalisation module.

Table 7 shows the results. Here, the strict (S) results take partially correct results as wrong whereas the lenient (L) results take them as being correct (the latter is probably more appropriate since the work in any case should be

---

20 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

21 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained

22 F-measure is calculated using precision and recall decimal values to 17 decimal points
Table 3  Precision, recall and F-measure of definitions and norms: preliminary experiment

| Type         | Number | Precision (Strict) | Recall (Strict) | F-Measure (Strict) | Precision (Lenient) | Recall (Lenient) | F-Measure (Lenient) |
|--------------|--------|--------------------|----------------|--------------------|---------------------|------------------|--------------------|
| Obligation   | 36     | 75.24%             | 77.45%         | 76.33%             | 78.24%             | 77.45%           | 76.33%             |
| Power        | 13     | 88.64%             | 61.90%         | 72.90%             | 95.35%             | 76.92%           | 85.15%             |
| Legal Effect | 10     | 53.33%             | 25.81%         | 34.78%             | 53.33%             | 25.81%           | 34.78%             |
| Definition   | 8      | 100%               | 87.10%         | 93.10%             | 100%               | 87.10%           | 93.10%             |
| Permission   | 6      | 52.94%             | 90%            | 66.67%             | 52.94%             | 90%              | 66.67%             |
| Scope        | 6      | 73.91%             | 77.27%         | 75.55%             | 94.74%             | 90%              | 92.31%             |
| Rationale    | 6      | 100%               | 100%           | 100%               | 100%               | 100%             | 100%               |
| Right        | 2      | 100%               | 100%           | 100%               | 100%               | 100%             | 100%               |
| Exception    | 1      | 100%               | 100%           | 100%               | 100%               | 100%             | 100%               |
| Hierarchy    | 1      | 100%               | 100%           | 100%               | 100%               | 100%             | 100%               |

checked by a knowledge engineer). These tests were conducted on legislation that was also used to develop the system, which account for the exceptionally high results. What this does show is the degree of linguistic consistency across one piece of legislation, which shows the potential for semi-automated definition and norm extraction.

The system was then tested on unseen legislation, Directive 95/46/EC\footnote{Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data} excluding the preamble and annexes. Here, almost all extracted norms were partially correct. To understand this better, table\footnote{F-measure is calculated using precision and recall decimal values to 17 decimal points} and table\footnote{F-measure is calculated using precision and recall decimal values to 17 decimal points} show the quantified results for each norm element. Again, the strict (S) results take partially correct results as wrong whereas the lenient (L) results take them as being correct, bearing in mind that the latter is probably more appropriate since the work in any case should be checked by a knowledge engineer.

The field NormType is relevant to all norms. Possible output were: Definition, Obligation, Permission, Power, Scope, Right, Hierarchy, Exception, Legal Effect and Unknown. All the norms classified as Unknown were in fact, actual norms of the relevant type, apart from one Proclamation which were not sought in the program. The evaluation also revealed a number of norms that should be classed as Liability, which are potential obligations arising from the Power of another to impose an Obligation. This reflects a level of uncertainty about whether such an Obligation will arise. On the other hand, it could be argued that Obligations arising from the Obligation of another to impose an Obligation have a greater level of certainty and should be (and have) been classed as Obligations. Most of the errors in determining the norm type (36%)
Table 4  Accuracy of SRL extraction of elements of definitions and norms

| Element          | Fully Correct | Partially Correct | Wrong (False Positive) | Missing (False Negative) | Misclassified as Different Element (False Negative) |
|------------------|---------------|-------------------|------------------------|--------------------------|-----------------------------------------------|
| Norm Type        | 184           | N/A               | 66                     | 17                       | N/A                                           |
| Definiendum      | 6             | 8                 | 0                      | 2                        | 0                                             |
| Definiens        | 5             | 3                 | 0                      | 1                        | 0                                             |
| Includes         | 5             | 1                 | 0                      | 0                        | 0                                             |
| Excludes         | 0             | 0                 | 0                      | 3                        | 10                                            |
| Action           | 46            | 113               | 56                     | 40                       | 2                                             |
| Active Role      | 113           | 6                 | 10                     | 33                       | 10                                            |
| Passive Role     | 13            | 15                | 5                      | 90                       | 4                                             |
| Condition        | 30            | 41                | 35                     | 38                       | 13                                            |
| Timeframe        | 11            | 11                | 9                      | 6                        | 0                                             |
| Exception        | 4             | 13                | 1                      | 7                        | 7                                             |
| Reason           | 2             | 2                 | 6                      | 9                        | 0                                             |
| Situation        | 0             | 0                 | 5                      | 1                        | 0                                             |
| Result           | 0             | 0                 | 3                      | 1                        | 0                                             |
| Object           | 1             | 0                 | 0                      | 0                        | 0                                             |
| ExcludesObject   | 0             | 3                 | 2                      | 3                        | 0                                             |
| HigherPriority   | 0             | 0                 | 3                      | 0                        | 3                                             |
| LowerPriority    | 0             | 0                 | 0                      | 0                        | 0                                             |

arose from mistaking Powers for Permissions. The problem is that both types of norms have the modal verb ‘may’. The module sought to deal with this by identifying ‘Power’ verbs that follow the modal, based on the corpus used to develop the system. However, the evidence of this evaluation shows that this is less than satisfactory. For EU legislation, it can be assumed that almost all norms involving the modal ‘may’ and having a Member State as an Active Role are Powers. But this is not a general solution to the problem. Also common were the misclassifying of Obligations as Unknown, or the extraction of too many Obligations from the sentence. Some Obligations were misclassified as Legal Effect or Scope.

The other elements in Table 4 and Table 5 are Definiendum, Definiens, Includes and Excludes, which are all elements that pertain to Definitions. The elements Action, Active Role, Passive Role, Condition, Timeframe, Exception and Reason pertain to norms of the type Obligation, Permission, Power and Right. The elements Situation and Result pertain to meta-norms of the type Legal Effect. The elements Object, Excludes Object and Active Role pertain to meta-norms of the type Scope. The elements Higher Priority and Lower Priority pertain to meta-norms of the type Hierarchy. There were few meta-norms in the legislation evaluated.

The results are very varied, and shows that further work is required to achieve acceptable results for a comprehensive norm extraction system. Nevertheless, the most important elements - Norm Type, Active Role - are obtained with good accuracy, which in itself should help compliance officers or
related personnel in their most important search i.e. which obligations need to be complied with.

One significant weakness, however, is the poor performance of the system on identifying Passive Roles. Moreover, apart from their identification, there are two aspects that require further consideration. Firstly, how to distinguish between beneficiaries of norms and other passive roles, such as agents who play an active role in a condition or exception. Secondly, how to relate the passive roles to the relevant parts of the norm. An example of a particularly intricate norm from this point of view is the following:

Any person acting under the authority of the controller or of the processor, including the processor himself, who has access to personal data must not process them except on instructions from the controller, unless he is required to do so by law.

The high Partially Correct results for the Action element reveals that it suffers the most from boundary errors. Boundary errors are also a problem for Conditions, Timeframes, Exceptions and Reasons. This is one particular area where the output of the SRL system is particularly disappointing. However, even when supplemented by pattern-matching, problems remain. 30 Fully Correct Conditions were identified via the SRL output as opposed to 21 via

### Table 5  Precision, recall and F-measure of extraction of elements of definitions and norms

| Element         | Precision (Strict) | Recall (Strict) | F-Measure (Strict) | Precision (Lenient) | Recall (Lenient) | F-Measure (Lenient) |
|-----------------|--------------------|-----------------|--------------------|---------------------|------------------|---------------------|
| Type            | 73.60%             | 91.54%          | 81.66%             | 73.60%              | 91.54%           | 81.60%              |
| Norm Type       | 42.86%             | 75.00%          | 54.55%             | 100%                | 87.5%            | 93.33%              |
| Definiendum      | 62.50%             | 83.33%          | 71.43%             | 100%                | 88.89%           | 94.12%              |
| Definiens       | 83.33%             | 100%            | 90.91%             | 100%                | 100%             | 100%                |
| Excludes        | 0%                 | 0%              | 0%                 | 0%                  | 0%               | 0%                  |
| Action          | 21.40%             | 52.27%          | 30.36%             | 73.95%              | 79.10%           | 76.44%              |
| Active          | 87.60%             | 72.44%          | 79.30%             | 92.25%              | 73.46%           | 81.79%              |
| Related         | 39.39%             | 12.15%          | 18.57%             | 84.85%              | 22.95%           | 36.13%              |
| Passive Role    | 28.30%             | 37.04%          | 32.09%             | 66.98%              | 58.20%           | 62.28%              |
| Condition       | 35.48%             | 64.71%          | 45.82%             | 70.97%              | 78.57%           | 74.58%              |
| Timeframe       | 22.22%             | 22.22%          | 22.22%             | 94.44%              | 54.84%           | 69.39%              |
| Exception       | 20.00%             | 18.18%          | 19.05%             | 40.00%              | 30.77%           | 34.78%              |
| Reason          | 0%                 | 0%              | 0%                 | 0%                  | 0%               | 0%                  |
| Situation       | 0%                 | 0%              | 0%                 | 0%                  | 0%               | 0%                  |
| Result          | 0%                 | 0%              | 0%                 | 0%                  | 0%               | 0%                  |
| Object          | 100%               | 100%            | 100%               | 100%                | 100%             | 100%                |
| Excludes-Object | 0%                 | 0%              | 0%                 | 0%                  | 0%               | 0%                  |
| Higher-Priority | 0%                 | 0%              | 0%                 | 0%                  | 0%               | 0%                  |
| Lower-Priority  | 0%                 | 0%              | 0%                 | 0%                  | 0%               | 0%                  |
pattern-matching (keywords such as ‘where’ or ‘when’). 41 Partially Correct were identified via SRL, 34 via pattern-matching. 35 elements were wrongly classified as Conditions via SRL, 27 via pattern-matching. 38 Conditions were missed altogether and 13 Conditions were wrongly classified as something else. The situation is similar for Timeframes and Exceptions, although there are fewer of those in the legislation evaluated. Some improvement could be made by deeper analysis of dependency trees. However, many of these problems arose due to problems with linking different elements of lists in the normalisation module, and this needs to be looked at further.

The low incidence of Scope and Hierarchy in this particular legislation makes it difficult to provide a proper evaluation of relevant elements.

8 Conclusions and Future Work

This section provides conclusions and contributions of the article, as well as possibilities for future work.

This article endorses the vision of Sartor (2011), page 7, of a future legal semantic web where legal content on the web is enriched with machine processable information that can be “presented in many different ways, according to the different issues at stake and the different roles played by its users” by focussing on extracting and linking structured legal information for compliance purposes, building on research in legal theory, ontologies and natural language processing. In addition to a thorough description of methodology and results, the article highlights important challenges that have emerged from this research which will hopefully prove useful for future work in the extraction and linking of information from legislative text.

To address the resource bottleneck problem of creating specialist knowledge management systems, in particular how to semi-automate the extraction of definitions, norms and their elements to populate legal ontologies. Sections 5 and 6 described a definition and norm extraction system using semantic representations from a general-purpose SRL module (Björkelund et al. (2009)), as well as dependency parse trees from the same system. This solution was pursued in order to simplify the sets of rules required to identify possible norms and definitions, classify their types, and map arguments to domain-specific slots in a legal ontology.

There are a number of observations that have emerged from this evidence-based research. While much theoretical work on norms have focused on obligations, and there are indeed plenty of them in the legislation studied, there are also certain kinds of norms that are less prominent in the literature but are, nevertheless, important to cover in a comprehensive norm extraction system. EU legislation in particular are full of norms bestowing a Power to impose
certain obligations (as well as obligations to impose obligations). It is a moot point whether the secondary obligations then should be considered liabilities or obligations, particularly since there is a greater element of uncertainty where the secondary obligations derive from a Power rather than an Obligation. Legislation also contain a number of meta-norms indicating scope, hierarchy among norms and exceptions, and these are arguably essential information to extract for an ontology of legal norms. Statements indicating the legal effect of a particular situation may not be norms in the strictest sense, but have important implications and therefore should also be included. As for elements of norms, many statements refer to the scope of the norm. In addition to the Condition element used in Eunomos and Legal-URN, it is submitted that a Timeframe element can also be of value. One element that requires further investigation is Passive Role, since this research found that there are many different kinds of passive roles, not only beneficiaries, and it would be useful to distinguish between them. Another element that requires further consideration is the Condition element. There are a number of constructions in legislative text regarding the applicability of norms that are somewhat similar to conditions, but have different effects. For instance, ‘in particular’ indicates that a norm applies in a particular scenario, but is not limited to that scenario. ‘Notwithstanding’ implies that what might be considered as a negative condition does not in fact apply.

Other future work that follows from this research are:

- greater study on distinguishing norm types that often use the same modal verbs e.g. Powers and Permissions, Rights and Obligations. This may involve a probabilistic approach, and consideration of additional factors beyond predicate verbs.
- intra- and extra-sentential anaphora resolution specifically for legislative text. There are particular challenges and opportunities here. Anaphora can take the form of pronouns, abbreviated forms, or references to entities in other articles. A related issue is how to determine ‘missing’ elements of a norm e.g. for normative sentences that lack an Active Role, it is often the case that the role is implicitly carried over from the previous sentence.
- detection of Is-A and Part-Of relationships among defined entities, and in particular, detection of Example type Definitions found in sentences containing the construction ‘such as’.
- generation of different views of a norm based on Hohfeldian correlatives, such as the Obligations that can be implicitly derived from Rights, since legislative text will usually provide only one view explicitly. This can be useful for a compliance-oriented system such as Legal-URN. Indeed, recent developments in research on Legal-URN ([Rabinia et al. 2020]) transforms all legal statements into Obligations, since this is the point of view most useful to the intended user.
- rigorous analysis of the relationship between norms to determine norms that may be satisfied in alternative ways, or must be satisfied as a group,
taking into account the impact of negatives on conjunctions, as highlighted by Buabuchachart et al. [Buabuchachart et al. (2013)].

- the development of SRL-based information extraction modules for other jurisdictions and other languages;

References

Ajani G, Ebers M (2005) Uniform terminology for European contract law, vol 1. Nomos Verlag
Ajani G, Boella G, Di Caro L, Robaldo L, Humphreys L, Praduroux S, Rossi P, Violato A (2016) The european legal taxonomy syllabus: a multi-lingual, multi-level ontology framework to untangle the web of european legal terminology. Applied Ontology 11(4):325–375
Bartolini R, Lenci A, Montemagni S, Pirrelli V, Soria C (2004) Automatic classification and analysis of provisions in italian legal texts: a case study. In: OTM Confederated International Conferences’ On the Move to Meaningful Internet Systems’, Springer, pp 593–604
Berland M, Charniak E (1999) Finding parts in very large corpora. In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, Association for Computational Linguistics, pp 57–64
Bertoldi A, de Oliveira Chishman RL (2011) The limits of using frameNet frames to build a legal ontology. In: ONTOBRAS-MOST, Citeseer, pp 207–212
Biagioli C, Francesconi E, Passerini A, Montemagni S, Soria C (2005) Automatic semantics extraction in law documents. In: Proceedings of the 10th International Conference on Artificial Intelligence and Law, ACM, New York, NY, USA, ICAIL ’05
Biemann C (2005) Ontology learning from text: A survey of methods. In: LDV forum, vol 20, pp 75–93
Bird S (2006) Nltk: the natural language toolkit. In: Proceedings of the COLING/ACL on Interactive presentation sessions, Association for Computational Linguistics, pp 69–72
Björkelund A, Hafidell L, Nugues P (2009) Multilingual semantic role labeling. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, Association for Computational Linguistics, pp 43–48
Bobbitt P (2006) Constitutional Interpretation. Oxford University Press., URL http://books.google.lu/books?id=UDZqAAAAAACAAJ
Boella G, Humphreys L, Martin M, Rossi P, van der Torre L (2012) Eunomos, a legal document and knowledge management system to build legal services. In: AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents, Springer, pp 131–146
Boella G, Di Caro L, Rispoli D, Robaldo L (2013a) A system for classifying multi-label text into eurovoc. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, pp 239–240
Boella G, Di Caro L, Robaldo L (2013b) Semantic relation extraction from legislative text using generalized syntactic dependencies and support vector machines. In: International Workshop on Rules and Rule Markup Languages for the Semantic Web, Springer, pp 218–225
Boella G, Tosatto SC, Ghanavati S, Hulstijn J, Humphreys L, Muthuri R, Rifaut A, van der Torre L (2014) Linking legal open data: breaking the accessibility and language barrier in european legislation and case law. In: Proceedings of the 15th International Conference on Artificial Intelligence and Law, ACM, pp 171–175
Boella G, Di Caro L, Humphreys L, Robaldo L, Rossi P, van der Torre L (2016) Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. Artificial Intelligence and Law 24(3):245–283

Brill E (1992) A simple rule-based part of speech tagger. In: Proceedings of the workshop on Speech and Natural Language, Association for Computational Linguistics, pp 112–116

Buabuchachart A, Charness N, Metcalf K, Morgenstern L (2013) Automated methods for extracting and expanding lists in regulatory text. In: Proceedings of the First JURIX Doctoral Consortium and Poster Sessions in conjunction with the 26th International Conference on Legal Knowledge and Information Systems, JURIX 2013, Bologna, Italy, December 11-13, 2013.

Cai Z, Yu E (2002) Addressing performance requirements using a goal and scenario-oriented approach. In: International Conference on Advanced Information Systems Engineering, Springer, pp 706–710

Ciaramita M, Attardi G, Dell’Orletta F, Surdeanu M (2008) Desrl: A linear-time semantic role labeling system. In: Proceedings of the Twelfth Conference on Computational Natural Language Learning, Association for Computational Linguistics, pp 258–262

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

Cunningham H, Maynard D, Bontcheva K, Tablan V (2002) Gate: A framework and graphical development environment for robust nlp tools and applications. In: Proc. 40th Anniversary Meeting of the Association for Computational Linguistics (ACL)

Curran JR, Clark S, Bos J (2007) Linguistically motivated large-scale NLP with c&c and boxer. In: ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic

Fillmore CJ (1968) The case for case. Holt, Rinehart, and Winston

Fillmore CJ (1976) Frame semantics and the nature of language. Annals of the New York Academy of Sciences 280(1):20–32

Fillmore CJ, Ruppenhofer J, Baker CP (2004) Framenet and representing the link between semantic and syntactic relations. Frontiers in linguistics 1:19–59

Gao X, Singh MP (2014) Extracting normative relationships from business contracts. In: International conference on Autonomous Agents and Multi-Agent Systems, AAMAS ’14, Paris, France, May 5-9, 2014, pp 101–108

Ghanavati S (2013) Legal-URN framework for legal compliance of business processes. http://hdl.handle.net/10393/24028, ph.D. thesis, UOttawa, Canada

Giuliano C, Lavelli A, Romano L (2006) Exploiting shallow linguistic information for relation extraction from biomedical literature. In: EACL, Citeseer, vol 18, pp 401–408

Gliozzo AM, Giuliano C, Rinaldi R (2005) Instance pruning by filtering uninformative words: an information extraction case study. In: Computational Linguistics and Intelligent Text Processing, Springer, pp 498–509

Governatori G, Olivieri F, Rotolo A, Scannapieco S (2013) Computing strong and weak permissions in defeasible logic. Journal of Philosophical Logic 42(6):799–829

Grabmair M, Ashley KD, Hwa R, Sweeney PM (2011) Toward extracting information from public health statutes using text classification machine learning. In: Legal Knowledge and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference, University of Vienna, Austria, 14th-16th December 2011, pp 73–82

Hajík J, Ciaramita M, Johannson R, Kawahara D, Martí MA, Márquez L, Meyers A, Nivre J, Padó S, Štěpánek J, et al. (2009) The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, Association for Computational Linguistics, pp 1–18

Harris Z (1968) Mathematical structures of language.

Hearst MA (1992) Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th conference on Computational linguistics-Volume 2, Association for Computational Linguistics, pp 539–545

Hepp M (2007) Ontologies: State of the art, business potential, and grand challenges. In: Hepp M, Leenheer PD, de Moor A, Sure Y (eds) Ontology Management: Semantic Web, Semantic Web Services, and Business Applications, Springer, pp 3–22
Hirst G (1992) Semantic interpretation and the resolution of ambiguity. Cambridge University Press

Hudson R (1984) A (1984) word grammar. Basil Blackwell, Oxford

ITU-T (2012) Recommendation Z.151 (10/12). User Requirements Notation (URN) - Language definition. http://www.itu.int/rec/T-REC-Z.151/en

Johansson R, Nugues P (2008a) Dependency-based semantic role labeling of propbank. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, pp 69–78

Johansson R, Nugues P (2008b) The effect of syntactic representation on semantic role labeling. In: Proceedings of the 22nd International Conference on Computational Linguistics—Volume 1, Association for Computational Linguistics, pp 393–400

Kaiser M, Webber B (2007) Question answering based on semantic roles. In: Proceedings of the Workshop on Deep Linguistic Processing, Association for Computational Linguistics, pp 41–48

Kingsbury P, Palmer M (2002) From treebank to propbank. In: LREC, Citeseer

Klein M (2001) Combining and relating ontologies: an analysis of problems and solutions. In: Workshop on Ontologies and Information Sharing, IJCAI'01, Seattle, USA

Lenci A, Montemagni S, Pirrelli V, Venturi G (2009) Ontology learning from Italian legal texts. In: Proceedings of the 2009 Conference on Law, Ontologies and the Semantic Web: Channelling the Legal Information FLOOD, IOS Press, Amsterdam, The Netherlands, The Netherlands, pp 75–94

Levin B (1993) English verb classes and alternations: A preliminary investigation. University of Chicago Press

Loos EE, Anderson S, Day D, Jordan P, Wingate J (2004) Glossary of linguistic terms, vol 29. SIL International

Lyding V, Chiocchetti E, Séraisset G, Brunet-Manquat F (2006) The lexalp information system: Term bank and corpus for multilingual legal terminology consolidated de Maat E, Winkels R, van Engers T (2009) Making sense of legal texts. Formal Linguistics and Law 212:225

Malerba A (2017) Interpretive interactions among legal systems and argumentation schemes. PhD thesis, alma

Marcus M, Kim G, Marcinkiewicz MA, MacIntyre R, Bies A, Ferguson M, Katz K, Schasberger B (1994) The penn treebank: annotating predicate argument structure. In: Proceedings of the workshop on Human Language Technology, Association for Computational Linguistics, pp 114–119

Márquez L (2009) Semantic role labeling: past, present and future. In: Tutorial Abstracts of ACL-IJCNLP 2009, Association for Computational Linguistics, pp 3–3

Márquez L, Carreras X, Litkowski KC, Stevenson S (2008) Semantic role labeling: an introduction to the special issue. Computational linguistics 34(2):145–159

Martin JH, Jurafsky D (2016) Semantic role labeling: draft of june 26, 2015. In: Speech and language processing

Maxwell KT, Oberlander J, Lavrenko V (2009) Evaluation of semantic events for legal case retrieval. Proceedings of the WSDM 9:09–09

Mel’cuk IA, et al. (1988) Dependency syntax: theory and practice. SUNY press

Mihalcea R, Moldovan D (2000) Semantic indexing using wordnet senses. In: Proceedings of the ACL-2000 workshop on Recent advances in natural language processing and information retrieval: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics—Volume 11, Association for Computational Linguistics, pp 35–45

Miller GA, Beckwith R, Fellbaum C, Gross D, Miller KJ (1990) Introduction to wordnet: An on-line lexical database. International journal of lexicography 3(4):235–244

Moldovan DI, Rus V (2001) Logic form transformation of wordnet and its applicability to question answering. In: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, pp 402–409

Palmirani M, Ceci M, Radicioni D, Mazzei A (2011) Framenet model of the suspension of norms. In: Proceedings of the 13th International Conference on Artificial Intelligence and Law, ACM, pp 189–193
Yangarber R, Grishman R, Taipanainen P, Huttunen S (2000) Automatic acquisition of domain knowledge for information extraction. In: Proceedings of the 18th conference on Computational linguistics-Volume 2, Association for Computational Linguistics, pp 940–946