Towards a Middle-out Approach for Building Legal Domain Reference Ontology

Mirna El Ghosh, H. Naja, H. Abdulrab, M. Khalil

To cite this version:


HAL Id: hal-01627322

https://hal.archives-ouvertes.fr/hal-01627322

Submitted on 1 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Towards a Middle-out Approach for Building Legal Domain Reference Ontology

M. El Ghosh, H. Naja, H. Abdulrab, and M. Khalil

Abstract—This article presents a middle-out approach to build legal domain reference ontology for a Legal Knowledge Based System (LKBS). The proposed approach is a combination of top-down and bottom-up strategies. In particular, we propose to develop legal domain reference ontology, splitted into modules or fragments, based on merging two processes: Conceptual Modeling Process, by reusing foundational ontologies (top-down strategy) and Ontology Learning Process from textual resources (bottom-up strategy).

Index Terms—Conceptual modeling, domain reference ontology, legal ontology, modularization, ontology learning.

I. INTRODUCTION

The aim of this article is to present a middle-out approach for building legal domain reference ontology for a Legal Knowledge Based System (LKBS). The domain application of this work is the Lebanese penal system. The proposed approach is a combination of top-down and bottom-up strategies. In particular, we propose to develop legal domain reference ontology splitted into modules or fragments. For this purpose, the development process is based on merging two different processes: 1) Conceptual Modeling Process and 2) Ontology Learning Process. For the Conceptual Modeling Process, we apply the top-down strategy which is based on reusing foundational and core ontologies. Meanwhile, the Ontology Learning Process is based on the extraction of knowledge from legal textual resources such as the Lebanese Penal Code.

Generally, two types of domain ontologies can be developed: reference and operational [1]. Domain reference ontology is defined as conceptual model that describes clearly and precisely the domain entities. Meanwhile, an operational ontology is a machine readable implementation version of the ontology.

In this article, we focus on developing a domain reference ontology that will be formalized and implemented as an operational ontology in the future work.

The article is organized as follows: In Section II, we start by providing a state of the art about the existing approaches for building legal ontologies. Section III discusses the related work in this domain. In Section IV, we present the architecture of our proposed middle-out approach. Finally, we conclude our work in Section V.

II. EXISTING APPROACHES FOR BUILDING LEGAL ONTOLOGIES

In this section, we study the existing approaches for building legal ontologies. The approaches are organized in two main categories: bottom-up and top-down approaches [2].

1) Bottom-up: start from the most specific concepts and build a structure by generalization [3]. In this approach, the building process of the ontology usually starts with linguistic study on existing data structures forms (documents, reports, etc.) in order to extract relevant concepts of the domain and relations among them with the semi-automatic support in document analysis (see Fig. 1). This approach results in a very high level of detail which makes it difficult to spot commonality between related concepts and increases the risk of inconsistencies [4]. Moreover, the bottom-up approach is limited by developing domain-specific or application ontologies that are not reusable. Meanwhile, it can support the refining and expanding of existing ontologies by incorporating new knowledge emerging from texts [2]. Terminae is an example of bottom-up approach to develop ontologies from textual resources using Natural Language Processing (NLP) techniques [5].

2) Top-down: start from the most generic concept and build a structure by specialization [3]. In this approach, the building process of the ontology starts by an analysis and study of relevant information sources about the given domain and then modeling the top level concepts which will be refined in next steps (see Fig. 2). This category of
approaches typically carried out manually by domain experts and lead to reusable and shareable upper level ontologies [2]. Although, it results in better control of the level of detail. However, starting at the top can result in choosing arbitrary high-level categories which lead to a risk of less stability in the model [4]. The methodology of Uschold and Gruninger [4] is an example of top-down approach for building ontologies.

Therefore, and based on what found in the literature, we conclude that each category of approaches, top-down or bottom-up, have disadvantages and advantages. For better results in building comprehensive ontologies, there is a need to combine the two categories [2]. This claim is based on the complementary of the two categories which is a fact acknowledged in the literature, specifically in the work of [3], [4] and [6].

What is a Middle-Out Approach?

A middle-out approach is a combination of top-down and bottom-up strategies [4]. In other words, it is an integration of theoretical modeling and text analysis [2]. This approach strikes a balance in terms of the level detail. Detail arises only as necessary, by specializing the basic concepts [4].

Few attempts have been made to build legal ontologies, such as LKIF-Core [7], LOIS [8], OPJK [9], and DALOS [10], by following the middle-out approach.

III. RELATED WORK

Our proposed middle-out approach tends to combine top-down and bottom-up strategies in a hybrid middle-out approach in order to build legal domain reference ontology. Similar approaches, used to build ontologies in general, are found in the literature and used in the work of [2], [6] and [11].

1) Francesconi and his colleagues 2008 [2]: proposed a methodology, for multilingual legal knowledge acquisition and modeling, which combines top-down and bottom-up strategies in the work of the DALOS KOS project.
   • Top-down: defines the conceptual language-independent structure of the legal domain under consideration on the basis of expert judgment. This structure is modeled as an ontology.
   • Bottom-up: use of (semi-)automatic NLP techniques for the knowledge acquisition from texts.

2) Saias and Quaresma [6]: proposed a middle-out, methodology to automatically create a legal ontology from a set of legal documents:
   • Top-down: Choose an already existent top-level legal ontology.
   • Bottom-up: Identification of concepts and its properties using NLP techniques.

3) Pellicer and his colleagues [11]: proposed a hybrid approach, to build domain ontology in the hydrographical domain.
   • Top-down: use of the MENTHONTOMETRY methodology that emphasizes the reuse of existing domain and upper-level ontologies.
   • Bottom-up: application of FCA techniques to output a hierarchy of concepts from the feature instances contained in the repositories used as data sources.

IV. PROPOSED MIDDLE-OUT APPROACH

Our main objective is to define an approach that combines top-down and bottom-up strategies, in a hybrid middle-out approach, in order to develop reusable modularized legal-penal domain reference ontology.

As aforementioned, the current work is developing domain reference ontology for the penal system. Reference ontology, as defined by Falbo [1], is a conceptual model that describes clearly and precisely the domain entities. This kind of ontology is built with the goal of making the best possible description of the domain (legal penal domain in our case).

In the building process of the domain reference ontology, we tend to use the modularization activity in order to split the ontology into small fragments or sub-ontologies that are reusable.

The proposed approach is a combination of top-down and bottom-up strategies in a middle-out approach. In other words, we tend to combine two different processes: (1) the Conceptual Modeling Process and (2) the Ontology Learning Process from texts.

   • Top-down: consists of the definition of the conceptual structure of the legal domain which is modeled as ontology modules (upper, core and domain). In this strategy, reusing other ontologies, that capture similar or complementary knowledge (foundational, upper-level or core-legal ontologies such as UFO [12], LKIF-Core [7]-[13]) can help in building well-founded ontology.
   • Bottom-up: consists of 1) building the legal concepts and relations among them, by using Ontology Learning methodologies and (semi-)automatic NLP techniques to extract the legal knowledge from textual resources, then 2) modeling this knowledge as a domain and, or domain-specific ontologies (modules).

Consequently, the architecture of the legal domain reference ontology (Fig. 3) is modularized in four modules which are themselves ontologies (upper, core, domain and domain specific or application).

At the highest level, the upper module represents the most general concepts and relations that cover all the domains (Fig. 4). For instance Act is a common-sense concept for all the domains.

The core module provides a definition of structural knowledge in the legal domain. Legal-Act represents any
legal act in all the legal domains (penal, civil, etc.).

4) **Use of OntoUML** (UML class diagram profile that incorporates important foundational distinctions made by UFO the Unified Foundational Ontology) as an ontologically well-founded modeling language to support the conceptual modeling process of the ontology [12].

A. **Bottom-up Strategy**

In this section, we present the bottom-up part of the proposed middle-out approach which is based on the Ontology Learning Process from textual resources. This strategy tends to extract concepts and relations from texts using NLP techniques in order to build the domain-specific and part of the domain modules.

The context of this work is the Lebanese penal code which is used as the textual resource for term extraction and knowledge acquisition.

Texts resources, usually, contain unstructured data not meaningful for a computer system. Different Natural Processing techniques (such as TreeTagger, GATE, YaTeA, etc.) and (semi-)automatic textual analysis or ontology learning tools (such as OntoGen, ASIUM, Text-To-Onto, Text2Onto and TERMINAE) could be used to facilitate the knowledge acquisition and to support the building process of an ontology from natural language texts by reducing the development time and costs. This process referred to as Ontology Learning [17].

In the Literature, there is a diversity of methodologies for Ontology Learning Process. We quote the work of [2], [18], [19].

For this part of the middle-out approach, we propose the use of the methodology Terminae [18]. In Terminae, the knowledge modeling is based on the knowledge extraction or terminology extraction from textual resources. This methodology is widely used in projects for the legal domain such as Legal ontology for a European community legislative text [20] and a legislative project based on the creation of legal micro-ontologies [5].

In addition, Terminae is offered with a tool (TERMINAE) to help the designer to apply the methodology semi-automatically, to extract the legal knowledge from the corpus, and to build the domain or domain-specific modules of the ontology.

For the application of Terminae, we defined a legal corpus which is an excerpt of the Lebanese penal code (300 articles). After using NLP tools and linguistics techniques, such as YaTeA, to extract the legal terms and their frequency from the corpus, we obtained a list of 1200 terms (single and multi-word terms).

Although, using TERMINAE, the designer can evaluate the acquired terms with the help of legal expert and link them to the concepts they express (see Table I).

<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infraction</td>
<td>60</td>
<td>Infraction</td>
</tr>
<tr>
<td>Crime</td>
<td>35</td>
<td>Crime</td>
</tr>
<tr>
<td>Offense</td>
<td>11</td>
<td>Offense</td>
</tr>
<tr>
<td>Violation</td>
<td>10</td>
<td>Violation</td>
</tr>
</tbody>
</table>

Some relationships hold between the concepts extracted
Violation

bottom-up

concepts of the domain module can be defined using the section IV and presented in Fig. 1 and Fig. 2, that part of the (see Fig. 5).

ontologies and ontology patterns [1]. In our work, we tend to existing domain ontologies, core ontologies, foundational development process [21].

Meanwhile, the concept Infraction, extracted from the corpus using the methodology Terminae and the tool TERMINAE, is located in the domain module.

Therefore, by applying the bottom-up part of the middle-out approach, the designer will obtain a list of legal domain-specific and domain concepts and relations among them. These concepts and relations tend to build the domain-specific and part of the domain modules.

B. Top-down Strategy

In this context, we clarify the idea aforementioned in section IV and presented in Fig. 1 and Fig. 2, that part of the concepts of the domain module can be defined using the bottom-up strategy. The concepts Crime, Offense and Violation are located in the domain-specific module. Meanwhile, the concept Infraction, extracted from the corpus using the methodology Terminae and the tool TERMINAE, is located in the domain module.

TABLE II: SOME RECENT TOP-DOWN APPROACHES

<table>
<thead>
<tr>
<th>Approach</th>
<th>Development Process</th>
</tr>
</thead>
</table>

At this point, we can clarify that our work is applied in the legal domain and more specifically in the penal system. In addition to this, the purpose of building the legal domain reference ontology is to make this reference ontology reusable as much as possible for LKBS and reasoning systems in the legal domain. Therefore, after studying the main concepts of the legal domain and more specifically in the penal system. In addition to this, the purpose of building the legal domain reference ontology is to make this reference ontology reusable as much as possible for LKBS and reasoning systems in the legal domain. Therefore, after studying the main concepts of some foundational ontologies, we found that UFO is the most convenient foundational ontology to use in our work. Specifically, we can cite some basic concepts in UFO that will be considered as foundations for the conceptual modeling of the legal penal domain reference ontology: Agent, Intentional_Moment, Action, Event, Normative_Description, Goal, and Situation.

2) Reuse and extend the concepts and relations of the defined foundational ontology in order to identify the main concepts and relations of the upper module. This step can be processed by analogy, specialization or analogical analysis [1].

3) Define legal-core ontology. Although it is a difficult task to choose the right legal-core ontology. We cite some legal-core ontologies found in the literature: LKIF-Core, LRI-Core, and FOLAW. Actually, LKIF-Core is the most recent legal-core ontology and contains essential legal concepts such as Medium, Document, Legal_Source, Legal_Document, and Code.

4) Reuse and extend the concepts and relations of the defined legal-core ontology by specialization, in order to define the concepts and relations of the core module.

5) Specialize the concepts of the core module to define new concepts and relations for the domain module.

Therefore, by applying these steps, the designer will obtain a list of concepts and relations organized in modules based on their level (general, legal or penal).

Since we decided to start the conceptual modeling process,
of the legal domain reference ontology, based on the foundational ontology UFO and using the modeling language ONTOUML, the representation and classification of the concepts and relations of all ontology modules (upper, core, domain and domain-specific) will be according to the types defined in ONTOUML such as Kind and Subkind. The Fig. 7 shows the classification of LKIF-Core concepts according to UFO concepts kind and Subkind.

Note that Kind represents a rigid concept, i.e., a class that applies necessarily to its instances. Subkind is a subtype of kind and is also rigid.

V. CONCLUSION

In this article, we have presented an overview on approaches for building legal ontologies and a proposed middle-out approach for the development of legal domain reference ontology, specifically for the penal system. The main purpose for building this legal domain reference ontology is to obtain a reusable ontology for LKBS and reasoning systems in the legal domain. For this reason, the main work was modularizing this ontology, and splitting it into small fragments or modules that are reusable. In addition to this, we focused on reusing foundational ontologies and legal-core ontologies in the conceptual modeling process of the reference ontology. Although, to make this ontology the most possible representative and reflective to the real penal domain, we used Ontology Learning Process techniques to extract the knowledge, concepts and relations, from textual resources, such as the penal code.

The proposed middle-out approach represents the merging process of the two different processes, the conceptual modeling (top-down strategy) and the Ontology Learning Process (bottom-up strategy).

This approach raises several perspectives such as:

- Define the alignment, or merging, process of the different modules (upper, core, domain and domain-specific).
- The possibility to use ontology patterns during the ontology conceptual modeling activity.

REFERENCES


Habib Abdulrah was born in 1956 in Yemen, Aden. He is a computer scientist. He obtained a masters degree in informatics from the University of Paris 6 in 1983, followed by a Ph.D. from the University of Rouen in 1987. He is currently a professor in the Mathematical and Software Engineering Department at Rouen and also at INSA de Rouen. He published several articles.

Hala Naja was born on February 22, 1970 in Tripoli, Lebanon. She obtained her DEA in 1993, computer science, University of Henri Poincare, France, followed by the PhD in 1997 University of Henri Poincare, France.

Currently, she is a full associate professor of computer science and the coordinator of master in software engineering at the Computer Science Department in the Faculty of Science, Lebanese University. Her research concerns Software engineering, software architecture and their evolution. Recently she is interested in ontologies and reasoning. Dr. Naja’s publications “Overview of an approach describing multi-views/multi-abstraction levels software architecture,” ENASE, 2013, pp. 140-148, etc.

Mohamad Khalil was born in 1973 in Akkar Atika, Lebanon. He obtained an engineering degree in electrical and electricity from the Lebanese University, Faculty of Engineering, Tripoli, Lebanon in 1995. He received the DEA in biomedical engineering from the University of Technology of Compiegne (UTC) in France in 1996. He received his Ph.D. from the University of Technology of Troyes and University of Technology of Compiegne in France in 1999. He received his HDR from UTC in 2006.

He is currently professor at the Lebanese University, Faculty of Engineering, Tripoli, and researcher in the Lebanese University. His current interests are the signal and image processing problems: detection, classification, analysis, representation and modeling of non stationary signals, with application to biomedical signals and images. Prof. Khalil published several articles.