

IT Support for the Judiciary: Use of Ontologies in the e-Court Project

Joost Breuker, Abdullatif Elhag, Emil Petkov, and Radboud Winkels

University of Amsterdam, Department of Computer Science and Law (LRI),
Postbus 1030, 1000 BA Amsterdam, The Netherlands
{breuker, elhaga, petkov, winkels}@lri.jur.uva.nl

Abstract. The e-COURT project is European project that involves academic, government and industrial partners. The project aims at developing an integrated system for the acquisition of audio/video depositions within courtrooms, the archiving of legal documents, information retrieval and synchronized audio/video/text consultation. The University of Amsterdam is mainly responsible for the development and utilization of (legal) ontologies in this system.

1 Introduction

The e-COURT project¹ aims at introducing information technology in European criminal courtrooms, in particular in indexing audio/video recording of courtroom sessions and their documentation.

- *Audio/Video/Text* synchronization of data from the court trials and hearings.
- *Advanced Information Retrieval*. Multilingual, tolerant to vagueness. Statistical techniques will be combined with ontology based indexing and search.
- *Database management*: multimedia documents (audio & video clips, text, pictures, etc.) supporting effective retrieval of (portions of) these.
- *Workflow management module* defines and manages rules for sharing relevant information and events among judicial actors.
- *Security management*.

2 Role of Ontologies in the Project

The University of Amsterdam is mainly responsible for the developing the (legal) ontologies in the e-COURT project for:

1. **Information retrieval.** The most widely used function of the e-COURT product will be document retrieval. An ontology repository will be used for:
 - *Specialization or extension of queries*. The result set can be automatically expanded or contracted by traversing the (multiple) class hierarchy for more specific or more general related query terms;

¹ <http://www.intrasoft-intl.com/e-court>

- *Translation of queries.* The query terms can be translated (in legal context, when necessary) in order to get documents in several languages in the result set;
 - *Clustering the result set* The typical problem in (WWW) information search is that the number of returned documents may be unmanageably large and heterogeneous. By using terms that are associated with the key-terms used (in general: the values of their attributes in the ontologies), this ‘return set’ can be (additionally) ordered by relevance and clustered by different meanings and views of terms.
2. **Meta-data generation and specification** Ontologies contain generic concepts, i.e. classes rather than individuals or instances, while documents and data-base entries in general contain specific information: individual facts (instances). Therefore, ontologies are suitable to describe the semantics of the stored documents and facts. In this way ontologies are a resource for meta-data descriptions (see [6], [2]). In e-Court ontologies are used to provide meta-data in two ways:
- *Document annotation (tagging).* The documents, important document sections or occurrences of words, designating legally significant entities will be annotated or tagged in XML to not only to improve the indexing and search effectiveness, but also to make the textual and semantic (topic) structure of the document explicit .
 - In the same way as the ontologies can be used to create meta-data in the documents by tagging these, ontologies are also used as a basis for constructing the RDB schemas for databases containing the e-Court documents.

It should be noted that the *structure* of terms in an ontology hardly ever corresponds one to one to that of the schemas of a database. Neither is that the case for the textual structures in documents. Therefore, full automation in tagging and DB schema construction is rather difficult. In [2] one may find a good example of semi-automatic tagging of web-pages. The solutions in e-COURT range from simply providing a standard vocabular to automatic tagging of dialogue turns in hearing documents.

3 Types of Ontologies to Be Used

The information in the criminal court hearings and other documentation is very diversified. There are legal issues, in particular related to the accusation and the formal criminal legal procedures (investigation, court procedures). Then there are the documentation standards of the criminal court. There are the discourse aspects: dialogue turns, questions-and-answers, etc. Further, a large part of the content of the hearing documents contains descriptions of what has happened: ‘stories’ which have an agent-causal structure. All these topics cannot be captured in one, all encompassing ontology. Therefore, we decided to develop a number of specialized ontologies. To prevent incoherence and to supply a more or less uniform point of view on these ontologies we also developed a legal *core ontology*. A core ontology is an intermediary between an upper ontology and a domain ontology [5]. In the next subsections we will present the various ontologies.

3.1 Two legal core ontologies

FOLaw. In previous projects we developed, applied and evaluated a core ontology of law, **FOLaw**, covering the various types of knowledge used in legal reasoning [5]. These types of knowledge are:

Normative knowledge This is probably the most typical legal knowledge: it refers to norms as indicated by deontic operators such as ‘permitted’, ‘forbidden’ and ‘obliged’ but also to concepts like rights and duties.

World knowledge The norms refer to worlds (legal domains) such as crime, or authorship which are assumed to exist and known by the legists and the concerned citizen.

Responsibility knowledge The law is not only concerned with trespasses of law but also who is responsible for trespasses and observing law in general. Often legislation may contain explicit statements of liability (eg parents who are responsible for the conduct of their children).

Reactive knowledge This concerns the kinds of punishments or rewards that the law has in stock.

Creative knowledge The law may create (virtual or real) agents or institutions with a legal status.

Legal meta-knowledge Norms may conflict: any regulation may be riddled with exceptions (on exceptions). To take care of how these conflicts are to be solved meta-legal principles have to be invoked.

This ontology worked fine in analyzing regulations (even very large ones, containing more than 15.000 provisions) and has been the basis of very practical applications, in particular in systems for assessing whether provisions are applicable to a case (see [7]). However, this ontology did not cover very well the legal system itself: in particular its procedures (formal law) and documents. Moreover the ontology is in fact rather an epistemology, i.e. it is concerned with the kinds of knowledge and reasoning in law, rather than the concepts (objects) themselves.²

LRI-Core. Therefore a more generic ontology, **LRI-core**, is constructed that includes FOLaw, but also knows about legal roles, procedures, documentation, communication and legal sources. In fact, because we need a good understanding of distinctions between persons and roles, between physical and mental events, etc. the **LRI-core** has – and includes – many upper-ontology terms [3]. We could not simply start with one of the currently available upper-ontologies (e.g. [4] or in particular the IEEE-Standard Upper Ontology that is under development³ because their focus is rather on describing the physical and formal-mathematical world: not the social/communicative world which is more typical for law. Besides this lack of sufficient covering, we did not agree about the physical part anyway. Our analysis of causation has made us aware that the notion of process as a change of (physical) attributes should be separated completely from the

² Similarly, the categories in knowledge representation, such as ‘concept’, ‘attribute’ etc. form an epistemology, philosophically speaking.

³ <http://suo.ieee.org>

notion of event: i.e. the things that actually occur. Events are understood as instances of processes [3]. The ontological assumptions are:

- Reality consists of physical objects.
- Mental entities behave largely analogous to physical objects: this reflects our ‘naïve’ psychology which assumes e.g. that if one is informed about some fact, this fact is stored in memory. Whether this fact is believed or not is the ‘illocutionary’ view on communication. Facts of belief and knowledge are mental objects.
- Communication proceeds via physical objects but contains mental objects (information).
- In the same way as the world of communication is made up of objects that have a physical and mental view, the world of (cognitive) agents have both a mental and physical (autonomous system) view.⁴
- Social organizations are composed of roles that are performed by agents that are identified as persons.

LRI-Core, and the subsequently presented specific e-COURT ontologies are developed using PROTEGE-2000 (RDF(S)).

3.2 Specific domain ontologies for e-COURT

Three types of ontologies are developed in e-COURT. For each legal system (jurisdiction) a separate version has to be developed because legal content, procedures and documentation are the prerogative of every nation state. Although the statutes and norms may refer to similar worlds and are largely constrained by international treaties and legal practice, there remain important differences. The mappings of these ontologies (worlds) is a problem by itself, which we will not discuss here. In e-COURT ontologies are developed for the Italian, the Polish and the Dutch jurisdictions. The ontologies under development are:

- *Criminal law terminology* This ontology covers the definition of actions that are considered to be criminal (substantive penal law). This ontology is connected to the LRI-core by notions of agent, action, intent and belief, but has also a grounding in physical processes.
- *Trial content and criminal court procedure* This ontology is concerned with ‘formal’ penal law, i.e. about the way the criminal legal procedures should be performed. In principle this ontology should cover the full procedure, i.e. from the start of a criminal investigation to the final verdict at the highest court to be involved. However, in e-COURT the focus is on the trials at the criminal courts. Formal penal law is connected to the LRI-core by the notions of role (judge, defendant, etc), and procedure (assemblies of actions and processes). The more dynamic aspects of a trial are in the ‘spontaneous’ dialogues in the hearing (turn taking, topics). Finally, the typical dialogues are disputes: i.e. their major structure is a dialectical exchange of arguments [1].

⁴ The multiple view evades the classical mind-body problem.

- *Document description entities* These ontologies cover the terms used to specify the document meta-data for the criminal trial documents such as the notions of author, date, authorization, version, structure (sectioning) etc. Only some of these terms map onto the LRI-Core (notably ‘date’ and ‘author’). It should be noted that for hearing and trial documents these notions are very simple. However, for legal documents in general, in particular for provisions (regulations) the structural descriptions may be very complex and well formalized.

In fact, these ontologies will still not fully cover all that is in the trial documents. Legal cases describe states of affairs and events in common sense terms and legal terms. The **LRI-core** covers a large portion of common sense upper terms, but the thousands of common sense terms that describe actual criminal situations are beyond the scope of the project. Therefore, we will try to extend the LRI-core with parts from linguistic ontologies like Wordnet which contain thousands of common sense terms (in a rather superficial way). The advantage of using (Euro-)Wordnet is that it allows for translation between many European languages.

4 Perspectives

The ontology acquisition and development effort for the e-COURT project will be an important step in the development of a legal core ontology that can be reused beyond the e-COURT project. In fact, we develop the core ontology both for the e-COURT project and for the e-POWER project⁵ – an European project aimed at supporting information management and legal drafting in tax-law. The ontologies are developed in RDF(S)/OIL within the framework of the Ontoweb network that is the European platform for the development of the ‘ontology language’ standard for the W3C Semantic Web initiative.

References

1. Gordon, T.F.: *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice.* Kluwer, Dordrecht (1995)
2. Handschuh, S., Staab, S., Ciravegna, F.: *S-CREAM semi-automatic creation of metadata.* In: *Proceedings of EKAW-2002.* Springer Verlag (2002)
3. Lehmann, J., Breuker, J.A.: *On defining ontologies and typologies of objects and processes for causal reasoning.* In: Pease, A., Menzel, C., Uschold, M., Obrst, L. (eds.): *Proceedings of the IJCAI-2001 workshop on IEEE Standard Upper Ontology.* Seattle (2001) 31–36
4. Sowa, J.F.: *Knowledge Representation: Logical Philosophical, and Computational Foundations.* Brooks Cole Publishing Co, Pacific Grove, CA, (2000)
5. Valente, A., Breuker, J.A., Brouwer, P.W.: *Legal Modeling and automated reasoning with ON-LINE.* *International Journal of Human Computer Studies* **51** (1999) 1079–1126
6. Vargas-Vera, M., Motta, E., Domingue, J., Buckingham Shum, S., Lanzoni, M.: *Knowledge extraction by using an ontology-based annotation tool.* In: *In K-CAP 2001 workshop on Knowledge Markup and Semantic Annotation* (2001)
7. Winkels, R., Boer, A., Hoekstra, R.: *Clime: Lessons learned in legal information serving.* In: van Harmelen, F., ed.: *Proceedings of the European Conference on Artificial Intelligence-2002, Lyon.* IOS-Press, Amsterdam (2002)

⁵ <http://www.lri.jur.uva.nl/research/epower.html>