

Ontologies, ICTs and Law

The International *Ontojuris* Project

Bibiana Luz Clara¹, Ana Haydée Di Iorio², Roberto Giordano Lerena³

1: Abogada. Profesora Investigadora de la Facultad de Ingeniería de la Universidad FASTA. Presidente del Instituto de Derecho Informático del Colegio de Abogados de Mar del Plata. Argentina. {bluzclara@ufasta.edu.ar}

2: Ingeniera en Informática. Profesora Investigadora de la Facultad de Ingeniería de la Universidad FASTA. Instructora Informática en Ministerio Público de la provincia de Buenos Aires. Argentina. {diana@ufasta.edu.ar}

3: Ingeniero en Sistemas. Profesor Investigador y Decano de la Facultad de Ingeniería de la Universidad FASTA. Argentina. {rogjord@ufasta.edu.ar}

Abstract. This article presents the experience of the International *Ontojuris* Project, modeled and developed to search and retrieve multilingual legal information based on ontologies and on the Universal Networking Language (UNL). It also presents the issue of multilingual information management, the importance of data processing from the semantic point of view and the possibility of semantic interoperability between systems, basically on Web search engines.

Key words: Ontology, Law, Artificial Intelligence, UNL, *Ontojuris*

Documentary Legal Informatics in Argentina

In the beginning, the development of legal informatics captured the attention of law practitioners to improve their working practices and increase accessibility to information and documents [1].

Due to the large amount of legal information in existence, it was necessary to find a support to facilitate access to this information, both to legal practitioners and citizens. The Documentary Legal Informatics would thus develop aiming at the automatic processing of legal information sources: legislation, jurisprudence and doctrine [2].

In Argentina, the best example of Documentary Legal Informatics is the Sistema Argentino de Informática Jurídica, SAIJ (Argentine System of Legal Informatics) created in 1979. The SAIJ is a government agency supervised by the Dirección Técnica de Formación e Información Jurídico-Legal (Technical Office for Legal Information), under the Subsecretaría de Justicia (Justice Subsecretariat) in the Ministerio de Justicia, Seguridad y Derechos Humanos (Ministry of Justice, Security and Human Rights). It provides normative, jurisprudential and doctrinaire information, whether national or provincial, taken from official sources.

The SAIJ¹ also coordinates the National Network of Legal Informatics, established in 1995. This net is constituted by all provinces that have signed agreements with the entity. Each province has a cooperation center in charge of providing and updating the provincial legal information².

In the Argentine legal system, jurisprudence is a formal source of law. For this reason, when a law practitioner carries out a jurisprudential search, he is seeking to reinforce the interpretation of standards or a personal point of view. In short, he attempts to present, by reference to verdicts, persuasive arguments to influence the judge's reasoning towards his side.

In addition to being limited by the syntactic search, most of these legal information search systems require the thorough knowledge of the verdict which the operator is trying to find: the year, actors involved, the court, and subject matter. At the moment of search, both in government initiatives and in private ones related to legal publishers, the legal practitioner frequently retrieves a large amount of irrelevant data that should be refined repeatedly until obtaining the desired result.

Many Artificial Intelligence (AI) techniques related to the representation of knowledge have tried to solve this problem. Among them, the representation by "ontologies" is noted; it refers to the formulation of a conceptual scheme within a given domain to allow the search of knowledge through meaning. This "ontological" representation is the basis for a real "Semantic Web" [3][4] by which the legal practitioner will be able to retrieve information from concepts, semantically, or by obtaining the exact data related to the search, all these independently from the possibility that in the referred text the specific term could be used at the moment of query [5][6][7].

In addition to the usual problems of legal search, the globality of law appears together with the complexity of varied conceptualization related not only to language but also to the particular cultures to which the concept refers.

Integration in a multilingual world

In a globalized, multicultural and multilingual world, access to resources is limited by multiple barriers, among them language and those originated in the interpretation of the real world to be conceptualized.

¹ Source www.saij.jus.gov.ar – Institutional

² There are also numerous initiatives of legislative, executive and judicial entities linked to the publication of legal information. Namely, the JUBA system of the Supreme Court of Justice of the Province of Buenos Aires, now accessible via the Web. It includes summaries and complete verdicts in the Province of Buenos Aires; the FANA system, also accessible on the Web including nationwide summaries and verdicts.

The use of ontologies, and conceptualization in itself, is complex, no matter which culture or language it is dealt with (or conceptualized). A higher level of complexity occurs when expanding the coverage of the proposed solution and the necessary conceptualization, to a multicultural and multilingual world.

Culture, terms, concepts, relationships between terms and term-concept relationships differ from one place to another, from one language to another. Whatever the domain, a given conceptualization that is valid in a certain language may not be acceptable in a different one. Thus, it is not possible to automate the process to a strict automatic translation, and, above all, it is impossible to homologate terms and concepts in different languages. There are words which cannot be translated in any language, simply because its strict meaning and use in its place of origin (in that specific culture) does not have a strict equivalent in another culture. Each term is the representation of a concept in a given language, and it may turn out that this concept will not have its equivalent in another language; hence, a possible word to represent that concept in that second language will not exist. Even in the same language (Spanish, for example), the same word may be used to represent different concepts in different countries or regions, and the same concept may be represented by different words in different regions using the same language.

The cultural complexity of languages and multilingualism are then transformed in a barrier to communication. In an interconnected and globalized world, it is urgent and necessary to undertake these issues from a technological point of view in order to facilitate intercultural communication.

Information on the web is growing daily and there is an urgent need to find “intelligent” searchers capable to work with semantics in the language and place where the search is performed. Users need to search by concept, not by term. Users think according to concepts, but must search the web for terms. The search is usually syntactic, not semantic. Browsers retrieve and return web pages containing search terms as they are spelt, textually, and not semantically. Some of them even propose pages in different languages where the specific term appears (having a different meaning in that language), without the ability to discriminate or prioritize on behalf of the concept.

There is also specific terminology in each domain or specialty which makes certain terms (words or set of words) have different meanings in a language, being the same country and culture. There are also concepts built on the basis of words that separately, have a certain meaning, but with a composition that does not mean the composition of those meanings. Traditional translators do not recognize this kind of compositions, namely, the representation of new concepts.

Users around the world need to see web pages from other countries, but on the basis of a specific concept, not terms representing it in their languages. This requires the development of search engines capable to understand and process the concept associated to the indicated term; with that concept (or meaning), engines should find pages having any of those terms or expressions representing it in their respective languages. These are known as intelligent search engines. They index and retrieve on-line meanings or concepts instead of words or terms. They include a conceptual infrastructure and ontological relationships that allow such management of search.

Given the importance of the term (actually, the concept) of the query, its vital correct interpretation, and since the above mentioned must be limited to the scope of law, the problem is increased. Misconception of a legal document is a very high risk that law practitioners cannot take; consequently, they need the support of technological tools to collaborate with their work and guarantee the correct interpretation of data, terms, information and documents involved in their decisions.

The UNL Program

In 1996, the United Nations General Assembly crated the Universal Networking Language Program (UNLP) as a project of the United Nations University (UNU).

The aim and activities of the UNLP are to develop and promote platforms and communication and information tools that will provide every nation the same opportunities to access, share and exchange scientific, cultural, social, and economic resources available in the global village. The UNLP has a flexible and dynamic net of persons and institutions devoted to developing, expanding, improving and multiplying the UNL System www.fi.unl.upm.es as a means of overcoming linguistic barriers; it is also a platform to collect and multiply human knowledge among people speaking different languages³.

Ultimately, the project aims at allowing any person to share and retrieve information in their own language, no matter the language originating it. The project counted with an initial participation of 15 languages: German, Arabic, Chinese, Spanish, French, Hindi, Indonesian, English, Italian, Japanese, Latvian, Mongolian, Portuguese, Russian and Thai. The UNL System basically consists of UNL servers, UNL editors, and UNL viewers. The UNL language consists in UNL relationships and their attributes, universal terms, and a knowledge database⁴.

³ Source: www.unl.fi.upm.es

⁴ source: www.undl.org

The Centro de Lengua Española (Spanish Language Center) and their group are working under the assistance of the Universidad Politécnica de Madrid (UPM) and they represent the Spanish language, not only in Spain, but also in every country sharing this language. Consequently, with this pretended universal program, the UNL language is adopted as the platform for the development of a multilingual legal server based on ontologies to pursue the International *Ontojuris* Project.

The *Ontojuris* Project

The International *Ontojuris* Project aims at facilitating a multilingual access to information about legal documents in the areas of Intellectual Property Law, Consumer Rights and Informatics Law. A consortium was then formed by researchers from Argentina, Brazil and Spain. Argentina was represented by Universidad FASTA; Brazil, by Instituto I3G; and Spain, by Universidad Politécnica de Madrid. Experts from Universidad de Chile are also collaborating with the project.

The overall objective of the program consists in the research and development of an intelligent multilingual system based on ontologies, for the retrieval of legal information, limited in a first stage to the domains of Intellectual Property Law, Consumer Rights and Informatics Law [8].

Broadly, the stages in this project are as follows:

- a) Selection of texts related to legislation, jurisprudence and doctrine of the domains involved in order to generate ontologies associated to each domain.
- b) Identification and definition of terms and patterns of relationship. Construction of specific ontologies⁵.
- c) English determination of Headword⁶ for each term.
- d) Conversion of each term to the UNL and construction of domain terms inexistent in dictionary⁷.

⁵ The Project is developed under its own ontology editor, provided by I3G which anticipates the definition of some relationships.

The definition of ontologies was based on the identification of terms and by linking them after the following relationships: Synonymy, Type of (category or class), and Part (fraction or component).

⁶ The UW (Universal Words) constitute the vocabulary of the UNL. They are concept labels, syntactic and semantic units that combine to form the UNL expression. Each UW represents a concept. A UW is formed by a Headword and a list of restrictions. The Headword may be a word, a compound word or a phrase in English. The list of restrictions is associated to the Headword to disambiguate and add specifications.

⁷ The retrieval of universal words to fulfil the Ontology was achieved by referring to the UNL dictionary available at the Centro de Lengua Española. Those not included in the data base were created.

- e) Definition of measures for indexing ontologies⁸.
- f) Definition of parameters allowing flow of ontologies.
- g) Development of procedures for the integration of the ontology editor with applications and tools for the web search.
- h) Modelling of the tool interphase.
- i) Integration of the UNL to the ontology editor.
- j) Expansion of search through the Universal Word (UW)⁹.
- k) Specification of the result presentation.

In developing the tool, the methodology of Knowledge Engineering, based on the semi-formal ontology description, was used to support the process of ontology engineering. In this methodology, instances of representation do not include the description of objects, only their relationships within a given domain.

The editor was designed to support the task and experience of the Knowledge Engineers when constructing the multilingual ontologies. It is a complex structure which connects terms taking into account concepts in the knowledge of their specific application. This allows the editor to determine the context of the documents in the query: they are contextualized.

The basic components of the ontology editor are: classes (taxonomically organized) and relationships (representing the type of interaction between concepts in a given domain). The ontology representation does not use axioms or instances.

Discussion. Aiming at the future.

Having concluded the *Ontojuris* prototype, it is now time to verify its utility “in the field”, with law practitioners from different countries validating the system.

⁸ The *Ontojuris* system uses the Ontology created by the editor to index and retrieve information in the specified legal documents (laws, decrees, doctrine). The terms established by the method of creation of Ontology are used in the indexing process. The terms considered as relevant in a given document are added to the list of terms. On the other hand a list of words is generated from a dictionary of the natural language of the document. Hence, each document is labeled with the indices of terms and the words it contains.

⁹ This phase, not yet completed, suggests a new expansion method based on domain ontologies and UW in order to retrieve multilingual information. This is an ongoing study based on the possibility of relying on a domain dictionary of UW for each natural language of the original documents. Each document to be searched is converted into a term vector and a word vector; besides it is mapped with a UW vector by which it is also indexed. That is to say, during indexing, the system converts each term into its corresponding UW, and it converts the original term into each different language associated to that UW.

Future steps must also be discussed. On the one hand, the expansion of the project to other disciplines given that this type of browser may be adjusted to a domain in any field provided that experts accomplish an accurate selection of ontologies.

On the other hand, members of other languages will be invited to the consortium to expand multilingual competences to other languages.

In the field of system integration, and given the level of knowledge embodied in the ontology of Law, it is necessary to work for an on-line integration with Law systems in Latin America certifying the semantic inter-operability among the systems.

Finally, it must be noted that the participation of the consortium Universities in the future course UNESCO “TECLIN” – Linguistic Technologies for Children Education in Aboriginal Communities will allow methodologies and the developed technology to extrapolate to other fields and contribute to the fulfilment of the United Nations goals for the millennium.

There is already a pilot project in Argentina for the recovery of endangered languages (particularly, the *Quechua*) which enables a “dialogue” between modern languages, such as Spanish, and aboriginal languages (declared World Heritage Site) favoring conservation. The project is developed on the same methodological and technological basis carried out for interaction between different languages in the field of Law. With very encouraging preliminary results, the Centro de Investigación CIPCO (CIPCO Research Center), La Buhardilla Foundation, in Tucuman, Argentina, is working in this direction with the support of the *Ontojuris* Universities.

Conclusion

After overcoming issues like the availability of digital information, connectivity and technical interoperability, cultural diversity and linguistics appear to be the real problems to reach a global knowledge society. It is at this point where technology of information has a fundamental role and an enthralling challenge.

The *Ontojuris* project reveals the potential of technological tools available to add up to the socialization of knowledge in the great “Global Village”.

Acknowledgements

The authors are grateful to the members of the *Ontojuris* project of Universidad FASTA and to the members of the *Ontojuris* Project in Spain and Brazil. Also, to Tania Bueno, Sonali Bedin, Hugo Hoeschl, Cesar Stradiotto and Jesús Cardeñosa.

We would like to thank Ana Inés Cosulich for her assessment in the English version of this paper.

References

- [1] Peñaranda Quintero (2001), *Iuscibernética, interrelación entre el derecho y la informática*, Ed. Miguel García e hijo, Caracas, Venezuela.
- [2] Luz Clara, Bibiana (2001), *Manual de Derecho Informático*, Ed. Nova Tesis, Rosario, Argentina.
- [3] Castells, Pablo, *La web semántica*, disponible en <http://arantxa.ii.uam.es/~castells/publications/castells-uclm03.pdf> (accedida 2 de Mayo de 2009)
- [4] Castells, Pablo, *Búsqueda semántica basada en conocimiento*, disponible en <http://nets.ii.uam.es/publications/castells-fds08.pdf> (accedida 15 de Abril de 2009)
- [5] Pompeu Casanovas (2005), *Ontologías jurídicas profesionales. Sobre conocer y representar el Derecho*, disponible en http://www.leibnizsociedad.org/secciones/mater/pon/textos/ontologias_pompeu.pdf (accedida el 17 de Junio de 2009)
- [6] Abian, Miguel Ángel (2005), *Ontologías, que son y para que sirven*, disponible en <http://www.wshoy.sidar.org/index.php?2005/12/09/30-ontologias-que-son-y-para-que-sirven> (accedida el 4 de Octubre de 2009)
- [7] Burners Lee, Tim (2000), *Conference on the Semantic Web* disponible en <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html> (accedida el 20 de Octubre de 2009)
- [8] Proyecto Ontojuris: Disponible en www.i3g.org.br/ontojuris/sistema.html