Ontology Learning Process as a Bottom-up Strategy for Building Domain-specific Ontology from Legal Texts

Mirna El Ghosh[1], Hala Naja[2], Habib Abdulrab[1] and Mohamad Khalil[3]
[1]*INSA, Rouen, France*
[2]*Lebanese University, Faculty of Sciences, Tripoli, Lebanon*
[3]*Lebanese University, Faculty of Engineering, Tripoli, Lebanon*

Keywords: Ontology Learning, Semi-automatic Extraction, Natural Language Processing, Legal Ontologies, Domain-specific Ontologies.

Abstract: The objective of this paper is to present the role of *Ontology Learning Process* in supporting an ontology engineer for creating and maintaining ontologies from textual resources. The knowledge structures that interest us are legal *domain-specific* ontologies. We will use these ontologies to build legal domain ontology for a Lebanese legal knowledge based system. The domain application of this work is the Lebanese criminal system. Ontologies can be learnt from various sources, such as databases, structured and unstructured documents. Here, the focus is on the acquisition of ontologies from unstructured text, provided as input. In this work, the *Ontology Learning Process* represents a knowledge extraction phase using Natural Language Processing techniques. The resulted ontology is considered as inexpressive ontology. There is a need to reengineer it in order to build a complete, correct and more expressive *domain-specific* ontology.

1 INTRODUCTION

It is commonly known that the knowledge of the legal domain is expressed and conveyed in texts using domain-specific terminology. However, this terminology does not provide a well-defined structure to be used by machines for reasoning tasks. Meanwhile, the extracting and mining of this terminology will lead to a certain domain representation model such as ontology (Mädche, 2000). Ontology is defined as a conceptualization of a domain into a human understandable, machine-readable format consisting of *entities*, *attributes*, *relationships* and *axioms* (Guarino, 1995). This definition imposes that the concepts and relations among them have to be explicitly represented and expressed using formal language such as Web Ontology Language (OWL). This formal structure representation leads to specify axioms for reasoning, in order to define constraints in ontologies (Wong, 2009). Building and maintaining ontologies manually remains a resource-intensive, time consuming and costly task. This is due to the difficulty in capturing knowledge, also known as the "knowledge acquisition bottleneck". Even with some reuse of *Core* or *Upper* ontologies. Therefore, there is a need to automatic or semi-automatic techniques that support the building process. These techniques have become to be known as *Ontology Learning* (OL) (Cimiano, 2004). OL has the potential to reduce the cost of creating and maintaining ontologies using semi-automatic methods and tools. Actually, we motivate to develop legal domain ontology for the Lebanese criminal domain. In a previous work (El Ghosh, 2016), a middle-out approach is proposed for building this ontology for a legal knowledge based system that performs reasoning and information retrieval tasks (Figure 1). Accordingly, we proposed to modularize the legal domain ontology into four modules or ontologies: *upper*, *core*, *domain* and *domain-specific*. The *upper* module represents the most general concepts and relations that cover all the domains (such as Agent, Act and Action). The *core* module provides a definition of structural knowledge in the legal domain. For instance, concepts, such as Legal_Source, Legal_Act and Legal_Document, are common for all the legal fields (criminal, civil, etc.). The concepts of the *domain* module, in turn, such as Offence, Infraction and Offender, describe the conceptualization of the criminal domain. Finally, in the *domain-specific*

El Ghosh M., Naja H., Abdulrab H. and Khalil M.
Ontology Learning Process as a Bottom-up Strategy for Building Domain-specific Ontology from Legal Texts.
DOI: 10.5220/0006188004730480
In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*, pages 473-480
ISBN: 978-989-758-220-2

module, we learn the knowledge of the Lebanese criminal system from textual resources such as the criminal code. Furthermore, an alignment process will be applied to complete the global ontology by linking the concepts of the different modules. In order to develop the different modules, two different strategies are applied (*top-down* and *bottom-up*). The *top-down* represents the conceptual modeling process based on reusing foundational and core ontologies (El Ghosh, 2016). Meanwhile, the ontology learning process from textual resources is depicted by the *bottom-up* strategy that aims to develop the *domain-specific* ontology module. *Domain-specific* ontologies specify formally concepts and relations of a specific subject domain (Hatala, 2012). They cannot be reused unlike other kinds of ontologies (upper and core).

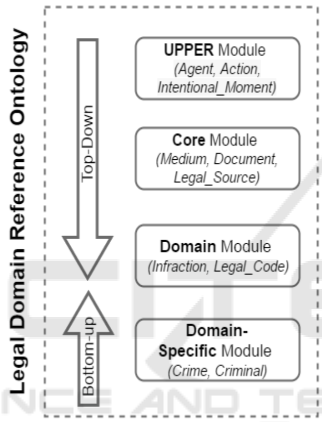

Figure 1: Middle-out approach for building modularized ontology.

What is important that these ontologies are useful in systems involved with artificial reasoning and information retrieval. In this context, the OL process from unstructured legal documents could be useful for building the criminal *domain-specific* ontology. Meanwhile, the main obstacle that exists is to reduce the efforts required for creating the ontology by defining a convenient semi-automatic development process and ontology learning tool. In order to achieve the goal, we started by discussing the ontology learning from unstructured texts in section II. In section III, we overviewed existent ontology learning methods and tools. The experimental work is presented in section 4. The section 5 discusses similar works. We finished by section 6 for the discussion and section 7 for the conclusion.

2 ONTOLOGY LEARNING FROM UNSTRUCTURED TEXT

The term *Ontology Learning* (OL) was introduced in (Mädche, 2005) and is considered as an important task in Artificial Intelligence, Semantic Web and Knowledge Management. It is the dynamic process of building ontologies. OL is a data model that represents a set of concepts and relations within a domain (Yang, 2008). More specifically, OL is considered as a subtask of Information Extraction (IE), which is a type of Information Retrieval (IR) (Rogger, 2010). The main purpose of OL process is to apply methods from various fields such as linguistic analysis, machine learning, knowledge acquisition, statistics and information retrieval in order to extract knowledge and support the construction of ontologies. This dynamic process, depicted in the Figure 2, takes as input implicit and unstructured knowledge and produces as output explicit structured knowledge (Cimiano 2005). Generally, OL is a semi-automatic process where the ontology engineer and the domain expert can be involved to achieve better results (Rogger, 2010). Thus, the techniques used in the ontology development process will be under their supervision. Their expertise and background knowledge helps in verifying the obtained information and decide the valuable information.

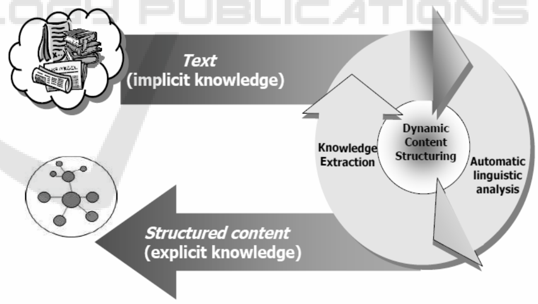

Figure 2: The dynamic process of ontology learning, (Buitelaar, 2005).

2.1 Input

As aforementioned, ontologies can be learnt, by applying the OL process, from various sources of data types: structured (such as databases), semi-structured (e.g. XML) and unstructured textual documents. The domain application of this work is the Lebanese criminal code which is an unstructured text resource. This type of resources is the most available format as input for ontology learning processes. They reflect mostly the domain knowledge for which the user is building the ontology. In addition, they describe the terminology, concepts and conceptual structures of the given domain. However, some authors, such as (Rogger, 2010), consider that processing unstructured data is the most complicated problem because most of the knowledge is implicit and allows conceptualizing it by different people in different manner. Specifically, in the legal domain, the implicit knowledge of the natural language is one of the main obstacles to progress in the field of artificial intelligence and law (McCarty, 2007).

2.2 Output

Ontology learning from text is the process of deriving concepts, relations and axioms from textual resources to build ontologies. The main output of the OL process is a structured content represented in an explicit formal way. For (Cimiano, 2004), the tasks in ontology learning from text are organized in a set of layers (Figure 3). These tasks aim at returning six main outputs: *terms*, synonyms, *concepts*, *taxonomic relations*, *non-taxonomic relations* and *axioms*. These outputs represent the main elements of ontology.

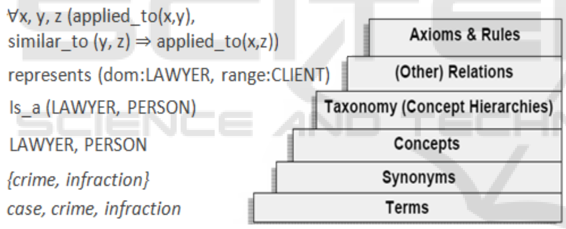

Figure 3: Ontology learning from text, layer cake (Buitelaar, 2005).

Terms are the most basic building blocks in ontology learning (Wong, 2009). Concepts can be abstract or concrete, real or fictitious. Concept hierarchies or taxonomies are crucial for any knowledge based system (Cimiano, 2005). Non-taxonomic or non-hierarchical relations represent the interactions between concepts (e.g. meronymy, thematic roles, attributes, possession and causality) (Wong, 2009). Finally, the axioms are defined as propositions or sentences that are always taken as true. Axioms act as a starting point for deducing other truth, verifying correctness of existing ontological elements and defining constraints (Wong, 2009).

3 ONTOLOGY LEARNING METHODS AND TOOLS

There are many works in the literature that deal with ontology learning from textual resources. The focus of this paper is to discuss and evaluate existent methods and tools to develop (semi-)automatically text-based domain ontologies. Furthermore, we will define a (semi-) automatic approach for building our legal *domain-specific* ontology.

3.1 Methods

In order to obtain high-quality ontologies, the development process has to be driven by a methodology (Hatala, 2012). In this section, we discuss briefly the most known ontology learning methodologies from textual resources. In the work of (Sabou, 2005), the ontology learning process is based on three major tasks: term extraction, conceptualization and enrichment. For (Mädche, 2005), the OL process is composed of four different phases: extract concepts, prune, refine and Import or reuse. In other studies, such as (Mazari, 2012) and (Ge, 2012), the ontology learning tasks are resumed in three: documents preprocessing, concepts extraction and relations discovery. Actually, these tasks discover only taxonomic relations (parent-child, hyponymy (is-a) and meronymy (part-of)). However, some authors such as (Novelli, 2012), (Balakrishna, 2010) and (Serra, 2012) propose methods to solve the problem of learning non-taxonomic relations of ontologies from text. In the legal domain, most of the methodologies focus on concepts extraction as a main step of the ontology development process (Lenci, 2009). The approach of (Walter, 2006) is based on the exploitation of the frequency of definitions in legal texts.

3.2 Tools

In the literature, a long list of ontology learning tools has been proposed. The existent tools differ according to input data types, output formats and mainly the methods and algorithms used in order to extract the ontological structures. The main goal of using ontology learning tools is to reduce the time and cost of ontology development process. In this section, we discuss mainly the existent ontology learning tools from unstructured textual resources. ***Terminae*** is a method and tool that generates standard OWL ontologies (Biebow, 1999). *Terminae* integrates linguistic and knowledge engineering tools to guide the knowledge acquisition from texts and to build terminological and ontological models. ***Text2Onto***, successor of *Text-to-Onto* (Mädche, 2001), is a data-driven, ontology learning tool that supports automatic development of ontologies from textual documents (Cimiano, 2005). *Text2Onto* is built upon the *GATE* [1] framework. Accordingly, *Text2Onto* implements linguistic processing and machine learning statistical techniques to extract domain concepts and relations. This tool features also algorithms for generating concepts, taxonomic and non-taxonomic relations. ***OntoGen*** is a semi-automatic and data-driven ontology editor that helps the users to build ontologies by suggesting concepts and relations. This system integrates machine learning and text mining algorithms. *OntoGen* offers two main features: concept suggestion and naming and ontology and concept visualization. ***T2K***[2] extracts domain–specific information from texts using natural language processing techniques in three main phases: preprocess text and extract terms, form concepts using POS patterns and relations or knowledge organization (Dell'Orletta, 2014). ***CRCTOL*** is Concept-Relation-Concept tuple-based ontology learning system from domain-specific text documents. The tool adapts a full text parsing technique and incorporates both statistical and lexico-syntactic methods (Jiang, 2005). We conclude that most of these tools rely on linguistic and statistic methods to learn ontologies. The focus is on extracting concepts and taxonomies. Thus, we need to learn more semantic relations and axioms.

Table 1: Summary of ontology learning tools.

Tool	Elements extracted	Techniques
Terminae *(2005)*	Terms, synonyms, concepts, taxonomies, non-taxonomic relations	Linguistic and knowledge engineering
Text2Onto *(2005)*	Terms, synonyms concepts, taxonomies, non-taxonomic relations, instances	linguistic processing statistical text analysis machine learning association rules
OntoGen *(2006)*	Terms, concepts, taxonomies	Machine learning text mining
T2K *(2008)*	Terms, concepts ,taxonomies	statistical text analysis and machine learning
CRCTOL *(2010)*	Concepts, taxonomies, non-taxonomic relations	Statistical lexico-syntactic association rules

[1] https://gate.ac.uk/

4 OUR WORK

Even after a comprehensive literature review, we found a difficulty to define a complete approach or tool that can totally extract domain-specific ontologies from textual resources. This is due to two reasons. First, we could not find a complete (semi-)automatic tool or approach that carries the ontology development process. Second, there is no guarantee that the (semi-)automatically generated ontology is correct and precise enough to characterize the domain in question (Rudolph, 2007). Since the focus of the current research is mainly on extracting the elements of a criminal *domain-specific* ontology from textual resources, using an existent semi-automatic ontology learning can help to extract an OWL ontology including the basic elements (concepts, taxonomies, relations and disjointness axioms). Meanwhile, and based on what is found in the literature, incomplete and not satisfactory results are expected. For this reason, the intervention of ontology engineer and legal expert during the ontology learning process is required in order to supervise the work and to verify the obtained information. Furthermore, a reengineering methodology is needed in order to enhance the results by transforming the resulted ontology into a new more correct, complete and expressive ontology. The general idea of the reengineering approach is depicted in figure 4. In the current work, mainly the ontology learning process, from texts, is discussed. The reengineering phase will be the study of further works.

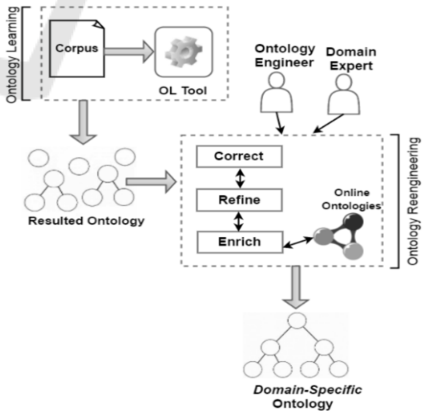

Figure 4: Reengineering phase for updating domain-specific ontology.

In this section, we introduce the main components of the ontology learning process used in the preparation and execution of the criminal *domain-specific* ontology.

4.1 Material Selection

Actually, the *domain-specific* ontology that we aim to build, using (semi-)automatic ontology learning tool guided by an approach, represents the *domain-specific* module in the modularized legal domain ontology. The context of interest is the Lebanese criminal system. The domain related material is the Lebanese penal code that consists of legal natural language texts. The Lebanese penal code contains the general penal laws of Lebanon. First enacted in 1943 and it remains in effect today. It is translated to French and English versions. Concerning the structure of the code, it is divided into two main books composed of 770 articles (Figure 5).

LEBANESE CRIMINAL CODE

Chapter IV
Responsibility

Part I
Responsible persons

Section II
Criminal complicity

Subsection 1 - The perpetrator

Article 212 – The perpetrator of an offence is anyone who brings into being the constituent elements of an offence or who contributes directly to its commission.

Figure 5: Excerpt of the Lebanese criminal code.

4.2 Tool Selection

After exploring the literature and collecting the state-of-the-art for the most frequently used ontology learning tools, we met some access difficulties in our experimentations. In fact, three of the tools were publicly available on the internet to download and install: *Terminae*, *OntoGen* and *Text2Onto*. In this section, we discuss briefly the usability of each tool. Concerning the input type, all the tools accept simple text files (.txt), *Text2Onto* and *Terminae* accept also PDF files (.pdf). For *OntoGen*, there are additional input file types that need to be pre-processed, such as Named Line-Document and Bag of Words. *Terminae* and *OntoGen* need preprocessing efforts. Starting with *Terminae* where the linguistic tool extract terms automatically from the corpus based on their occurrences. Meanwhile, the rest of the steps are processed manually which is too resource demanding and too time consuming. For this reason, this tool is discarded. Furthermore, we face some difficulties while using *OntoGen*. We could not control the system that generates sequences of terms that are not well related. In addition to this, the suggestion of concepts is limited to single-word terms, proposed only from the input documents (no external resources), and the relations extraction is limited as well to taxonomic. Meanwhile, *OntoGen* provide a visualization and exploration of concepts only and not of the whole ontology. OntoGen is discarded too. We finished our experiments by *Text2Onto*. According to (Gherasim, 2013), *Text2Onto* is an ontology learning tool that covers the entire process of extracting OWL ontologies. Furthermore, it provides a long list of proposed concepts and relationships along with their weights in a tabular form. Meanwhile, *Text2Onto* does not have any mechanism to filter the concepts irrelevant to goal (Hatala, 2012). The user input is limited to removing concepts and relationships extracted from the supplied course. In *Text2Onto*, the visualization of the structure of the resulted ontology is missing. Regarding the external resources, *Text2Onto* uses *WordNet* to improve and enrich the algorithms of pattern-based relation extraction. However, some authors found that *WordNet* lacks the richness of named relations (Fouad, 2015). For this reason, they decided to use online ontologies as an alternative to *WordNet*. Regarding the limitations of *Text2Onto*, this tool still answers the main requirements of our work: automatic extraction, usability, scalability, and reusability. Based on this selection, we proposed to apply a reengineering phase that consists of evaluating the ontology extracted using *Text2Onto*, correcting the detected errors, refine the ontology model and finish by enrich the semantic relations and axioms. We will study deeply this point in further works.

Table 2: List of experimented tools.

Tool	Terminae	OntoGen	Text2Onto
User Input	Add, remove, modify	Add, remove, modify	Remove
Visuali-zation	Not available	Concepts	Not available
External Resources	Not available	Not available	WordNet

4.3 Ontology Extraction Process

In this section, we present the main phases of the criminal *domain-specific* ontology extraction process using *Text2Onto*. Actually, the process is composed of two main phases: linguistic preprocessing and extraction of modeling primitives. In the following, we discuss briefly each phase and the algorithms used to achieve the resulted *domain-specific* ontology.

4.3.1 Preprocessing

The purpose of the preprocessing phase (Figure 6) is to prepare the corpus and remove the ambiguity by filtering out worthless symbols and words, in order to extract meaningful textual content from the input documents. In *Text2Onto*, there is a combination of machine learning approaches with basic linguistic processing such as tokenization or lemmatizing and shallow parsing (Cimiano, 2005). In addition to this, *Text2Onto* benefits from GATE by the integration of JAPE that provides finite state transduction over annotations based on regular expressions (Mädche, 2001).

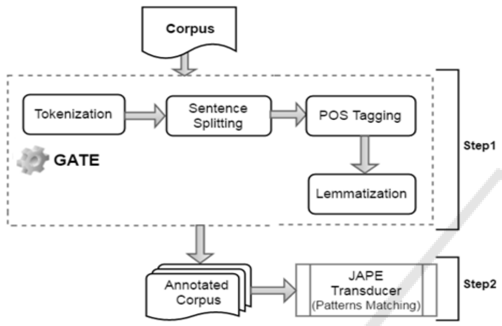

Figure 6: Preprocessing phase in Text2onto.

4.3.2 Extraction of Modelling Primitives

In this section, we describe briefly the extraction phase of the ontology modelling primitives. For this purpose, *Text2Onto* implements series of algorithms. Five main modeling primitives are considered in this tool: concepts, instances, taxonomies, general relations and disjoint axioms. In this section, the extraction process of each primitive is discussed briefly. For extracting concepts, three algorithms are implemented. Based on experiments, TFIDFConceptExtraction algorithm is selected. 486 single and multi-word concepts are extracted such as: Probation, Criminal, Crime, Term Penalty and Violence. Concerning the taxonomies (subclass-of relations), *Text2Onto* provides three algorithms to classify concepts based on Vertical Relations, WordNet, and Patterns. For better results, the three algorithms are combined.

Table 3: Excerpt of the hierarchies extracted using Text2Onto.

Domain	Range
Divorcee	Wife
Offender	Person
Death penalty	Penalty

Regarding the Instances, *Text2Onto* identifies proper nouns as instances. Technically, it filters the terms tagged as *Instance* from the GATE result. Long list of instances are extracted such as Lebanon, April and Friday. *Text2Onto* relies on SubcatRelationExtraction algorithm to extract general relations. This algorithm uses syntactic pattern matching technique to extract general relations.

Table 4: Excerpt of general hierarchies extracted using Text2Onto.

Label	Domain	Range
involve	Residence	Placement
require	Activity	License
exceed	Offence	Bound

For the disjointness axioms, they are extracted in Text2Onto based on lexico-syntactic patterns.

Table 5: Excerpt of disjoint axioms extracted using Text2Onto.

Domain	Range	Confidence
Measure	Penalty	0.013
Felony	Disposal	0.013
Person	Association	0.06

4.3.3 Ontology Visualization

After applying the algorithms of *Text2Onto*, the results are exported, as output, in OWL format. Subsequently, we have looked for an ontology visualization tool to visualize the resulted ontology. Different tools are tested such as OWLViz[2], a plug-in for Protégé, and COE cmap tool[3], and OWLGrEd[4]. The resulted ontology is visualized correctly in OWLGrEd (Figure 7).

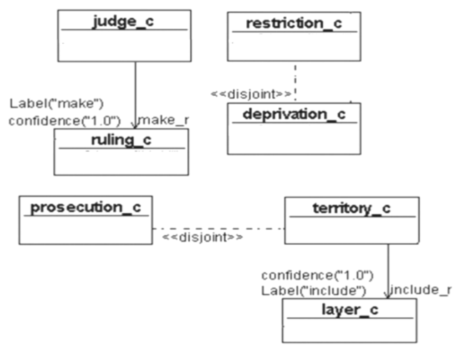

Figure 7: Ontology visualization using OWLGrEd.

[2]http://protegewiki.stanford.edu/wiki/OWLViz
[3]http://coe.ihmc.us/
[4]http://owlgred.lumii.lv/

5 RELATED WORK

There are many works in the literature that proposed the (semi-)automatic building of ontologies from textual resources using ontology learning methods and tools. The most related works are (Francesconi, 2010) in the legal domain and (Ortiz, 2007) in the political domain. In the work of (Francesconi, 2010), the authors have used two different tools for term extraction: GATE for English texts and T2K for Italian. The rest of the phases, such as evaluation of terms and link them to concepts, extraction of lexical relations were processed under the supervision of ontology engineers and domain experts. For the work of (Ortiz, 2007), the authors applied *Text2Onto*for creating domain ontology from texts. They concentrated mainly in their study on concepts extraction. In addition to this, the authors proposed a reengineering methodology based mainly on reusing online ontologies. What differs our work first is the domain application, which is the Lebanese criminal code. The context of the code is composed of legal norms written in legal language. Secondly, we used *Text2Onto* to extract all the essential elements of a *domain-specific* ontology. Finally, we expect to build an expressive domain-specific ontology for reasoning system, which is difficult using only an ontology learning tool, for this reason we have proposed a reengineering approach, based not only on online ontologies, to correct the errors and to enrich the extracted ontology with relations and axioms in order to make it more expressive.

6 DISCUSSION

The aim of this paper if to extract domain-specific ontology elements from texts using ontology learning tool. *Text2Onto* is selected for this purpose. The tool applies an automatic extraction process based on list of algorithms and NLP techniques using GATE applications. In addition to this, the results can be exported as OWL ontology ready to edit and update in ontology editor frameworks such as Protégé. After applying list of algorithms to extract the elements of the *domain-specific* ontology, we obtained some results to discuss. Starting with concepts, the tool extracted 486, single and multi-word, concepts. The domain expert filtered the list and removed the errors. We can resume the identified errors in some examples. Some verbs like *stay*, *incur* and *abort* were identified as concepts by *Text2Onto*. Some *domain-specific* concepts were identified as instances such as *Confiscation*, *Detainee* and *Terrorist*. For the instances, the extraction is limited because of the corpus quality. Actually, the experiment is based on criminal code written in legal language, which is authoritative and contains legal speech acts accompanied by rituals of various types. *Text2Onto* identified only 20 semantic relations and 86 disjoint axioms. A reengineering phase is needed to enrich the extracted ontology. From this perspective, the reengineering methodology is proposed to correct, enrich and refine the resulted ontology and to build correct, complete and more expressive *domain-specific* ontology.

7 CONCLUSIONS

In this paper, we have briefly described the field of ontology learning from textual resources as a bottom-up approach for building a domain-specific ontology for the criminal law. The mechanism of ontology learning process from unstructured text was identified. Furthermore, we have presented an overview of the existent ontology learning methods and tools. We also discussed our work followed by a summarizing comparison of the ontology learning tools used in our experiments. Based on the experiments, *Text2Onto* is selected as a tool for the ontology learning process. In fact, this tool answers the main requirements of the study. Using *Text2Onto*, the main elements of the *domain-specific* ontology are extracted (concepts, taxonomies, relations and axioms). The results were essentials, but inexpressive. A reengineering process is needed to build a more expressive ontology.

ACKNOWLEDGEMENTS

This project has been funded with support from the European Union with the European Regional Development Fund (ERDF), the National Support from the National Council for Scientific Research in Lebanon (CNRS) and Lebanese University

REFERENCES

Mädche, A. and Staab, S., 2000. “Mining ontologies from text,” Proceeding of *EKAW '00*, pp. 189-202.

Guarino, N. and Giaretta, P., 1995. "Ontologies and Knowledge Bases: Towards a Terminological Clarification," Proceeding of *Toward Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, pp.25-32.

Wong, W., Y., 2009. "Learning Lightweight Ontologies from Text across Different Domains using the Web as Background Knowledge," *PhD thesis*, University of Western Australia, School of Computer Science and Software Engineering.

Cimiano, P. Mädche, A, Staab, S., and Volker, J., 2004. *Ontology Learning. Handbook on Ontologies*, Springer.

Hatala, M., Gašević, D., Siadaty, M., Jovanovic, J. and Torniai, C., 2010. "Ontology Extraction Tools: An Empirical Study with Educators", *IEEE Transactions on Learning Technologies*, vol. 5, no. 3, pp. 275-289.

El Ghosh, M., Naja, H., Abdulrab, H. and Khalil, M., 2016. "Towards a Middle-out Approach for Building Legal Domain Reference Ontology," *International Journal of Knowledge Engineering*, vol. 2, no. 3, pp. 109-114, September.

Gašević, D., Jovanovic, J. and Devedzić, V., 2007. "Ontology-based annotation of learning object content," *Interactive Learning Environments*, vol. 15, pp. 1-26.

Serra, I., Girardi, R. and Novais, P., 2010. "Reviewing the Problem of Learning Non-Taxonomic Relationships of Ontologies from Text," *International Journal of Semantic Computing*, vol. 6, December.

Mädche, A. Staab, S., 2005. Ontology learning for the semantic web, *IEEE Intelligent Systems*, vol 16, pp. 72-79.

Yang, H. and Callan, J., 2008. "Human-Guided Ontology Learning," Proceeding of *HCIR*, pp. 26-29.

Rogger, M. and Thaler, S., 2010, "Ontology Learning," *Seminar paper, Applied Ontology Engineering*, Leopold–Franzens–University Innsbruck.

Cimiano, P., Hotho, A. and Staab, S., 2005. "Learning concept hierarchies from text corpora using formal concept analysis" *Journal of Artificial Intelligence research*, vol. 24, pp. 305-339.

Buitelaar, P., Cimiano, P. and Magnini, P., 2005. "Ontology Learning from Text: Methods, Evaluation and Applications", Vol. 123, IOS Press, July.

Sabou, M, 2005. "Visual Support for Ontology Learning: an Experience Report," Proceeding of *IV05*, London.

Mazari, C., Aliane, H. and Alimazighi, Z., 2012. "Automatic Construction of Ontology from Arabic Texts," Proceeding of *ICWIT*, pp.193-202.

Ge, J., Li, Z. and Li, T., 2012. "A Novel Chinese Domain Ontology Construction Method for Petroleum Exploration Information," Journal of Computers, Vol 7, No 6, pp. 1445-1452.

Novelli, A. and Oliveira, J., 2012. "Simple method for ontology automatic extraction from documents," *International Journal of Advanced Computer Science and Applications*, v. 3, p.44-51.

Balakrishna, M., Moldovan, D., Tatu, M. and Olteanu, M., 2010."Semi-Automatic Domain Ontology Creation from Text Resources," Proceeding of *LREC'10*.

Lenci, A., Montemagni, S., Pirrelli, V. and Venturi, G., 2009. Ontology learning from Italian legal texts, in Law, Ontologies and the Semantic Web – Channelling the Legal Information Flood, *Frontiers in Artificial Intelligence and Applications*, Springer, Volume 188, pages 75–94.

Walter, S. and Pinkal, M., 2006. "Automatic extraction of definitions from German court decisions," Proceeding of *Workshop on Information Extraction Beyond The Document*, pp. 20–28, Sydney, Australia.

Biebow, B. and Szulman, S., 1999. "TERMINAE: A linguistics-based tool for the building of a domain ontology," in *Proc. EKAW '99* - Proceeding of *the 11th European Workshop on Knowledge Acquisition, Modeling, and Management*, Berlin, Germany, pp. 49–66.

Dell'Orletta, F., Venturi, G., Cimiano, A. and Montemagni, S., 2014. "T2K²: a System for Automatically Extracting and Organizing Knowledge from Texts," proceeding of *LREC*, pp. 26-31, Iceland.

Jiang, X. and Tan, A. H., 2005. " Mining Ontological Knowledge from Domain-Specific Text Documents," proceeding of *IEEE International Conference on Data Mining*, USA.

Cimiano, P. and Volker, J., 2005. "Text2Onto," in Natural Language Processing and Information Systems, ed: Springer, pp. 227-238.

Mädche, A. and Volz, R., 2001. "The Text-To-Onto ontology extraction and maintenance system," proceeding of *1st International Conference on Data Mining*.

Rudolph, S., Volker, J. and Hitzler, P., 2007. "Supporting lexical ontology learning by relational exploration," Proceeding of *ICCS*, pages 488–491.

Fouad, Z., Grigoris, A., Mathieu, A., Giorgos, F., Haridimos, K. and Enrico, M., 2015. "Ontology evolution: a process-centric survey," *The Knowledge Engineering Review*, pp. 45–75.

Gherasim, T., Harzallah, M., Berio, G. and Kuntz, P., 2013. "Methods and Tools for Automatic Construction of Ontologies from Textual Resources: A Framework for Comparison and Its Application," *Advances in Knowledge Discovery and Management.* Springer, pp. 177–201.

McCarty, L., T., 2007. "Deep semantic interpretations of legal texts," proceeding of *ICAIL*, pp. 217-224.

Ortiz, A., 2007. "Polionto: Ontology reuse with automatic text extraction from political documents," proceedings of *the 6th doctoral symposium in informatics engineering*.

Francesconi, E. ,Montemagni, S., Peters, W., Tiscornia, D., 2010. "Integrating a bottom–up and top–down methodology for building semantic resources for the multilingual legal domain," Semantic Processing of Legal Texts: where the Language of Law Meets the Law of Language, Springer-Verlag, Berlin, Heidelberg.