An Ontological Representation of EU Consular Law

Erich Schweighofer
University of Vienna, Centre for Legal Informatics

Abstract. At present, EU consular law is under legal scrutiny by the European Commission. The CARE study reveals good pragmatic application but also significant implementation problems. As a site effect of our analysis, we have developed a concept of a legal ontology for knowledge description, multilingual information retrieval and semi-automatic application of consular law using a dialogue system. First experiments show the potential of this approach.
Keywords: Consular assistance, diplomatic protection, EU law, legal ontologies, dynamic electronic legal commentary, dialogue systems, multilingual information retrieval

1. Challenges of EU Consular Law

Article 23 of the Treaty on the Functioning of the European Union (TFEU) gives every citizen of the Union the right to consular and diplomatic protection if his or her Member State is not represented in a specific third country. Whichever mission (of another EU member state) the EU citizen ends up asking for support, the mission has to provide support on the same conditions as for their own nationals.

Article 46 of the Charta on Human Rights lays down the same right. The Green Paper "Diplomatic and consular protection of Union citizens in third countries", presented by the Commission in 2006, focuses on strengthening this right: In it, the European Commission points out that European citizens are not fully aware of this right, and that the legal consequences of it are far from being fully implemented by the Member States. After the consultation phase of the Green Paper, the Action Plan 2007-2009 "Effective consular protection in third countries: the contribution of the European Union" was adopted. One important measure is the examination of Member States' legislations and practices on consular protection and the assessment of the extent and nature of the observed discrepancies between Member States.

The CARE (Citizens Consular Assistance Regulation in Europe) project (http://www.careproject.eu) aims at offering tools to the Commission which support the European Commission in performing this examination. The CARE database collects relevant legal materials on diplomatic and consular protection adopted in each EU Member State. Various types of documents are collected: legislation, case law, administrative directives and guidelines, and also other informative materials made available by national governments for their citizens. The database contains full text documents in their original language, enriched by a metadata set, i.e. information about

the documents. Metadata are translated into English and French. Texts of the most relevant documents are translated into English and French as well. The database is accessible by all European citizens via the Internet (http://www.careproject.eu/database). A comprehensive report analyzes the legal framework in the EU Member States based on assessments of 27 national correspondents.

From a legal point of view, significant insufficiencies of implementation of Article 23 TFEU exist, in particular concerning legal frame work, standards of legal rules, reimbursement etc. These problems are solved in practice with a pragmatic implementation.

An ontological analysis shows that conceptualisation of consular law remains sketchy. Neither International treaties nor national laws have developed a strong terminology on consular law. Even a lexical ontology may provide important assistance.

Further, an ontology can be considered as an approach for solving the problem of multilingual (e.g. in 23 Community languages) handling of consular cases (see for the long list of functions Article 5 of the Vienna Convention on Consular Relations), taking into account the 27 different consular protection laws and policies. The ontology can provide required equivalence of concepts but can be linked also to a dialogue system.

For these reasons, experimental research on legal ontologies and dialogue systems has been undertaken. The remainder of this paper is organized as follows: Section 2 describes the consular law legal information system, section 3 the ontology of EU consular law, section 4 the dynamic legal electronic commentary, section 5 first experiments and, last but not least, in section 6, tentative conclusions are presented.

2. Legal Information System CONSUL

Handbooks in paper have long ceased to constitute best practice for dissemination of information. Websites and information systems are able to very nicely present the complex knowledge while coping very efficiently with often daily updates (e.g. travel recommendations). For finding materials, legal search constitutes an indispensable tool. Legal retrieval remains the best solution for determining the similarity between documents and queries (Manning et al 2008, Turtle 1995, Schweighofer 1999). For smaller domains like consular law with a complex structure, hypertext systems are a powerful tool. The flexible way of access with a non-linear representation of knowledge allows a user-friendly access to this body of knowledge.

The existing CARE database already allows full text information retrieval and browsing in the document collection. Our more powerful document retrieval system is going to be built using Apache Lucene (Apache Lucene

2010, Gospodnetic & Hatcher 2006), which offers state-of-the-art text retrieval capabilities but also allows fine-tuning of the information retrieval system according to the document properties of our text collection. Apache Lucene fulfils also the requirement of easy maintenance of the text corpus but also an efficient handling of the various versions.

3. Ontology of EU Consular Law

Since the 1990ies, ontologies as a conceptualisation of a domain are considered as tool for organising legal knowledge. Later, the idea of a semantic web (Berners-Lee 2001) with a mark-up that makes the text intelligent and active energized the concept of legal ontologies. For a long time, the University of Amsterdam has set the standards of legal ontologies with LRI-Core and now LKIF (Hoekstra, Breuker, De Bello/Boer 2007). Legal ontologies were implemented for tasks of conceptual information retrieval, knowledge representation, multilingual information retrieval or exchange of information and knowledge (see (Casanovas et al. 2007) and (Casellas et al. 2009)).

In our case, we consider using two ontologies: a lexical ontology like in the LOIS project (Dini et al. 2005) and a much more developed Dynamic Electronic Legal Commentary Ontology (Schweighofer 2006, Schweighofer 2010a) (see below).

A thesaurus for indexing contains a list of every important term in a given domain of knowledge and a set of related terms for each of these terms. A lexical ontology builds up from this basis with works on glossaries and dictionaries, extends the relations and makes this knowledge computer-usable in order to allow intelligent applications. Lexical ontologies provide this formalized description of a domain that can be understood and re-used by a knowledge system.

Based on already existing indices and sketchy conceptual structures, a lexical ontology CONSUL with about 200 legal and factual descriptors with definitions and relations has been established. Content relations will be taken from standard WordNet relations (especially hyperonymy and hyponymy). For all concepts, an ILI will be created in order to support multilingual use but also multilingual retrieval. Methodology is mostly derived from the previous LOIS project.

4. Dynamic Electronic Legal Commentary (DynELCom) CONSUL

The Dynamic Legal Electronic Commentary (DynELCom) (Schweighofer 2006, Schweighofer 2010a) CONSUL consists of a textual, e.g. syntactic representation of consular law that is supplemented by a semantic representation of the legal rules (e.g. conceptual representation of rules), a

semantic representation of the world (e.g. conceptual representation of facts) and a legal link structure between these repositories of knowledge. Knowledge acquisition is supported by semi-automatic text summarisation and text classification. A sketchy inference machine allows automated reasoning in "easy cases". A dialogue system establishes the facts but also handles the interface with the citizen.

The easier formalisation of knowledge and semi-automatic knowledge acquisition allow dynamic semi-automatic updating of the knowledge base.

The goal is an ontological index like that in legal commentaries, however, without the textual components. It is obvious that the readability of such ontological structures is limited and will require some training. However, the exact representation of the underlying conceptual und logical structure of the legal system is much better represented.

The DynELCom CONSUL is a model of a semantic legal knowledge system. Legal knowledge is formalised with tools of the semantic web and of legal ontologies. Browsing and handling of the legal text corpus is supported by a conceptual structure with links.

The main difference to existing approaches of legal ontologies lies in the fact that world ontologies (e.g. consular factual situations) are also included in this conceptual structure. As many quite developed ontological representations of world knowledge already exist, such knowledge can be used for enrichment of an ontological representation of the legal system.

A major part of the DynELCom CONSUL consists of the link structure between the facts (world ontology) and rules (legal ontology). Thus, legal reasoning is supported that may be sufficient in "easy cases and a valuable support in contradictory situations.

The formalisation of a legal knowledge domain with the DynELCom CONSUL allows also semi-automatic and automated applications. Conceptual search of links to factual and legal concepts are obvious results of this representation. This search can be supported by dialog systems that support the user in establishing relevant facts of a case. Thus, a sketchy form of automated legal reasoning can be offered, e.g. a "simplified legal syllogism". The facts of a case are properly refined by a dialog system leading to a factual concept but also a legal concept.

The DynELCom CONSUL faces the dynamics of the legal system. The indispensable indexing and analysis process is supported by semi-automatic categorisation and text analysis. Computational linguistics, text extraction, document categorization and text summarization tools are now sufficiently powerful so that good results can be achieved in very short time.

The analysis of the DynElCom is based on a co-operative work model between the man and the computer. The legal information system provides the basis for the commentary. The knowledge base with the ontology and semi-automatic text analysis provides extensive knowledge of the text

corpus of the legal information system. Software tools are information retrieval, hypertext, knowledge management, text summarisation, text categorisation and the inference machine. Manually, ontologies have to be established and maintained, semi-automatic indexing must be constantly fine-tuned and inference engines must be supervised. Such work is presently done by legal authors and practitioners. With the DynELCom CONSUL, a concentration of such analysis takes place in a semi-automatic way. The main advantage is real time delivery, higher quality and lower costs.

The main advantage of the DynELCom CONSUL seems to be obvious: in "easy cases", much of the work can be automated. Consular services would become cheaper with higher quality. Existing pressures on public budgets may lead to cuts in consular networks. With semi-automatic systems, much work can be outsourced to other consular posts of other Member States or honorary consuls.

Text corpus: The basis for the text corpus is the CARE project database. Only few modifications are envisaged; mostly hypertext links to the ontology, visual representations and a list of document types.

Ontology CONSUL: The ontology consists of a legal ontology, a world ontology and links (anchors) between the legal and the world ontology. Elements of the ontology are 3 types of frames: legal frame, fact frame and anchor frame. A frame contains a header, definitions (with sources), classification codes, and relations (to other frames, e.g. synonym, homonym, polysem, hyponym, hyperonym, antonym etc. but also to an anchor). The anchor frame can best be described as a citation with a header, the identification (abbreviation or number) and links to facts and legal concepts. For the representation, existing standards of the semantic web and legal ontologies are implemented; in particular OWL, RDF and LKIF.

This first step with a frame-like representation of legal concepts, factual concepts and the anchors between facts and rules will be followed by a second step that intends a more sophisticated ontological representation of the legal system. This representation focuses on space, persons, actions, material rules and procedural rules.

Action space for persons will be the real space and the cyberspace. Persons can be natural persons (and quasi-persons, e.g. robots or software agents). Objects in the space are physical objects (thinks), energy and quasi-physical objects (e.g. web store). Actions can be physical processes (actions or non-actions in real space or quasi-physical processes (actions on the web). Mental objects and mental processes consist of combinations of these elements. Due to social practice, such "virtual" sets are considered as a unified object or process (e.g. organizations, enterprises, associations, families etc.) Law builds on the existing physical and social structure of persons, objects and processes but modifies it or adds particular elements.

Persons can be natural or legal (e.g. limited company, international organization, state), objects are physical, mental or legal, and actions are physical, mental, or legal. It is obvious that the differences between the social reality and law (as representation as the world should be) are a high interest in any legal system. The representation is structured in concepts (an ontology), rules and factual situations. Isomorphism is respected via direct links to norms but also its logical representation. Legally relevant links between the world ontology and the legal ontology provide support for legal reasoning (e.g. possible factual situations or legal consequences of certain facts). In the LKIF terminology, such a function is called anchors (LRI-Core Ontology). They provide an anchor function as links between the social spectrum of actions and legally.

Knowledge acquisition tools: Text extraction and summarisation tools are decisive for the knowledge acquisition. The tools consist of a knowledge base containing the extraction, summarization and classification rules with header, rule, definition and relations and several tools for semi-automatic text analysis providing information on relevant documents, extract important text passages, classify documents, deliver definitions etc.

We have developed prototypes and applications on corpora-based text analysis for about 20 years now. Due to space restrictions, we can provide only a very short overview of the methods. A pre-defined list of descriptors can be checked against a text corpus with the KONTERM method (Schweighofer 1999). The various term occurrences are clustered according to the context allowing a structuring of homonyms and polysems. Thus, the various meanings in the text corpus can be analyzed. The self-organising map is a general unsupervised tool for ordering high-dimensional data in such a way that alike input items (e.g. documents) are mapped close to each other (Schweighofer 1999). In such a map, similar documents are grouped together. An extension allows the building of various layers and clusters of the map (growing hierarchical self-organising map). Further, common similarities of a cluster can be described with keywords (labelling of self-organising maps). Further, we have also taken advantage of the GATE library for text analysis. The GATE ANNIE (A Nearly New Information Extraction System) tool is very helpful for a more detailed analysis: segmentation of documents (tokenizer), words, gazetteer, sentence splitter and semantic tagger. The GATE JAPE tool (Regular Expressions Over Annotations) is implemented for a similar purpose (Gate 2010).

Sketchy inference engine: In first step with a simplified ontology, the inference is not much more than a hint of relevance like in the information retrieval system. Factual concepts are matched with legal concepts and vice versa. In case of a more complex ontology, an inference engine is required. Decision trees are represented as complex IF-THEN-statements with a mechanism for prioritizing rules. Such statements are interpretations of

facts, rules, concepts and anchors in the ontology. Such an inference engine allows the representations of a legal syllogism and a quicker handling of relevant information.

Dialogue system: Such a system is intended to converse with a human in a coherent structure (Wikipedia: Dialogue Systems 2010, Schweighofer 2010b). In the beginning, the dialogue will be text-based with a graphical user interface. A spoken dialogue system is in consideration. Natural language understanding is supported by a robust parser. The purpose of a dialogue system consists in the establishing of facts but also in the clarification of applicable legal rules.

5. Establishing the Ontologies and First Experiments

Due to time and financial restrictions, the implementation of the DynELCom CONSUL has mostly remained a concept. However, due to our ongoing involvement in the CARE project, we have worked for more about one year on a partial experimental application. The following presentation provides first experiments.

Existing text corpora (RIS, EUR-Lex, CARE) and ontologies resulting from our daily work with European, international and Austrian law forms the basis for these experiments. For Austrian and European law, we have established an ontology with a sufficient granularity of an ontological representation of a jurisdiction: about 10,000 thesaurus entries, 5,000 citations, up to 200 document types, a classification structure, 100 text extraction and summarization rules. This meta data is stored and updated in a database with different types of knowledge frames:

Fact and legal descriptors: header, definition (with sources), examples (with sources), relations (synonym, homonym, polysem, hyponym, hyperonym, antonym etc.), classification, other information.

Anchors: header, identification (abbreviation or number), synonyms, classification, author, links, other information.

Document types: header, identification (abbreviation), use, format, other information.

Classification: header, code, definition, relations, other information.

Extraction and summarization rules: header, rule, definition, relations, other information.

Concepts: header, definition (with sources), related thesaurus entries and citations, relations (synonym, homonym, polysem, hyponym, hyperonym, antonym etc.), classification, legal conceptual structure (ontological model), other information.

Rules: header, quasi-logical expression, source, type, classification, legal conceptual structure (ontological model), other information.

Procedures: header, flowchart, source, type, classification, legal conceptual structure (ontological model), other information.

This ontology was extended to the consular and diplomatic protection.

The following examples may show the lexical ontology (note: an (L), (F) or (A) is added to the header in order to distinguish between legal and fact descriptors as well as anchors). The attribute "legal conceptual structure" indicates relevant branches of law.

Legal concept:
```
Header: Evacuation (L)
Definition: In case of a catastrophe (e.g.
earthquake), EU Member States will evacuate their
citizens (and family members) as a matter of law
or policy. EU Member States will co-operate and
support each other for this goal (Art. 23 TFEU).
Source: Article 23 TFEU, national laws, CARE
project report
Relations: BT catastrophes (F), BT consular
assistance (L), catastrophes (A)
Classification: CAT:EVA
Legal conceptual structure: consular assistance,
catastrophes
Other information: none
```

Fact concept:
```
Header: 2010 Earthquake in Haiti (F)
Definition: Earthquake of 12 January 2010 with an
epicentre near the town of Léogâne affecting about
3 million people in Haiti.
Relations: Catastrophe (F), evacuation (L)
Source: English Wikipedia
Classification: CAT.EAR
Legal conceptual structure: Evacuation (L)
Other information: none
```

Anchor (link):
```
Header: Catastrophes (A)
Links: Terrorism (F), earthquake (F), tsunami (F),
hurricane (F), flooding (F), international
conflict (F), consular assistance (L), Article 23
TFEU (L), evacuation (L)
etc.
```

Figure 1: Examples of frames

At present, we are in the process of finishing the first prototype of this representation of Austrian consular law. The next step would be a verification of the conceptual structure using the knowledge acquisition and text analysis tools. Such a process is very time-consuming and requires financial resources not available so far. However, the existing ontology provides already a very helpful tool for legal work as it represents legal and fact concepts and its links.

6. Conclusions and Further Work

In this paper, we have given an outline of a system for semi-automatic application of consular law in a multilingual and multinational environment, focusing on the underlying legal ontology. For the moment, we are working on a more sophisticated and extended ontological representation.

Acknowledgements

References

Apache Lucene (2010), *High-performance, full-featured text search engine library.* http://lucene.apache.org/java/docs/ (accessed 14 February 2010).

Berners-Lee, T. et al.(2001), *The Semantic Web*, Scientific American Vol. 284, No. 5, pp. 34-43.

Casanovas, P., Biasiotti, M. A., Francesconi, E., Sagri, M. T. (eds.) (2007), *Proceedings of LOAIT 07,* II. Workshop on Legal Ontologies and Artificial Intelligence Techniques, http://www.ittig.cnr.it/loait/LOAIT07-Proceedings.pdf.

Casellas, N., Francesconi, E., Hoekstra, R., Montemagni, S. (Eds.) (2009), *Proceedings of LOAIT 2009,* 3rd Workshop on Legal Ontologies and Artificial Intelligence Techniques joint with 2nd Workshop on Semantic Processing of Legal Text. Barcelona: IOT Series.

Dini, L. et al. (2005), *Cross-lingual Legal Information Retrieval Using a WordNet Architecture*, in: Proc. Int. Conf on Artificial Intelligence & Law 2005, ACM Press, New York, NY, pp. 163-167.

GATE (2010), *General architecture for text engineering,* http://gate.ac.uk/ (last accessed 20 June 2010).

Gospodnetic, O., Hatcher, E. (2005), *Lucene in Action,* Manning Publications.

Hoekstra, R., Breuker, J., De Bello, M., Boer, A. (2007), *The LKIF Core Ontology of Basic Legal Concepts*, in: Casanovas, P., Biasiotti, M. A., Francesconi, E., Sagri, M. T. (eds.) "Proceedings of LOAIT 07, II. Workshop on Legal

Ontologies and Artificial Intelligence Techniques", pp. 43-64. http://www.ittig.cnr.it/loait/LOAIT07-Proceedings.pdf (accessed 16 May 2010).

Manning, C.D., Raghavan, P. & Schütze, H. (Eds.) (2008), *Introduction to Information Retrieval*. Cambridge, Cambridge University Press (2008).

Schweighofer, E. (1999), *Legal Knowledge Representation, Automatic Text Analysis in Public International and European Law.* The Hague, Kluwer Law International.

Schweighofer, E. (2010a), *Indexing as an ontological support for legal reasoning,* in: Yearwood, J. and Stranieri, A. (eds.), "Technologies for Supporting Reasoning Communities and Collaborative Decision Making", IGI Global (In print).

Schweighofer, E., (2006), *Computing Law: From Legal Information Systems to Dynamic Legal Electronic Commentaries,* in: Magnusson Sjöberg, D. and Wahlgren, P. (eds.), "Festskrift till Peter Seipel". Norsteds Juridik AB, Stockholm, pp. 569-588.

Schweighofer, E., Geist A. (2010b), *Semi-automatic application of consular protection law,* in: Schweighofer, E. et al., "Globale Sicherheit und proaktiver Staat – Die Rolle der Rechtsinformatik. Tagungsband des 13. Internationalen Rechtsinformatik-Symposions IRIS 2010". Wien, books@ocg.at (Band 266), pp. 509-513.

Turtle, H. (1995), *Text Retrieval in the Legal World,* Artificial Intelligence and Law, Vol. 3, Nos. 1-2, pp. 5-54.

Wikipedia (2010), *Dialogue systems*. http://en.wikipedia.org/wiki/Dialogue_systems (accessed 20 June 2010).