

Legal RDF Dictionary

Murk Muller

Attorney at Law, DataArchitect

Kanzlei Murk Muller

Berlin

Germany

mm@mmrecht.com

<http://www.lexml.de>

Biography

Attorney at Law (Netherlands and Germany) and DataArchitect in Berlin, initiator of LEXML

Table of Contents

The standardisation process in the legal domain

RDF Dictionary

Hierarchy, multiple inheritance, DAML+OIL

Archetypes

Co-ordination and Stakeholders

The standardisation process in the legal domain

XML requires datastructures to realise its full potential. The legal world recognised this need at an early stage. LegalXML <http://www.legalXML.org>, was founded in the USA in the same month as XML 1.0 was released as a recommendation. LEXML <http://www.lexml.de> was founded July 2000. Both organisations have as their goal to ensure that datastructures for XML are created and to coordinate this process of "standardisation" of legal information. At XML Europe 2001 LEXML was confirmed as the European network for standardisation in the legal domain, coincidentally the same place where LEXML was founded: Berlin!

LegalXML and LEXML walk different ways to reach their goals. LegalXML strives at agreeing amongst a substantive number of actors from the American legal world on one DTD for all documents of a particular kind, e.g. contracts, court filing documents, statutes, transcripts. LEXML does not aim at one datastructure per document type, but rather allows and encourages a greater number of XML-datastructures to be created by any community which is willing and capable to do so. The European legal landscape is too diverse, by virtue of its number of jurisdictions, its legal systems (roman, napoleontic, common law), number of languages and (legal) culture, that it would be unrealistic to strive for one structure per document type. Furthermore the expectation is that by encouraging smaller communities to establish their own structures, the process of agreeing on such structures will be simpler and shorter in time.

LEXML's approach seems thereby to ignore a challenge, which comes apparent when one looks at the main feature of XML: exchange of data between heterogenous systems. Exchanging legal data marked up with the same datastructure poses no significant problems. For data stemming from different structures an interface is needed, which "maps" one structure to the other in order to make these data comparable. If LegalXML reaches its goal of creating a limited set(10-20) of DTD's it would also have to provide for mapping between these DTD's, but it would be a task of limited nature, certainly if during the process of creating these DTD's, a good coordination takes place to ensure that element names, attributes and nesting-structures are the same or can easily be matched.

RDF Dictionary

At this point John McClure's proposal for tackling the standardisation challenges should be mentioned. His path holds to a certain extent the middle between LegalXML's and LEXML's approach. John McClure advocates the creation of a legal RDF Dictionary which contains all possible "atoms" needed to create legal datastructures. By the example of the RDF Dictionary which he built for the Dataconsortium <http://www.dataconsortium.org> he has shown that it is possible to create any DTD from the RDF Dictionary by a click on the button. The corpus of atoms of the Dataconsortium RDF Dictionary has a top level of, at present, fifteen prime terms (or base terms), to be brought down to five in the future. A second layer consists of approximately 109 terms, which are categories of each prime term. All other 9000 terms follow from the first two layers, not in an hierarchical fashion, but in the form of a network of terms linked with the multi-inheritance feature of the RDF language. 5000 of these terms are in fact compound terms of 4000 true "atom terms". John McClure is proposing a similar RDF Dictionary to contain all legal terms which would appear in the Legal XML DTD's. In the present draft it would have eight prime terms and around 100 categories. The amount of atom and compound terms is open ended.

Creating a dictionary for Europe containing all possible atoms of present and future datastructures would, for the same reasons of diversity mentioned above be an almost unsurmountable task. A larger group of key actors of the European legal domain is unlikely to agree on the content of such a dictionary, a smaller group is unlikely to produce a dictionary which has enough authority to steer the standardisation process. So, here again the natural choice is for smaller communities and an organic growth of a number of RDF Dictionaries. These RDF Dictionaries serve as an interface for datastructures which are already there, but at the same time is a source for new datastructures. The RDF Dictionary concept is applicable on many levels: from the level of one small particular domain, or a small geographic area, to a national level, bilateral level, going on to an international, supranational and finally global level. It is possible and desirable, that on all of these levels RDF Dictionaries will come into existence. These Dictionaries complement and reinforce one another by forming a network. Each RDF Dictionary can take advantage of the work which has been done for other RDF Dictionaries by the simple, but very effective, namespace mechanism provided by XML/RDF. The architecture of the RDF Dictionary allows for organic growth not only of one particular RDF Dictionary itself, but also of the network, a network of structure.

Hierarchy, multiple inheritance, DAML+OIL

Keywords: hierarchy

The traditional classification of legal terms is a hierarchy. The broadest term stands at the top, refined by narrower terms. Each (narrow) term has no more than one broader term as its "parent". Hierarchical classification is, after a flat list, the easiest way to classify terms. For a long time people have been aware that hierarchical classification is a rather inadequate way of describing legal reality. The reason that hierarchical classification remains the most practised way of classifying probably lies in the means of storage. Any piece of information should be stored just in one place. If one stores it in more than one place, the risk of inconsistencies is created. The strict hierarchical structures in traditional systems are often softened by card indexes, thesauri and other cross linking methods. But the basis is and remains a hierarchical structure. Legal databases have, so far, not brought any significantly other method of storing and retrieving, apart from full text search. Full text search offers relief in some cases, in many cases it is a rather inaccurate and inefficient way to find one's way in an information surplus.

Legal ontologies hold promises in this field. Where they apply RDF based ontology languages like DAML+OIL ontologies break through hierarchy by allowing multiple inheritance. A (narrower) term can have more than one broader term as a parent. A legal ontology contains a structured view of the legal system, which view potentially comes closer to legal reality than traditional structured views.

What does the RDF Dictionary add to these developments? The RDF Dictionary is not a legal ontology, at least not in the traditional sense. It does in itself not try to describe the legal system. What it does, is link structures to one another. It facilitates the interoperability, the communication, between data structures. Such a structure is in most cases an XML Schema or DTD, describing the structure of a particular kind of legal document, like judgement or contract. There is, however, no reason to exclude DAML+OIL ontologies from the benefits of the RDF Dictionary. The legal RDF Dictionary not only link structures, it also offers a user the possibility to compare various structures and use the structure which best suits his needs at that moment for the particular task he is performing. A structure once made is in principle static. As things change over time a structure may in parts loose its usefulness. The RDF Dictionary makes sure a user can always choose to use the latest structure, or build a custom structure from parts of existing structures.

Archetypes

One to one translation of terms originating from different jurisdictions often lacks the desirable precision. The RDF Dictionary therefor uses the concept of "Archetypes" to achieve a more precise translation of, for instance, the German term "Urteil", the similar Dutch term "vonnis" and the English term "judgement" or "verdict". A set of "Archetypes" defines aspects of these terms. If in an XML instance document one finds a term "Urteil", by virtue of the instance document having been marked up according to a datastructure which is mapped to the network of RDF dictionaries, one is able to establish a precise as possible meaning of that term in the context of one's own legal system. The mechanism to establish the meaning uses the fact that one own legal system is also contained in datastructures linked to the RDF Dictionary network. The mechanism is illustrated with the example of Urteil/vonnis/judgement/verdict in a very first draft of a legal RDF Dictionary to be found at <http://rdf.lexml.de>.

Co-ordination and Stakeholders

LEXML, and LegalXML both host a project for the development of an RDF Dictionary. The American and European efforts are coordinated. The undersigned is co-chair of the Dictionary Workgroup of LegalXML, of which Workgroup John McClure is the Chair. John McClure closely follows the development of the European version of the RDF Dictionary. The undersigned and John McClure have spent many hours in face-to-face discussions ensuring the end-convergence of the developments of the American and the European legal RDF Dictionaries and their compatibility in the intermediate period.

A number of European governments is considering to develop an RDF Dictionary for their national legal system. The EU has recognised the potential of the legal RDF Dictionary for the integration of Europe. The legal publishing industry, by virtue of its SGML history a major source for legal datastructures, has shown interest in the legal RDF Dictionary. Also, the W3C has expressed interest in the RDF Dictionary concept. Their support will contribute to acceptance and further development. An overview of parties involved with the development and occasions where the legal RDF Dictionary was discussed, can be found at <http://eu.lexml.de>.