Terminae Method and Integration Process for Legal Ontology Building

Sylvie Despres[1] and Sylvie Szulman[2]

[1] CRIP5, Equipe IAD, Université Paris 5
sd@math-info.univ-paris5.fr
[2] LIPN, UMR7030 Université Paris 13
ss@lipn.univ-paris13.fr

Abstract. This paper describes the contruction method of a legal application ontology. This method is based on the merging of micro-ontologies built from European community directives. The terminae construction method from texts enhanced by an alignment process with a core legal ontology is used for building micro-ontologies. A merging process allows constructing the legal ontology.

1 Introduction

This paper presents the contruction method of a legal application ontology. The aim of this ontology is to support the development of formal models of legislation to be used in legal knowledge-based systems. This construction method is based on micro-ontology building and their merging. A micro-ontology refers to a set of concepts and properties describing a domain's restricted context. Each micro-ontology from a European directive is achieved both by using the semi automatic terminae method [3] and by aligning it with a legal core ontology clo [11]. A core ontology covers a field such as law which may consist of many subdomains like criminal law, private law, European Union law, etc. Terminae is based on knowledge elicitation from texts and allows creating a domain model by analyzing a corpus with Natural Language Processing (NLP) tools as for the sekt project [18]. The method combines knowledge acquisition tools based on linguistics with modeling techniques so as to keep links between models and texts. During the building process [1], it is assumed that: (1) the ontology builder should have a comprehensive knowledge of the domain, so that she/he will be able to decide which terms (nouns, phrases, verbs or adjectives) are domain terms and which concepts and relations are labeled with these domain terms; (2) the ontology builder knows well how the ontology will be used. The alignment process takes place during the micro-ontology construction. [16], [14] and [10] defined ontology alignment as follows: ontology alignment consists in bringing two or more ontologies into mutual agreement, making them consistent and coherent and allowing the aligned ontologies to reuse information one from the other. In alignment, the original ontologies persist, with links established between them. Alignment is usually performed when the ontologies cover complementary domains. The alignment process was performed mostly

by hand, with the TERMINAE tool. Therefore our ontology is structured around CLO. TERMINAE provides easy import of concepts among CLO but doesn't check whether consistency is maintained once the operations are performed. The merging process [8] is achieved on the micro-ontologies. It consists in creating a single coherent ontology from two or more existing ontologies with overlapping parts.

The article is structured as follows: Section 2 describes our ontology construction method; Section 3 implements the method on two European directives around the concepts TRAVAILLEUR (EMPLOYEE), CITOYEN (CITIZEN); Section 4 presents a comparaison with related works. The article ends with a discussion between ontology and text.

2   Construction Method

Our construction method for building the application ontology includes two steps: - building micro-ontologies from texts using TERMINAE and alignment with a core ontology; - merging micro-ontologies obtained at the first step.

2.1   Building Micro-ontologies

Our micro-ontology construction is based on the semi-automatic TERMINAE method enhanced by an alignment process with a core ontology. The interest of the alignment with a top-ontology as DOLCE and a core legal ontology as LRI-Core [5] has been shown in [9]. The legal core ontology concepts constitutes the common conceptual denominators of the field, therefore, normally, their reuse is facilitated. In this work, CLO has been chosen to achieve an alignment from the natural language definition of the concepts. CLO was chosen because it is built from DOLCE and is available in OWL language whereas LRI-Core is still under development.

The TERMINAE method works with the terms that occur in the analyzed texts so as to describe the thus constructed concepts in a formal ontology. This ontology involves two kinds of concepts: terminological and non terminological. The terminological concepts are created through the ontology elaboration process. They are linked to their term occurrences in the corpus through their terminological forms. These forms contain a definition of the concepts in natural language and a list of synonyms. The non terminological concepts are created to structure the ontology. The TERMINAE tool provides material support to the method. Moreover, TERMINAE allows importing and exporting ontologies in OWL language, which aids the alignment and merging processes. SYNTEX [4] is used in the linguistic study activity to obtain the domain terms. The LINGUAE concordancer included in TERMINAE allows pattern recognition. MFD [7] uses association rules to find lexical relations.

2.2   Merging Micro-ontologies

Ontology merging aims to create a unique ontology from the orginal ontologies. The original ontologies will be unified and replaced by the new ontology. Before

merging micro-ontologies there has to be a correspondance between the concepts of the ontologies to merge.

[15] define the ontology integration as the iteration of the following steps: (1) find the places in the ontology where they overlap; (2) relate concepts that are semantically close via equivalence and subsumption relations; (3) check the consistency, coherence and non-redundancy of the results.

[16] defined the set of basic operations for ontology merging as: removing a class from the ontology; removing a concept from the list of parent concepts because a more appropriate parent exists in another ontology; adding a concept to the list of parents of the considered class because an appropriate parent concept has been found in another ontology; renaming a concept to conform to naming conventions in another ontology; moving a concept from one ontology to another; removing a slot from the list of slots of a concept because a similar slot was inherited from a new parent concept; moving a slot because it is more appropriately defined in a parent concept; renaming a slot. The problems that arise when merging are the mismatches that may occur between separate ontologies. Ontologies may differ at two levels: language or meta-level (level of the primitives that are used to specify ontologies). In our context, mismatches at the language level do not occur because the micro-ontologies are written in the same language (OWL). The mismatches occur at the ontological and terminological levels.

3   Results

In his section, our first results about the legal application ontology are described. The contruction of this ontology was initiated by using two directives (hereafter refered to as "D-travailleur" and "D-citoyen").

First, the application of the linguistic method (cf 2.1) to the micro-ontology construction from "D-travailleur" is detailed. Then, the micro-ontology established from "D-citoyen" is presented. Finally the merging of these two micro-ontologies is studied.

3.1   Micro-ontology Building

The studied directives were written by the European Union Council which is the first decision organ of the Union. The French version is the support of the linguistic study but we used the English version to translate our examples. The first one is "Council directive 2001/23/EC of 12 March 2001 on the approximation of the laws of the Member States relating to the safeguarding of employees' rights in the event of transfers of undertakings, businesses or parts of undertakings or businesses". The second one is "Council Directive 2004/58/EC on the right of citizens of the Union and their family members to move and reside freely within the territory of the Member States". Henceforth, the "D-travailleur" is used to present our construction method of micro-ontologies.

The Corpus Constitution. The corpus consists of a single text ("D-travailleur") of seven pages (3133 words) selected by jurists. This corpus is imposed, closed

and small sized. This small size is not an obstacle to the model elaboration because a directive is self-contained and cohesive.

Linguistic Study. This step consists in selecting the terms and lexical relations that will be modeled. The results of this stage are quite raw and will be further refined. SYNTEX, LINGUAE and MFD are the used tools. SYNTEX yielded 900 candidate-terms. The most used terms were: *travailleur* (*employee*) (56 occurrences), *transfert* (*transfer*) (41 occurrences), *cédant* (*transferor*) (15 occurrences), *cessionnaire* (*transferee*) (14 occurrences). Because we work on a small corpus, there are many hapax (a term which appears only once in the corpus) which are significant and have to be kept such as the syntagm "community charter" of the Fundamental Social Rights or the verb "to abrogate". LINGUAE and MFD tools have been used to explore the relations between central concepts. For example, MFD detects two patterns (notify to transferee, transfer to transferee). Then the relation between transferor and transferee is given by a pattern of LINGUAE (see Figure 1).

(lemme : cédant) (*) (lemme : cessionnaire)	(lemma: transferor) (*) (lemma: transferee)
cédant notifie au cessionnaire	(transferor notifies transferee)

Fig. 1. Relation study with LINGUAE

Normalization. The normalization step is a semantic study and constitutes a bridge between the lexical and syntactical study toward modeling. It consists in a particular conceptualization process based on corpus analysis. The concept description is based on the semantic analysis of the term occurrences. The semantic interpretation of the terms is driven by expert knowledge and application requirements. The method starts with the study of the central concepts of the model. These central concepts are commanded by the aim of the ontology building namely the description of the TRAVAILLEUR (EMPLOYEE) concept in the context of the outsourcing of his/her enterprise in the European Community. We start with three central concepts (TRAVAILLEUR (EMPLOYEE); TRANSFERT (TRANSFER), LICENCIEMENT(REDUNDANCY)).

a. Modeling bootstrap
From the terminological form of the term bound to a central concept, we find terms and relations describing the concept. Terms are translated into concepts and relations into roles. A natural language definition is either found in the directive or established from term occurrences in the directive. The concept definition is given, in comprehensive form, from the term occurrences found in the text with structural andor functional properties. A structural property references a link toward an ancestor concept (for instance, subsumption link). A functional property describes a domain property between concepts. The following figures

present concept definitions obtained by using the text. Only some definitions are found in the directive. The used syntax simplifies the linguistic expression to remain closer to the ontology language.

* Study of TRAVAILLEUR (EMPLOYEE)
The terminological form involves the following definition found in the directive "D-travailleur"(see Figure 2).

toute personne qui, dans l'état membre concerné est protégé en tant que travailleur dans le cadre de la législation nationale sur l'emploi.	employee shall mean any person who, in the Member State concerned, is protected as an employee under national employment law.

Fig. 2. Definition of the *travailleur* (*employee*) term

The definition of the concept EMPLOYEE has been established from the occurrences of the term *employee* in the text(see Figure 3). Three important points are stressed in this definition: (1) the European community is constituted of Member States which have national laws; (2) an employee is protected by national employment laws; (3) there is no explicit link between national laws and European laws. All these elements have to be represented in the ontology.

TRAVAILLEUR	EMPLOYEE
PROPRIÉTÉS STRUCTURELLES	STRUCTURAL PROPERTIES
est une personne	is a person
appartient à une entreprise	belongs to an enterprise
appartient à un Etat membre	belongs to a member state
PROPRIÉTÉS FONCTIONNELLES	FUNCTIONAL PROPERTIES
possédant des droits et des intérêts	having rights
ayant des représentants	having representatives
ayant des conditions de travail	having working conditions
est l'objet du transfert de son entreprise	is the object of the transfer of his/her enterprise
est protégée par une législation nationale	is protected by national employment law

Fig. 3. Concept TRAVAILLEUR(EMPLOYEE) definition

b. Modeling consolidation
During the modeling consolidation process, four kinds of activities occur:

- generalization, that is to say, the search of the ancestor concepts. It consists in studying and inserting new concepts along the bottom up axis.
- specialization, that is to say, the search of the child concepts. It consists in studying and inserting new concepts along the top-down axis in respect of differentiation principles [2].

– clustering consists in creating new concepts and regrouping identical properties.
– alignment may occur during these previous activities and requires concept and role meanings. The used ontology is CLO. The domain modeling requires the introduction of non terminological concepts for structuring the ontology. The work according to the top-down axis is dedicated to specialization and differentiation of the concepts already defined. The roles describe the functional properties and some of them restrict the inherited roles.

* Generalization of the EMPLOYEE concept

The EMPLOYEE ancestor concepts are found by working on the ascending axis. The structural properties are the linguistic expressions from which the research of the father concepts is achieved.

The structural property *is a person* leads to the study of the term *person* in the legal domain. In the directive, the term *natural or legal person* exists and is used in the *transferor* definition. Recourse to the expert helped us to create three meanings for this term. Each meaning is represented by a concept. The concepts of NATURALPERSON, LEGALPERSON and NATURALLEGALPERSON are integrated in the micro-ontology.

The linguistic syntagm *employee is a person* is translated into a subsumption link between the EMPLOYEE and NATURALPERSON.

The structural property *belong to* of the EMPLOYEE concept leads to the study of the terms *enterprise* and *member states*. This property is described by the role BELONG TO.

* The study of the functional properties

The functional properties are used for defining the roles. The terms *rights*, *representatives*, *working conditions*, *national employment law* are modeled as concepts in the micro-ontology. These concepts are bound to the EMPLOYEE concept by a role which expresses their linguistic property.

* The role definition

The created bottom-up structuring concept JURIDICALOBJECT gathers all of the juridical concepts described in a directive. It is defined by the role GOVERNED BY which takes its values in the range defined by the SOCIAL-OBJECT concept from CLO. The opposite role TO BE GOVERNED BY has been created.

Lexical properties as *covered by* or *is protected by* that constitute specializations of the property *governed by* (see figure 4) were also used in the text.

Name	Domain concept	Value concept
isGoverned	redundancy	nationalEmploymentLaw
isGoverned	socialObject	juridicalObject
governs	juridicalObject	socialObject
isProtected	employee	nationalEmploymentLaw
belongs to	employee	memberStates

Fig. 4. An excerpt of the defined generic roles

Therefore a role hierarchy was created between the rolesTO BE GOVERNED BY and TO BE PROTECTED BY.

* The alignment with CLO
An alignment with CLO is achieved for each buttom-up structuration concepts as JURIDICALOBJECT and NATURALPERSON . The NATURALPERSON concept represented in CLO defined by *Cognitive objects have a specific dependence on agentive physical objects (e.g. a natural person)* is identified to the PHYSICALPERSON concept. The DOCUMENT concept represented in CLO defined by *An information realization that realizes (at least) a text* . A subsumption link is defined between the DOCUMENT concept and the JURIDICALOBJECT concept.The figure 5 presents an excerpt of the micro-ontology "D-travailleur".
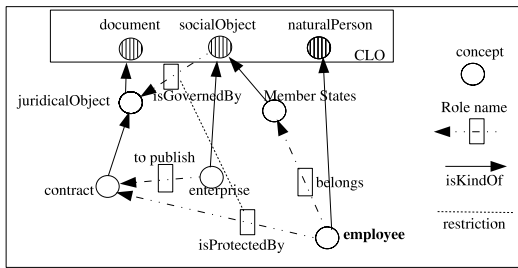

Fig. 5. An excerpt of the worker micro-ontology

"D-citoyen" Micro-ontology Building. The same method has been used to build the "D-citoyen" micro-ontology around the central concepts CITOYEN (citizen), LIBRE CIRCULATION (freely moving), ETATS MEMBRES (Member States). Figure 6 shows an excerpt of this micro-ontology.
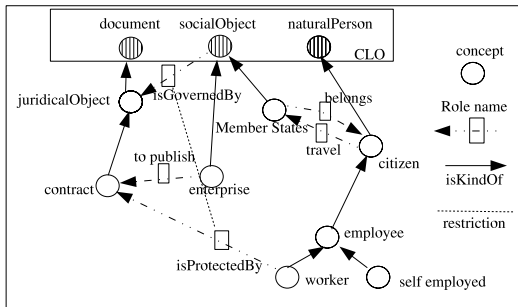

Fig. 6. An excerpt of the citizen micro-ontology

3.2   Merging Micro-ontologies: First Results

For merging union and intersection are the most prominent approaches. In the union approach, the merged ontology is the union of all entities in both source ontologies where differences in the representation of similar concepts have been resolved. In the intersection approach, the merged ontology consists only of the parts of the source ontology which overlap [8].

The aim of this work is to obtain an ontology of the European Union law around the TRAVAILLEUR(WORKER)concept. The jurists' requirements led to the building of a minimal ontology from the merging of micro-ontologies. The merging process is an interesting means both to reveal the inconsistencies between the models used to describe the concepts in the different directives and, if possible, to build a shared model of the studied concepts throughout the directives. Therefore, the intersection approach was chosen to build this ontology.

The merging process for our micro-ontologies is only beginning but some issues already emerge. Our micro-ontologies represent a model of one legal issue of the European law domain. Therefore in order to obtain a unique ontology of the TRAVAILLEUR(WORKER) concept, merging is necessary. The process was achieved with the two above mentionned micro-ontologies as shown in figures 5 and 6. The "D-citoyen" micro-ontology constitutes the reference for the merging process. The two micro-ontologies are written in the same language. Therefore, there are no mismatches at this level. The mismatches occur at the ontological and terminological levels. The merging is around the terms *travailleur (employee)* and *travailleur salarié (worker)*. These two terms refer to the same concept denoted by *travailleur salarié (worker)*. The resort to the terminological form allows the detection of the differences between the definitions associated to these concepts. The results of this study consists in moving the TRAVAILLEUR SALARIÉ (EMPLOYEE) concept from the "D-travailleur" micro-ontology to the "D-citoyen" micro-ontology under the TRAVAILLEUR (WORKER) concept. The generic role *appartient à (belongs to)* attached to the TRAVAILLEUR(EMPLOYEE) concept is removed because it exists at the citizen concept level and it is inherited. The terminological form associated to the WORKER concept contains all linked terms.

4   Comparison with Related Works

Our work is based on texts enhanced by an alignment process and a merging of the resulting micro-ontologies. This approach is clearly different from the legal ontology building presented in [5], [12]. Our approach is the only one in juridical domain that permits the traceability from texts to concepts and relations of ontology building. Thus it presents a great advantage for the management of ontology evolution and normally, it should aid to the enrichment of the constructed ontologies.

Our approach can be compared to the Ontology of Professional Legal Knowledge (OPLK) project [6]. A middle out approach is adopted and NLP tools for the linguistic study are used. By focusing on relations between concepts our

method differs from the OPLK project centered on adjectives to construct relations between concepts. The constructed ontologies differ because of the nature of the corpus. If we consider as [17] that juridical language includes several levels: the law-maker's language, the judge's language that interprets the law-maker's language and the language of the jurisprudence, our work is situated at the first level whereas the OPLK project is at the second level since the ontology is built from pragmatic knowledge. In fact, directives constitute normative texts that define objects and doers of the selected reality, and control the EU law.

5   Conclusion

The TERMINAE method gives a central role to the text and permits to establish relations between text and ontology. It differentiates the linguistical and conceptual levels. Therefore terms and concepts are distinct. A terminological network is built from these terms and their linguistical relations. The ontology concepts are built from the terms relevant for the application. The sructural and functional properties are used for elaborating a local concept description. This local concept is linked to a core ontology concepts with an alignment process. During this step, a correspondance is established between a signified of a term and a formal concept. A formalization process such as OntoSpec [13] should permit the alignment between formal concepts.

Acknowledgments

References

1. N. Aussenac-Gilles, B. Biébow, and S. Szulman. Revisiting ontology design: a methodology based on corpus analysis. In R. Dieng and O. Corby, editors, *Knowledge Engineering and Knowledge Management : Methods, Models, and Tools. Proc. of the 12th International Conference, (EKAW'2000)*, LNAI 1937, pages 172–188. Springer-Verlag, 2000.
2. B. Bachimont. Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. In J. Charlet, M. Zacklad, G. Kassel, and D. Bourigault, editors, *Ingénierie des Connaissances, évolutions récentes et nouveaux défis*, pages 305–323, Paris, 2000. Eyrolles.
3. B. Biébow and S. Szulman. TERMINAE: A linguistics-based tool for building of a domain ontology. In D. Fensel and R. Studer, editors, *Proc. of the 11th European Workshop (EKAW'99)*, LNAI 1621, pages 49–66. Springer-Verlag, 1999.
4. D. Bourigault and C. Fabre. *Approche linguistique pour l'analyse de corpus*, volume 25, pages 131–151. Université Toulouse Le Mirail, 2000.

5. J. Breuker. Constructing a legal core ontology: LRI-core. In *WONTO'2004*, 2004.

6. P. Casanovas, N. Casellas, C. Tempich, D. Varandecic, and R. Benjamins. OPKJ modeling methodology. In *LOAIT- Legal Ontologies and Artificial Intelligence Techniques*, IAAIL Workshop series, pages 121–133, 2005.

7. V. Ceausu and S. Després. Une approche mixte pour la construction d'une ressource terminologique. In *IC 2004*, pages 211–223, 2004.

8. J. de Bruijn, F. Martin-Recuerda, D. Manov, and M. Ehrig. Technical report.

9. S. Després and S. Szulman. Construction of a legal ontology from a europen community legislative text. In Thomas F. Gordon, editor, *Jurix 2004*, pages 79–88. IOS press, 2004.

10. J. Euzenat, T. Le Bach, J. Barrasa, P. Bouquet, J. De Bo, R. Dieng, M. Ehrig, M. Hauswirth, M. Jarrar, R. Lara, D. Maynard, A. Napoli, G. Stamou, H. Stuckenschmidt, P. Shvaisko, S. Tessaris, S. Van Acker, and I. Zaihrayeu. Technical report.

11. A. Gangemi, A. Prisco, M.T. Sagri, G. Steve, and D. Tiscornia. Some ontological tools to support legal regulatory compliance, with a case study. In *Workshop WORM Core*. LNCS, Springer Verlag, 2003.

12. A. Gangemi, M.-T. Sagri, and D. Tiscornia. *A Constructive Framework for Legal Ontologies*, pages 97–124. LNAI 3369. Springer, 2005.

13. G. Kassel. Une méthode de spécification semi-informelle d'ontologies. In *Actes des 13 èmes journées francophones d'Ingénierie des Connaissances (IC 2002)*, pages 75–87, 2002.

14. M. Klein. Combining and relating ontologies: an analysis of problems solutions. In A. Gomez-Perez, M. Gruninger, H. Stuckenschmidt, and M. Uschold, editors, *Workshop on Ontologies and Information Sharing, IJCAI'01*, pages 309–327, Seattle, USA, 2001.

15. D.L. McGuinness, R. Fikes, J. Rice, and S. Wilder. An environment for merging and testing large ontologies. In *Seventeenth International Conference on Principles of Knowledge Representation And Reasoning KR-2000)*, pages 483–493, 2000.

16. N. Fridman Noy and M.A. Musen. An algorithm for merging and aligning ontologies: Automation and tool support. In *Proc.of the Workshop on Ontology Management at the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, Orlando, 1999. FL: AAI Press.

17. W. Peters, D. Tiscornia, and M.T. Sagri. The structuring of legal knowledge in lois. In *LOAIT - Legal Ontology and Artificial Intelligence Techniques*.

18. URL: http://www.gate.ac.uk/projects/sekt.