

The LOIS Project

Wim Peters

University of Sheffield
Regent Court, Portobello Street
Sheffield S1 4DP
United Kingdom
w.peters@dcs.shef.ac.uk

The main task of the EU funded eContent project LOIS¹ (Lexical Ontologies for legal Information Sharing) is the development and connection of 6 legal WordNets based on the EuroWordNet (EWN) framework (Vossen et al, 1997). Using the EWN framework assures compatibility of the LOIS WordNets with EWN, allowing them to function as an extension of EWN for the legal domain. Ten partners from six European countries (seven universities/research centres and three enterprises) participate in this project. Within the approved project duration of 24 months, around 5000 synsets will be localized for each language involved. The LOIS project primarily aims at providing easy mono- and multi-lingual access to European legal databases for legal experts as well as for laymen. Further research will focus on improved techniques for information retrieval, on providing standards for the representation of legal documents, on the commercial use of public sector information, on showcase applications for test and demonstration purposes, and on product placement for integration of the result into commercial applications. To reach this goal, WordNets of six different languages (Italian, Dutch, Portuguese, German, Czech, English) will be localized and cross-linked through an unstructured inter-lingual index (ILI).

The existing Italian legal WordNet ‘JurWordNet’ (JWN) (Gangemi et al. 2003), which was developed as an extension of the Italian EWN, provides the basis for the LOIS lexical database (the first module of the LOIS database). Furthermore, the Italian JWN is used as the ILI to map the different WordNets. Before the start of manual localization, an automatic intersection of the 1695 synsets of the Italian JWN with EuroDicAutom² was made. Subsequently, a mapping was created between the English result list of 579 literals and the Princeton WordNet 1.6. The WordNet structures of the different WordNets have been established analogously to the Italian JWN. Up till the present moment, the manual revision, adding of definitions, and integration have been going on.

The legislative database (the second module of the LOIS database) is based on legal definitions extracted from EU sources and, for the sub-domain consumer protection law, also from the national transformations and other relevant national provisions. For this purpose, a tool was developed to

extract legal definitions from European directives. Definitions of different language versions have been automatically connected and national implementation measures are being added manually. As a result of the distinction of a lexical database and a legislative database, two different types of concepts are represented within LOIS: *lexical concepts*, designated by terms and the lexical meanings assigned to them, and *legal concepts*, designated by terms and their definitions from legal documents.

Regarding language internal relations, the complete set of EWN relations is available, but primarily the lexical relations (near-)synonymy and the taxonomic relations hyponymy/hyperonymy are used. Equivalence relations between synsets in each language are made explicit in the ILI, whereas each synset in monolingual WordNets has – either directly or indirectly by related synsets – at least one equivalence relation with an ILI-record. For demonstration purposes, the sub-domain of consumer protection law will be further structured with other WordNet relations.

Figure 1 shows a schematic presentation of the modular LOIS architecture, with the Italian legal database (IT) as example. The main LOIS module is the National Legal WordNet. This is composed of lexical and legal concepts. The first type consists of lexical concept representations. The second type covers legal terminology. These occur in national legislation, and therefore, they are part of the National Legal WN (NC2 in Figure 1), and in EU legislation (NC1 in Figure 1), in which case they are, because of their pan-European character, part of the National Legal WN on the one hand, and the ILI on the other. Each National Legal WordNet concept representation has a number of information fields associated with it. These provide information on e.g., language, orthography, definition and associated field of law. Any of these National Legal WordNet concept representations present in language specific synsets (LSS in Figure 1) of the corresponding EWN language components are linked to these synsets by means of plug-in relations (Magnini and Speranza 2001).

All National Legal WordNet concept representations are linked into the Inter-Lingual-Index by means of equivalence relations. Furthermore, an ‘implemented as’ relation has been introduced to indicate the link between EU concept representations and their nation-specific implementations. The Legal Document Index contains keys into national

¹See <http://www.loisproject.org/>

²See <http://europaeu.int/eurodicautom/>

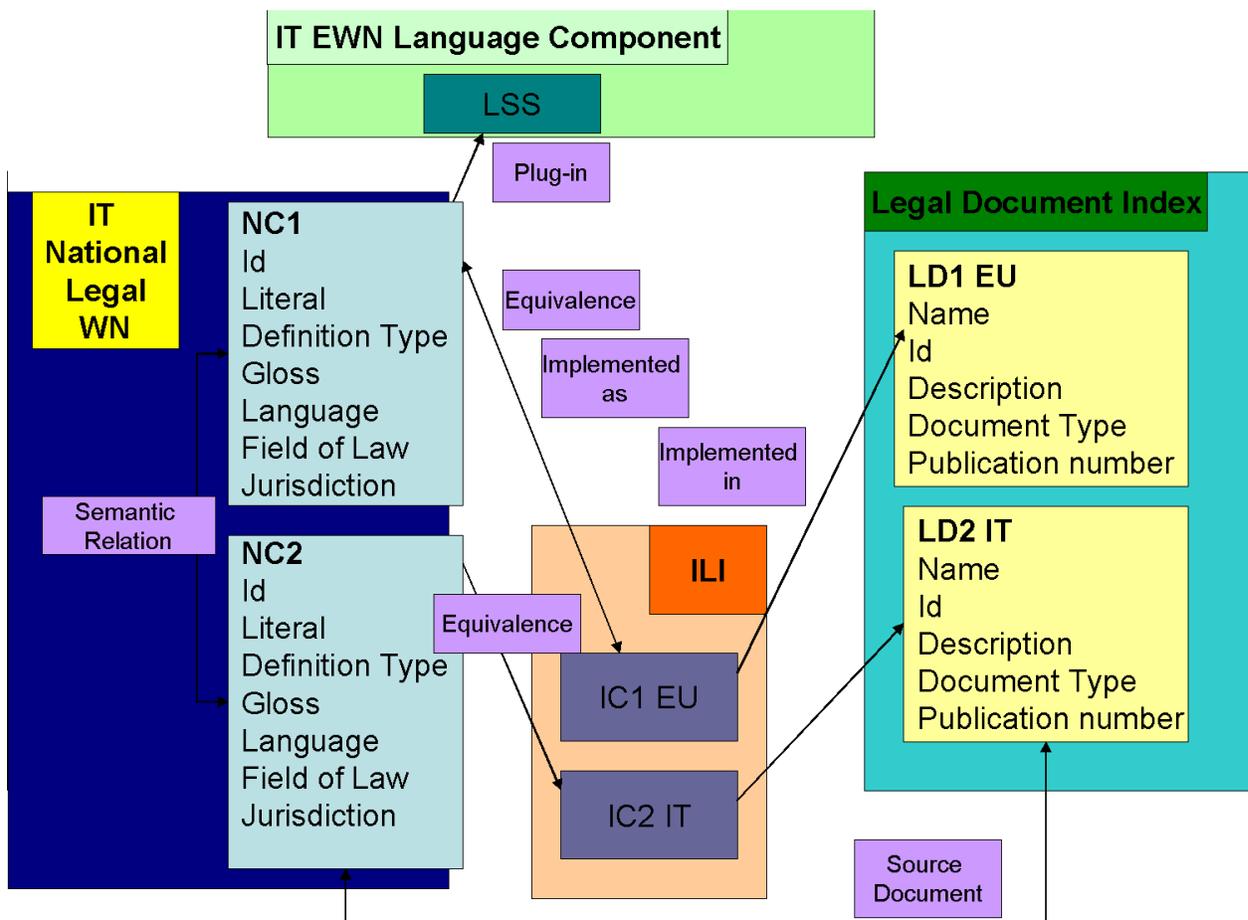


Figure 1: The LOIS database lay-out

and European legislative texts in which the legal terms are explicitly used and defined. Currently, we anticipate to establish a consolidated legislative database, comprising current (thus, no historical) versions of statutes. Each legal document (LD in Figure 1) has a number of information slots associated with it that further specify its nature. The main information is provided by the identity number that is taken from the CELEX database³ for EU documents, and local categorizations for national documents.

Overall, the LOIS architecture will allow users to investigate a wide range of legal issues, such as:

- multiple senses of terms, due to different legislative sources;
- differences between definitions of concepts in EU and national legislation through the *implemented_as* relation;
- relations between EU and national legislative documents through the *implemented_in* relation;
- comparisons of national legal systems;

- lexical definitions of concepts, if no terminological definition is given;
- *comparisons between common language meaning and terminological legal meaning through available plug-in links.*

References

- Gangemi, A, M.-T. Sagri, D. Tiscornia (2003), *Jur-Wordnet, a Source of Metadata for Content Description in Legal Information*, in: Proceedings of the Workshop *Legal Ontologies & Web based legal information management*, The International Conference of Artificial Intelligence and Law (ICAIL 2003), Edinburgh, June 24, 2003.
- Magnini, B and Speranza, M. (2001), *Integrating Generic and Specialized WordNets*, in: *Proceedings of the Euroconference RANLP 2001*, Tzigrav Chark, Bulgaria.
- Vossen, P., Peters, W. and Díez-Orzas, P. (1997), *The Multilingual design of the EuroWordNet Database*, in: Mahesh, K. (ed.), *Ontologies and multilingual NLP, Proceedings of IJCAI-97 workshop*, Nagoya, Japan, August 23–29.

³See http://europa.eu.int/celex/htm/celex_en.htm.