Use of OWL in the Legal Domain
Statement of Interest

Rinke Hoekstra

Leibniz Center for Law, University of Amsterdam
PO Box 1030, 1000 BA, Amsterdam, The Netherlands
hoekstra@uva.nl

Abstract. The Leibniz Center for Law at the University of Amsterdam is involved in several research projects that deal with the integration of knowledge representation with legal texts. For many of these projects, the use of ontologies in OWL DL plays an important, if not a central role. In this statement of interest we give a short sketch of the kind of application we see for OWL in the legal domain, and discuss several state-of-the-art developments by leaders in the OWL DL field, that are of primary interest to us.

1 Introduction

The Leibniz Center for Law[1] at the University of Amsterdam is involved in several research projects that deal with the integration of knowledge representation with legal texts, both in the form of semantic annotations and for the development of full-blown knowledge based systems. For many of these projects, the use of ontologies in OWL DL plays an important, if not a central role. In this statement of interest we give a short sketch of the kind of application we see for OWL in the legal domain, and discuss several state-of-the-art developments by pioneers of uncharted territories in the OWL DL field, that are of primary interest to us.

2 Text and Representation

In many ways, the corpus of legal information available today is the world wide web's little sister, at least qua structure. It consists of a huge volume of heterogeneous, but closely inter-linked documents. These documents are increasingly being made available in digital form as part of public accessibility projects by governments.[2] However, a major difference is that the relations between legal

[1] The Leibniz Center for Law is part of the Faculty of Law of the University of Amsterdam and participates in the OWL Working Group. See http://www.leibnizcenter.org.

[2] An example is the portal of the Dutch government http://www.wetten.nl that discloses currently active legislation

texts are typically expressed in natural language only. Also, these references are not always absolute, typically point to *parts* of documents, and often import an externally defined meaning of a term [7, 8]. Consolidation of such semantic references into a single representation introduces a significant maintenance issue, as legal texts are very dynamic and undergo change independently from each other. In fact, the meaning of terms in law imposes an ordering on entities in reality that can change over time, but *stays applicable* to older cases. In short, law adopts an intricate versioning scheme [2, 4]. We are involved in the MetaLex/CEN[3] XML standard for legislative sources [5, 1] that provides an XML schema for representing the structure and dynamics of legal texts.

It is our firm conviction that a semantic representation – be it for the purpose lightweight annotation, consistency checking or legal knowledge based reasoning (planning, assessment) – should take the dynamic and structural properties of legal texts into account. This is most directly reflected in the principle of *traceability*: any representation of some legal text should be traceable to its original source. A representation of some (part of) legislation is dependent on that legislation, and is therefore essentially always an *annotation* on that text. The MetaLex/CEN initiative provides a standard transformation of XML encoded legal texts to RDF/XML. More elaborate, formal representations of the contents of the texts in OWL are then related to this RDF representation. An exciting application area for this methodology is that of *compliance*, where the business processes of organisations (businesses and governments alike) need to be aligned with respect to some body of regulations.[4]

We are currently investigating means to represent definitions of concepts in OWL in such a way that their semantic interpretation mimics the structure and applicability of the texts. This includes means to *scope* definitions with respect to particular *parts of a text*, as in e.g. deeming provisions, regarding the *temporal validity* of a text [14], and concerning its *jurisdiction*.

One could argue that such requirements indicate the need for knowledge representation languages specific to law (as in e.g. deontic logics). However, the legal field is in one respect wholly analogous to the web in that legal information is used and incorporated in a wide variety of systems, each using the information in different ways. Also, the whole body of legal information is not maintained by a single issuer, but rather by a significant number of authorities that each publish, incorporate, extend, comply with, enforce and implement regulations. Therefore, the requirements for knowledge representation on the Semantic Web hold for representation of legal texts as well. Especially as the information exchange between those parties can benefit enormously from a well designed standard.

We are currently involved in an effort to develop a legal knowledge interchange language (LKIF) that allows for the interchange of legal knowledge be-

[3] CEN is the European Committee for Standardization, See `http://www.metalex.eu`, `http://legacy.metalex.eu` and `http://www.cen.eu`

[4] This is the subject of the recently granted AGILE project.

tween commercial vendors [3, 5].[5] LKIF is an example of a hybrid approach in that it combines OWL DL (for the representation of concepts) with a rule formalism.

3 Representation and Reasoning

The LKIF includes a core ontology (LKIF Core) that provides a vocabulary and a set of standard definitions of concepts common to all legal fields based on commonsense [6, 10].[6] In the development of this ontology, we frequently encountered the limitations of OWL DL, both with respect to reasoner performance and expressiveness. Especially the extensive use of a combination of Generic Concept Inclusion axioms (equivalent class statements on existential restrictions) and inverse properties turned out to be quite taxing for (in our case) the Pellet reasoner. As at this time we used the reasoner primarily for debugging purposes, a single inconsistency in the TBox could cause the reasoner to stall, making it hard to debug the ontology. Although this problem was remedied by lifting some of the restrictions on classes, it indicated that real time performance of reasoners on ontologies that use the full expressiveness of $\mathcal{SHOIN}(D)$ can still be improved.

Structured Objects On the other hand, for many common concepts in law OWL DL is still too limited in its expressiveness. To be sure, some of these restrictions are alleviated by the OWL 1.1 proposal – in particular, we applaud the role inclusions and qualified cardinality restrictions of $\mathcal{SROIQ}$ [12]– but for several common patterns we currently need to resort to rule-based solutions. This is not desirable for many reasons. To give an example, law mainly governs the actions of people, and is especially detailed where they interact, as in transactions. Transactions can be conceptualised in a straightforward manner as two interdependent actions. For instance, a sales transaction contains of the two actions of buying and selling, each of which involves its own actor, recipient and object. This pattern is analogous to that of *structured objects*, as described in [16], and is essentially diamond shaped, where OWL DL only handles tree (or forest) like structures. A similar problem occurs when expressing the complementarity of rights and duties.

Arguably, several of such patterns could be expressed using DL safe rules, but this solution is in many cases not satisfactory as it requires us to represent additional information in rules that *can* be represented using description logics (cf. [11]). Also, rules only take into account the individuals that were explicitly asserted in an ABox, and these as such do not necessarily express a valid model of the theory expressed by the TBox. Furthermore it seems more cognitively intuitive to indicate corresponding identity between relata using pronouns than by

[5] LKIF is developed as part of the ESTRELLA project, see `http://www.estrellaproject.org`

[6] LKIF Core currently contains about 205 named class definitions defined using 114 properties. See `http://www.estrellaproject.org/lkif-core`

means of variables or property reflexivity. We are currently investigating means to approximate such structures using role inclusion axioms, and are very much looking forward to progress in the direction of *description graphs* to describe structured objects [16].

Hybrid Approaches Notwithstanding our preference for DL-based concept definitions, we believe that any useful application of OWL DL in a knowledge based system will inevitably require interplay with different formalisms. For obvious reasons, a combination with rule-based solutions is the most likely, not only because most legacy systems (as e.g. developed by commercial vendors) are based on this paradigm. We feel that the current discussion on rule-like fragements of OWL (in the OWL WG), such as Oracle's OWL-Prime and OWL DLP (cf. [17, 9])[7] is therefore very important. Progress in the specification of combinations between DL and rule-based approaches is closely watched by us and we are hopeful that in the end, RIF and OWL will get along.

Conditional Classification A relatively uncharted territory in the field of legal knowledge representation is the combination of regulations with a geospatial jurisdiction, as in spatial planning. In recent projects we have experimented with semantic annotation of maps in combination with legal texts in MetaLex and the Dutch IMRO standard vocabulary for zoning plans [18, 19]. Spatial plans essentially enforce a particular type of *use* in some area expressed as designations, e.g. 'housing', 'water', 'greenery' etc. However, regulations may be in place that further refine those designations with additional restrictions. We envisage applications where users can describe their intended use, run it as query on a suitably represented body of regulations, and have the areas available for that use depicted on a map.

However, usage designations are not only enforced exclusively. Because of intricate interplay between regulations of different authorities (EU directives, national legislation, provincial and local directives) land use is open to *compensation*. For instance, a particular lot may have both the designation 'greenery' and 'housing' but each only to a varying degree. This means that someone applying for a permit to build a house on that particular lot is bound to some measure of compensation if the 'housing'-use of that lot exceeds the designated maximum. A permit will only be issued if the damage to existing greenery is compensated in a different area.

We hope that recent developments with respect to probabilistic extensions to OWL DL along the lines of [15], and currently implemented in Pronto,[8] can be usefully applied to indicate necessity of possibility for compensation of land use. We furthermore feel that the way in which annotations are used to incorporate a non-intrusive extension of the OWL DL semantics is a very sensible approach that deserves further thought.

[7] Recently named OWL-R Full and OWL-R DL respectively, see `http://www.w3.org/2007/OWL/wiki/Fragments_Proposal`

[8] See `http://pellet.owldl.com/pronto`

Explanation Reasoning in law is all about *justification*: the rational reconstruction of a case is often the most convincing argument.[9] For this reason, a legal knowledge based system needs to be equipped with elaborate explanation facilities. The current state-of-the-art in explanation and justification of DL entailments as e.g. supported by Pellet ([13]) is therefore a very welcome addition to standard OWL DL reasoning services.

4 Conclusion

As we discuss above, in our view OWL plays (or at least, should play) a central role in knowledge representation in the legal domain. We feel that law is an excellent example of a domain where the combination of semantic web technology and traditional knowledge representation can make a difference.

In particular, we hope to see progress in the areas dealing with:

Expressivity especially with respect to 'diamond shaped' class descriptions.
Performance both on ABox and TBox reasoning with highly expressive ontologies.[10]
Explanation of DL entailments for the purpose of justification and traceability.
Annotation with respect to a transparent connection between the axioms in an OWL ontology, and structural elements represented as RDF.
Extensions possibility to extend the OWL DL semantics (as used by Pronto) using a standard extension mechanism.
Versioning of both ontologies and concepts in the ontology.
Interaction with Rules for the purpose of building hybrid knowledge based systems.

[9] Of course this is not always the case, and legal argumentation is often performed in such a way as to make it *appear* rational.
[10] Wouldn't we all...

Bibliography

[1] A. Boer, R. Hoekstra, R. Winkels, T. van Engers, and F. Willaert. ^{META}lex: Legislation in XML. In T. Bench-Capon, Aspassia Daskalopulu, and R.G.F. Winkels, editors, *Legal Knowledge and Information Systems. Jurix 2002: The th Annual Conference*, pages 1–10, Amsterdam, 2002. IOS Press.

[2] A. Boer, R. Winkels, T. van Engers, and E. de Maat. A content management system based on an event-based model of version management information in legislation. In T. Gordon, editor, *Legal Knowledge and Information Systems. Jurix 2004: The 17th Annual Conference.*, pages 19–28, Amsterdam, 2004. IOS Press.

[3] Alexander Boer, Thomas F. Gordon, Kasper van den Berg, Marcello Di Bello, András Förhécz, and Réka Vas. Specification of the legal knowledge interchange format. Deliverable 1.1, Estrella, 2007.

[4] Alexander Boer, Radboud Winkels, Tom van Engers, and Emile de Maat. Time and versions in ^{META}lex XML. In *Proceeding of the Workshop on Legislative XML*, Kobaek Strand, 2004.

[5] Alexander Boer, Radboud Winkels, and Fabio Vitali. Proposed XML standards for law: Metalex and LKIF. In Arno R. Lodder and Laurens Mommers, editors, *Legal Knowledge and Information Systems. Jurix 2007: The Twentieth Annual Conference Annual Conference*, volume 165 of *Frontiers in Artificial Intelligence and Applications*, pages 19–28. IOS Press, December 2007.

[6] Joost Breuker, Rinke Hoekstra, Alexander Boer, Kasper van den Berg, Rossella Rubino, Giovanni Sartor, Monica Palmirani, Adam Wyner, and Trevor Bench-Capon. OWL ontology of basic legal concepts (LKIF-Core). Deliverable 1.4, Estrella, 2007.

[7] E. de Maat, R. Winkels, and T. van Engers. Making Sense of Legal Texts. In G. Grewendorf and M. Rathert, editors, *Formal Linguistics and Law*, Trends in Linguistics - Studies and Monographs (TiLSM). Mouton, De Gruyter, Berlin, (in press) 2008.

[8] Emile de Maat, Radboud Winkels, and Tom van Engers. Automated detection of reference structures in law. In Tom M. van Engers, editor, *Legal Knowledge and Information Systems. Jurix 2006: The Nineteenth Annual Conference*, volume 152 of *Frontiers in Artificial Intelligence and Applications*, pages 41–50. IOS Press, December 2006.

[9] Benjamin Grosof, Raphael Volz, Ian Horrocks, and Stefan Decker. Description logic programs: Combining logic programs with description logics. In *In Proceedings of the 12th International World Wide Web Conference (WWW 2003)*, 2003.

[10] R. Hoekstra, J. Breuker, M. Di Bello, and A. Boer. The LKIF Core ontology of basic legal concepts. In Pompeu Casanovas, Maria Angela Biasiotti, Enrico Francesconi, and Maria Teresa Sagri, editors, *Proceedings*

of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007), June 2007.

[11] R. Hoekstra, J. Liem, B. Bredeweg, and J. Breuker. Requirements for representing situations. In Bernardo Cuenca Grau, Pascal Hitzler, Conor Shankey, and Evan Wallace, editors, *Proceedings of the OWLED'06 workshop on OWL: Experiences and Directions 2006*, volume 216 of *CEUR Workshop Proceedings*, Athens, Georgia (USA), November 10-11 2006.

[12] Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The even more irresistible SROIQ. In *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning*, pages 57–67, 2006.

[13] Aditya Kalyanpur, Bijan Parsia, Matthew Horridge, and Evren Sirin. Finding all justifications of OWL DL entailments. In Karl Aberer, Key-Sun Choi, Natasha Fridman Noy, Dean Allemang, Kyung-Il Lee, Lyndon J. B. Nixon, Jennifer Golbeck, Peter Mika, Diana Maynard, Riichiro Mizoguchi, Guus Schreiber, and Philippe Cudré-Mauroux, editors, *ISWC/ASWC*, volume 4825 of *Lecture Notes in Computer Science*, pages 267–280. Springer, 2007.

[14] Szymon Klarman, Rinke Hoekstra, and Marc Bron. Versions and applicability of concept definitions in legal ontologies. In Kendall Clark and Peter F. Patel-Schneider, editors, *Proceedings of OWL: Experiences and Directions (OWLED 2008 DC)*, Washington, DC (metro), April 2008.

[15] Thomas Lukasiewicz. Probabilistic description logics for the semantic web. INFSYS Research Report 1843-06-05, INFSYS, Technische Universität Wien, 2007.

[16] Boris Motik, Bernardo Cuenca Grau, and Ulrike Sattler. Structured objects in OWL: Representation and reasoning. Technical report, University of Oxford, UK, 2007.

[17] Boris Motik, Ian Horrocks, Riccardo Rosati, and Ulrike Sattler. Can OWL and logic programming live together happily ever after. In *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, number 4273 in LNCS, Athens, GA, USA, November 5-9 2006.

[18] R. Peters and T.M. van Engers. The legal atlas: Map-based navigation and accessibility of legal knowledge sources. In *Knowledge Management in Electronic Government; 5th IFIP International Working Conference, KMGov 2004, Krems Austria*, pages 212–220. Springer Verlag, 2004.

[19] Radboud Winkels, Alexander Boer, and Erik Hupkes. Legal Atlas: Access to legal sources through maps. In Radboud Winkels, editor, *Proceedings on the 11th International Conference on Artificial Intelligence and Law (ICAIL 2007)*. IAAIL, ACM, June 2007.