




A Bottom-Up Approach for Moroccan Legal Ontology Learning from Arabic Texts

Kaoutar Belhoucine^(✉) , Mohammed Mourchid, Samir Mbarki,
and Abdelaaziz Mouloudi

Faculty of Science, MISC Laboratory, Ibn Tofail University, Kénitra, Morocco
kaoutar.belhoucine@gmail.com, mourchidm@hotmail.com,
mbarkisamir@hotmail.com, mouloudi_aziz@hotmail.com

Abstract. Ontologies constitute an exciting model for representing a domain of interest, since they enable information-sharing and reuse. Existing inference machines can also use them to reason about various contexts. However, ontology construction is a time-consuming and challenging task. The ontology learning field answers this problem by providing automatic or semi-automatic support to extract knowledge from various sources, such as databases and structured and unstructured documents. This paper reviews the ontology learning process from unstructured text and proposes a bottom-up approach to building legal domain-specific ontology from Arabic texts. In this work, the learning process is based on Natural Language Processing (NLP) techniques and includes three main tasks: corpus study, term acquisition, and conceptualization. Corpus study enriches the original corpus with valuable linguistic information. Term acquisition selects tagged lemmas sequences as potential term candidates, and conceptualization drives concepts and their relationships from the extracted terms. We used the NooJ platform to implement the required linguistic resources for each task. Further, we developed a Java module to enrich the ontology vocabulary from the Arabic WordNet (AWN) project.

The obtained results were essential but incomplete. The legal expert revised them manually, and then they were used to refine and expand a domain ontology for a Moroccan Legal Information Retrieval System (LIRS).

Keywords: Ontology learning · Taxonomies definition · Arabic WordNet · NooJ · Legal field · Arabic text

1 Introduction

Ontologies hold great importance for modern knowledge-based systems. They serve as explicit, conceptual knowledge models to share a common understanding of information in a domain and make that knowledge available to information systems [1]. However, the manual construction of ontologies is an expensive and time-consuming task because of the difficulty in capturing knowledge, an issue also known as the “knowledge acquisition bottleneck.” A solution for this issue is providing automatic or at least semi-automatic support for ontology construction. This operation is usually referred to as Ontology Learning (OL) [2].

Cimiano [3] compares the tasks involved in OL to forming a layered cake. The cake is composed, in ascending order, of term acquisition, synonym acquisition, concept formation, taxonomy definition, relation definition, and finally, axiom definition (see Fig. 1). Several ontology-learning tools are proposed in the literature for accomplishing these tasks [4–6]. They differ according to input data types (format and language), output formats, and mainly the methods used in order to extract the ontological structures. Unfortunately, the Arabic language is still not supported by these tools, even though it is one of the most common languages spoken worldwide.

In this paper, we deal with ontology learning from Arabic legal texts. We use the NooJ linguistic platform to semi-automatically process the identified steps: corpus study, term acquisition, and conceptualization. Then we use the AWN project to accomplish the ontology enrichment. Section 2 presents the overall ontology learning process from text: input, output, existing approaches, and prominent ontology learning tools. Section 3 discusses related works in the legal domain. In Sect. 4, we describe the proposed learning process and its implementation in NooJ. Section 5 comments on the learning process and the obtained results. Finally, in Sect. 6, we present our conclusions and plans for future work.

2 Ontology Learning

The term ontology learning refers to the automatic or semi-automatic support for the construction of an ontology [7]. It aims at extracting ontological elements (conceptual knowledge) from a given input text with limited human exertion. Techniques from established fields, such as NLP, data mining, and information retrieval, have been fundamental in developing ontology learning methods [8]. This section presents the inputs used to learn ontologies, the ontology learning tasks, and outputs, existing approaches, and most prominent ontology learning tools.

2.1 Input

There are three different kinds of ontology learning input data [9]: structured (such as databases), semi-structured (e.g., XML), and unstructured (natural language text documents). Unstructured data is the most widely available format for ontology learning input and presents the most common sources for ontology extraction [10]. However, processing unstructured data is a tedious task; indeed, human language is mostly very implicit and allows different people to conceptualize it in different manners [11]. The legal domain is strictly dependent on its linguistic expression and therefore inherits all the challenging problems that this implies. McCarty overtly claimed, “one of the main obstacles to progress in the field of artificial intelligence and law is the natural language barrier” [12].

2.2 Tasks and Outputs

Ontology learning is primarily concerned with defining concepts, relations, and (optionally) axioms from texts. Although there is no standard regarding this development

process, Cimiano [3] describes the tasks involved in ontology learning as forming a layer cake (see Fig. 1). These tasks aim at returning six main outputs: terms, sometimes synonyms, concepts, taxonomic relations, non-taxonomic relations, and finally, axioms.

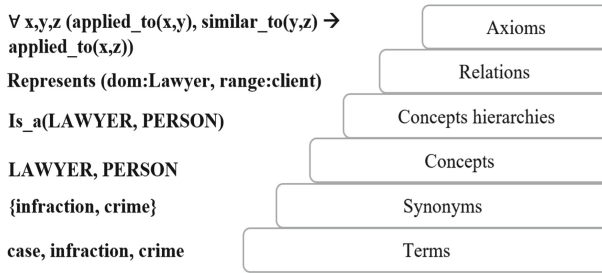


Fig. 1. Ontology Learning “Layer Cake” from [25].

Terms are the most basic building blocks of ontology learning [13]. They can be simple (i.e., single-word) or complex (i.e., multi-word), and are considered linguistic realizations of domain-specific concepts. There are many term extraction methods in the literature. Most of them are based on terminology and NLP research [14–16]; others, on information retrieval methods for term indexing [17].

Synonym discovery consists of finding words that denote the same concept [18]. The synonym layer addresses the acquisition of semantic term variants in and between languages. It is either based on sets, such as WordNet synsets [19] (after sense disambiguation), on clustering techniques [20–23], or on other similar methods, including Web-based knowledge acquisition.

Concepts can be abstract or concrete, real or fictitious. However, the consensus in this field is that concepts should include the following:

- Intension: a formal definition of the set of objects that this concept describes;
- Extension: a set of objects that the definition of this concept describes;
- Lexical realizations: a set of linguistic realizations, (multilingual) terms for this concept.

Most of the research in concept extraction addresses the question from a clustering perspective, regarding concepts as clusters of related terms [3]. This approach overlaps almost entirely with that of term and synonym extraction [24] and can be found in [25].

Concept hierarchies (generalization and specialization) or taxonomies are crucial for any knowledge-based system [24]. There are three main paradigms to induce concept hierarchies from texts:

- Lexico-syntactic patterns, as proposed in [26],
- Harris’s distributional analysis using clustering algorithms [27],
- The document-based notion of term subsumption, as proposed in [28].

Relations refer to any relationship between concepts except taxonomical relations. This includes specific conceptual relationships such as synonymy, possession, attribute-of, and causality, as well as more general relationships referring to any labeled link between a source concept. In the literature, few approaches have addressed the issue of relations extraction from texts, such as the use of an association rules extraction algorithm [29] and the use of syntactic dependencies [30].

Lastly, axioms are propositions that are always taken as true. They act as a starting point for deducing other truths and verifying the correctness of existing ontological elements. The extraction of axioms from the text occurs at an early stage [31]. Initial blueprints of this task can be found in [32]. This work proposes an unsupervised method based on an extended version of Harris's distributional hypothesis in order to discover inference rules.

2.3 Approaches

Several approaches deal with ontology learning from textual resources in the literature. We briefly discuss the most relevant ones for our concerns. Aussenac-Gilles [33] proposed an ontology learning approach based on knowledge elicitation from technical documents. This approach enables the creation of a domain model by analyzing a given corpus using natural language processing (NLP) tools and linguistics techniques. It includes four main activities: corpus constitution, linguistic study, Normalization, and Formalization. Sabou [34] proposed a natural language processing approach that uses syntactic patterns to discover the dependency relations between words. This approach consists of four main steps: term extraction, conceptualization, and enrichment. Mazari [35] proposed an automatic construction approach that uses statistical techniques to extract elements of ontology from Arabic texts. The ontology learning tasks are carried out in three steps: preparing the corpus, extracting concepts, and discovering relations. In the legal domain, all ontology learning experiments mainly focus on concept extraction as the primary step in the ontology development process [36].

2.4 Tools

Ontology learning tools aim to reduce both the time and cost of the ontology development process. They differ in terms of input data types, output formats, and mainly the methods and algorithms used in order to extract the ontological structures. In this subsection, we present the most relevant ontology learning systems from unstructured textual resources.

TERMINAE [6] is a tool based on a methodology elaborated from practical experiments of ontology building. Its originality is to integrate linguistic and knowledge engineering tools. The linguistic engineering part allows term acquisition from textual resources. The knowledge engineering part provides knowledge-base management with an editor and browser for the ontology. This tool helps to represent a notion as a concept, which is called a terminological concept.

Text2Onto [7], is a framework for learning ontologies from textual resources. Text2Onto represents the learned knowledge into a meta-level model called a probabilistic ontology model (POM), which stores the learned primitives independently of a

specific Knowledge Representation (KR) language. It calculates confidence about the correctness of the ontology elements and updates the learned knowledge each time the corpus is changed to avoid processing it from scratch.

Text-to-Knowledge (T2K) [8], is a generic computer platform for data and text mining. T2K extracts domain-specific information from texts by combining linguistic technologies and statistical techniques in three main phases: preprocess text and extract terms, form concepts, and relations or knowledge organization (Table 1).

Table 1. A summary of ontology learning tools.

Tools	Language	Elements extracted	Techniques
Terminae (2005)	French	Terms, synonyms, concepts, taxonomies, non-taxonomic relations	Linguistic and knowledge engineering
Text2Onto (2005)	English Spanish German	Terms, synonyms concepts, taxonomies, non-taxonomic relations, instances	Linguistic processing statistical text analysis, machine learning, association rules
T2K (2008)	Italien	Terms, concepts, taxonomies	Statistical text analysis and machine learning

Unfortunately, most of the existing ontology learning tools do not support Arabic language processing, while a few others lack support.

3 Related Work

Our proposed approach aims to use NLP techniques and tools in order to build a domain-specific ontology from Arabic textual resources. The most closely related works in the legal domain are Francesconi [37] and El Ghosh [10]. Francesconi [37] performed the term extraction task with two different acquisition tools: GATE for English texts and T2K for Italian.

The other tasks, such as evaluating terms, linking them to concepts, and defining relations, were processed under the supervision of ontology engineers and domain experts. For El Ghosh [10], the ontology extraction process has used Text2Onto and is composed of two main phases: linguistic preprocessing and extraction of modeling primitives (concepts, instances, taxonomies, general relations, and disjoint axioms). The resulting ontology is considered an inexpressive ontology and needs to be re-engineered.

Our work differs from previous work in the following aspects. First, we are processing Arabic, one of the most challenging natural languages in the NLP field. Second, we use the NooJ platform to implement the linguistic resources needed for term acquisition and conceptualization. Finally, we are developing a Java module to enrich the ontology vocabulary from the AWN project.

4 Our Work

After a comprehensive literature review, we can see that most of the approaches proposed for learning ontologies from text strongly depend on their specific environment, consisting of language, input, domain, and application. Thus, there is no standard regarding the ontology learning process and no guarantee that the (semi-) automatically generated ontology is sufficiently correct and precise to characterize the domain of interest [10].

For this reason, domain expert intervention throughout the learning process is highly necessary in order to control, complete, and validate the extracted elements. From this perspective, we defined a semi-automatically learning process that involves legal expert intervention and comprises mainly four tasks: corpus study, term acquisition, conceptualization, and enrichment. This section presents the corpus and the platform used to learn the ontology, introduces each learning task, and discusses the obtained results.

4.1 Corpus Definition

We constituted the corpus from the Moroccan family code (Fig. 2), which consists of Arabic natural language texts and includes seven main books composed of 400 articles of law, about 2,700 text units, and 18,000 different tokens.



Fig. 2. Moroccan family code Corpus excerpt.

4.2 Tool Selection

Arabic is a Semitic language that has a very complex morphology [38]; it is a highly inflected and agglutinative language; and, due to this complex morphology, it requires a set of preprocessing routines to be suitable for manipulation.

In the current project, we used NooJ [39] as a natural-language processing tool in order to formalize inflectional and derivational morphology, lexicon, regular grammars, and context-free grammars. NooJ uses an annotation mechanism (stored in each Text Annotation Structure, or TAS) that integrates every single piece of linguistic

information, making it possible to combine morphological constraints in syntactic rules. NooJ is also a powerful corpus processor that supports sophisticated operations, such as information extraction, concordances, and statistical analyses.

4.3 Ontology Learning Process

Corpus Study. This step consists of a lexico-syntactic analysis of Moroccan legal texts. First, we built a legal domain-specific dictionary based on the family code dictionary available on the ADALA Morocco legal and judicial Portal [40]. The built dictionary comprises more than 1,000 entries, consisting of simple terms (nouns and adjectives), compound nouns, pronouns, prepositions, adverbs, and conjunctions. Furthermore, we added to the simple terms the required related inflectional and derivational forms. Below are some examples of the dictionary’s entries (Table 2):

Table 2. Excerpt of dictionary entries.

ميزاٲ	N+Juri+FLX=FlxM+DRV=Mawarit:Flx
مدد بة حالة	N+UNAMB+Juri
شُرعي	ADJ+Juri+FLX=FlxFM+DRV=DrvFM

Second, inspired by Mesfar [41], we modeled a set of morphological grammars that recognize the component morphemes of the agglutinative forms. For instance, the morphological grammar in Fig. 3 allows the identification of the agglutinative word, including various prefixes {[definite article (the, ال)], [prepositions (for, ل), (by, ب)], (conjunctions (and, و))}, and the suffix [pronoun (her, ـه), e.g.: (Her husband, زوجها), (By its expiration, اءءءن اب)].

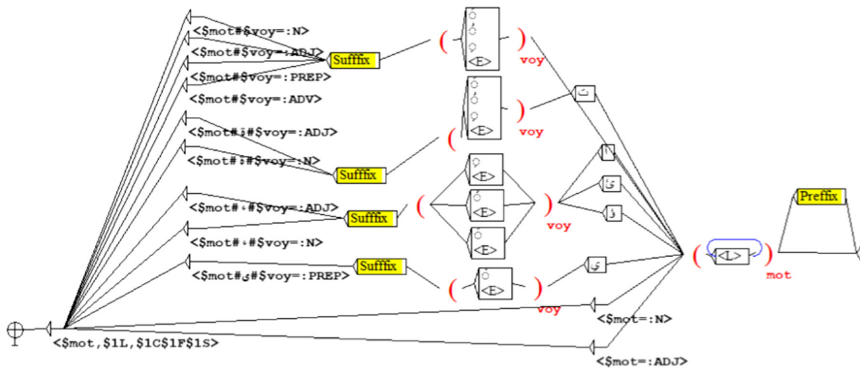


Fig. 3. A morphological grammar of tokenization.

Finally, to solve multi-word unit ambiguities, we modeled local grammars using the feature “ + UNAMB”. The local grammar in Fig. 4 recognizes as nouns both (Son, ابن) and (Son of son, ابن الابن). The corpus was annotated with a lexical coverage rate of 81.83%, which we consider to be a very satisfactory result.

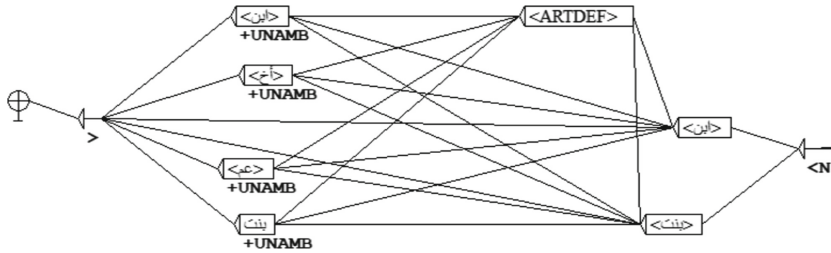


Fig. 4. A syntactic grammar for kinship relationship.

Term Acquisition. After preparing the corpus, we moved to extract the ontology elements. We identified manually, with the legal expert’s help, about 13 patterns of nominal compositions that reference the potential candidate terms (see Table 3). We modeled these patterns using the NooJ local grammars and applied them to extract the corresponding sequences in the corpus. Finally, to keep only the relevant terms, we employed TF-IDF measures of the NooJ statistical module. As a result, we acquired 398 single and multi-word candidate terms.

Table 3. Patterns of the potential candidate terms.

1 unit Pattern	2 units Patterns	3 units Patterns
Noun	Noun_adjectif Adjectif_noun Noun_prep_noun Noun_prep_adjectif	Noun_noun_noun Noun_noun_adjectif Noun_adjectif_noun Noun_noun_conj_noun Noun_noun_prep_noun Noun_adj_prep_noun

Conceptualization. In this step, concepts and their relations are derived from the extracted terms. We elaborated a cascade of local grammars that identifies the candidate terms sharing a large number of syntactic contexts, for instance, those sharing the same head or the same expansion (see Table 4.).

The legal expert used the obtained clusters to define the concepts, their properties, and semantic relationships between them – for instance, hyponymy, hypernymy, and synonymy. For example, the lexical units (daughter, بنت), (wife, زوجة), and (father, أب) share the same syntactic context [(فقفن), noun] and specialize the concept (Close relative, قريب). The lexical units (طلاق, divorce) and (زواج, marriage) share

several syntactic contexts – [noun, (types, أنواع)], [noun, (date, تاريخ)] and [pre-pNoun, (registration of, بتوثيق)] – and specialize the concept (Situation, حالة). Two hundred and thirty single and multiword concepts and 10 semantic relations were identified single and multiword, concepts were identified and 10 semantic relations.

Table 4. Excerpt of the clustered terms.

The terms sharing the expansion (طلاق, divorce)	The terms sharing the expansion (زواج, marriage)	The terms sharing the head (نفقة, expense)
الفسخ أو الطلاق	مستندات الزواج	نفقة القرابة
أسباب الطلاق	أهلية الزواج	نفقة المحضون
بتوثيق الطلاق	بتوثيق الزواج	نفقة نفسه
وثيقة الطلاق	رسم الزواج	نفقة الزوجة
أنواع الطلاق	ولي الزواج	نفقة الأقارب
عدة الطلاق	فترة الزواج	نفقة الأب
تاريخ الطلاق	انواع الزواج	نفقة البنات
عدة الطلاق	إثبات الزواج	نفقة الآباء
رسم الطلاق	تاريخ الزواج	

An excerpt can be seen in Fig. 5, below.

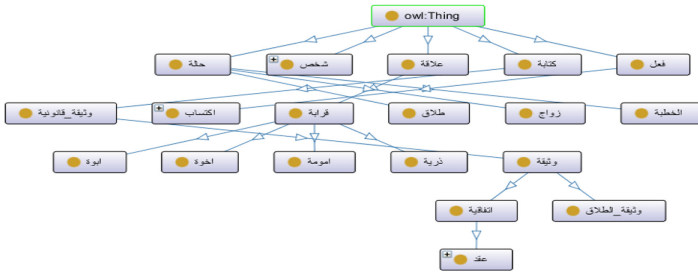


Fig. 5. Excerpt of the taxonomy identified.

Enrichment. At the end of the previous step, we added to the NooJ dictionary the semantic properties that referred to the concepts and their reference hypernym trees. In the current task, we identify the concept synonym sets from the AWN Project [42]. AWN is a lexical database for the Arabic language that groups words into clusters of synonyms called synsets that are linked by semantic relationships. Based on JAWS API [43], we developed a Java module that located for each simple word concept the corresponding synsets in AWN. If a concept had multiple senses, the module constructed an AWN hypernym tree for each and calculated their semantic similarity to the reference hypernym tree. Finally, the module adds the most similar sense’s synonyms to the concept as semantic property in the NooJ dictionary. In following the structure of our lexicon:

Entry, GrammaticalCategory+Concept+HypernymTree=listOfString+Synonyms=istOfString

Example:

رُوج, N + Concept + HypernymTree = رُوج أَقْرِبَ إِشْخَصْ + Synonyms = رُوج أَقْرِبَ إِتْمَلِ إِخْلِيلَ | شَرِيكَ حَيَاةِ

5 Discussion

This section briefly highlights the main issues and remarks identified throughout the learning process from Arabic legal texts. First, the Arabic language's complexity and lack of an ontology learning tool make the learning process from Arabic texts more complicated and challenging than learning from Romance languages. Second, the acquired pieces of information using lexical analysis and term extraction are essential but inexpressive. They need to be revised by a domain expert and re-engineered into the following ontological elements: concepts, concept properties, and relations. Third, analyzing a legal domain-specific corpus can identify relevant concepts and relationships relating to a regulated domain, which provides significant indications for building a legal domain ontology. Last, the NooJ platform offers all the linguistic tools required to implement the ontology learning methods proposed in the literature. Regrettably, it does not support knowledge engineering tools to model the ontological model.

6 Conclusion

In this article, we have presented an overview of ontology learning from text and proposed a bottom-up approach to building a legal domain-specific ontology from unstructured Arabic text. The learning process was identified. We used the linguistic platform NooJ as an NLP tool to extract the ontology elements (concepts and relations) and the AWN project to enrich the ontology vocabulary. The obtained results were validated and completed manually by the legal expert. Future work will focus on the formalization and implementation of the designed ontology. We will also focus on developing our LIRS according to available information in the ontology. We expect that using the ontology will help the results be more semantically related to the query than other related works.

References

1. Grimm, S., Abecker, A., Völker, J., Studer, R.: Ontologies and the semantic web. In: Domingue, J., Fensel, D., Hendler, J.A. (eds.) *Handbook of Semantic Web Technologies*, pp. 507–579. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-540-92913-0_13
2. Cimiano, P., Völker, J., Studer, R.: Ontologies on demand? - A description of the state-of-the-art, applications, challenges and trends for ontology learning from text. *Inf. Wissenschaft und Praxis* **57**(6–7), 315–320 (2006)

3. Cimiano, P.: *Ontology Learning and Population from Text. Algorithms, Evaluation and Applications*. Springer, New York (2006). <https://doi.org/10.1007/978-0-387-39252-3>. ISBN 978-0387-30632-2
4. Biébow, B., Szulman, S., Clément, A.J.B.: TERMINAE: a linguistics-based tool for the building of a domain ontology. In: Fensel, D., Studer, R. (eds.) *EKAW 1999. LNCS (LNAI)*, vol. 1621, pp. 49–66. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48775-1_4
5. Cimiano, P., Völker, J.: Text2Onto. In: Montoyo, A., Muñoz, R., Métais, E. (eds.) *NLDB 2005. LNCS*, vol. 3513, pp. 227–238. Springer, Heidelberg (2005). https://doi.org/10.1007/11428817_21
6. Dell’Orletta, F., Venturi, G., Cimiano, A., Montemagni, S.: T2K²: a system for automatically extracting and organizing knowledge from texts. In: *Proceeding of LREC*, pp. 26–31, Iceland (2014)
7. Drumond, L., Girardi, R.: A survey of ontology learning procedures. *Proceedings of the 3rd Workshop on Ontologies and their Applications*, vol. 427 of *CEUR Workshop Proceedings*, Salvador, Bahia, Brazil (2008)
8. Wong, W., Liu, W., Bennamoun, M.: Ontology learning from text: a look back and into the future. *ACM Comput. Surv.* **44**, 1–36 (2011)
9. Benz, D.: *Collaborative ontology learning*. Master’s thesis, University of Freiburg (2007)
10. El Ghosh, M., Naja, H., Abdulrab, H., Khalil, M.: Ontology learning process as a bottom-up strategy for building domain-specific ontology from legal texts. *ICAART* **2**, 473–480 (2017)
11. Rogger, M., Thaler, S.: *Ontology Learning*. Seminar paper, Applied Ontology Engineering, Leopold–Franzens–University Innsbruck (2010)
12. McCarty, L., T.: Deep semantic interpretations of legal texts. In: *Proceeding of ICAIL*, pp. 217–224 (2007)
13. Wong, W.Y.: *Learning Lightweight Ontologies from Text across Different Domains using the Web as Background Knowledge*. Ph.D. thesis, University of Western Australia, School of Computer Science and Software Engineering (2009)
14. Borigault, D., Jacquemin, C., L’Homme, M.C. (eds.): *Recent Advances in Computational Terminology*. *Natural Language Processing Series*, vol. 2, pp. 328–332 John Benjamins Publishing Company, Amsterdam (2001)
15. Frantzi, K., Ananiadou, S.: The C-value/NC-value domain independent method for multiword term extraction. *J. Nat. Lang. Process.* **6**, 145–179 (1999)
16. Pantel, P., Lin, D.: A statistical corpus-based term extractor. In: Stroulia, E., Matwin, S. (eds.) *AI 2001. LNCS (LNAI)*, vol. 2056, pp. 36–46. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45153-6_4
17. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **24**(5), 515–523 (1988)
18. El Ghosh, M.: *Automation of legal reasoning and decision based on ontologies*. Ph.D. thesis, Web. Normandie Université (2018)
19. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Introduction to WordNet: an on-line lexical database. *Int. J. Lexicogr.* **3**, 235–244 (1990)
20. Bourigault, D., Jacquemin, C.: Term extraction+ term clustering: an integrated platform for computer-aided terminology. In: *Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics*, pp. 15–22 (1999)
21. Faure, D., Nédellec, C.: Knowledge acquisition of predicate argument structures from technical texts using machine learning: the system ASIUM. In: Fensel, D., Studer, R. (eds.) *EKAW 1999. LNCS (LNAI)*, vol. 1621, pp. 329–334. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48775-1_22

22. Maedche, A., Staab, S.: The text-to-onto ontology learning environment. In: Software Demonstration at ICCS-2000-Eight International Conference on Conceptual Structures, August 2000
23. Drymonas, E., Zervanou, K., Petrakis, E.G.M.: Unsupervised ontology acquisition from plain texts: the OntoGain system. In: Hopfe, C.J., Rezgui, Y., Métais, E., Preece, A., Li, H. (eds.) NLDB 2010. LNCS, vol. 6177, pp. 277–287. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13881-2_29
24. Buitelaar, P., Cimiano, P., Magnini, B.: *Ontology Learning from Text: An Overview, Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press, Amsterdam (2005)
25. Reinberger, M., Spyns, P.: *Unsupervised Text Mining for the Learning of DOGMA-inspired Ontologies. Ontology Learning from Text*. IOS Press, Amsterdam (2005)
26. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th Conference on Computational Linguistics, vol. 2, pp. 539–545, Association for Computational Linguistic (1992)
27. Harris, Z.: *Mathematical Structures of Language*. Wiley, Hoboken (1968)
28. Sanderson, M., Croft, B.: Deriving concept hierarchies from text. In: *Research and Development in Information Retrieval*, pp. 206–213 (1999)
29. Maedche A., Staab, S.: Discovering conceptual relations from text. In: Horn, W. (ed.) *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000)*, (2000)
30. Gamallo, P., Gonzalez, M., Agustini, A., Lopes, G., de Lim, V.S.: Mapping syntactic dependencies onto semantic relations. In: *ECAI Workshop on Machine Learning and Natural Language Processing for Ontology Engineering* (2002)
31. Lin, D., Pantel, P.: DIRT - Discovery of inference rules from text. In: *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 323–328 (2001)
32. Lin, D., Pantel, P.: Induction of Semantic Classes from Natural Language Text. In: *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 317–322 (2001)
33. Aussenac-Gilles, N., Biébow, B., Szulman, S.: Revisiting ontology design: a method based on corpus analysis. In: Dieng, R., Corby, O. (eds.) *EKAW 2000*. LNCS (LNAI), vol. 1937, pp. 172–188. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-39967-4_13
34. Sabou, M.: Visual support for ontology learning: an experience report. In: *Proceeding of IV 2005*, London (2005)
35. Mazari, C., Aliane, H., Alimazighi, Z.: Automatic construction of ontology from Arabic texts. In: *Proceeding of ICWIT*, pp. 193–202 (2012)
36. Lenci, A., Montemagni, S., Pirrelli, V., Venturi, G.: Ontology learning from Italian legal texts. In: Breuker, J., et al. (eds.) *Law, Ontologies and the Semantic Web – Channelling the Legal Information Flood*, *Frontiers in Artificial Intelligence and Applications*, vol. 188, pp. 75–94. Springer
37. Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D.: Integrating a bottom-up and top-down methodology for building semantic resources for the multilingual legal domain. In: Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (eds.) *Semantic Processing of Legal Texts*. LNCS (LNAI), vol. 6036, pp. 95–121. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12837-0_6
38. Gharib, T.F., Habib, M.B., Fayed, Z.T.: Arabic text classification using support vector machines. *Int. J. Comput. Appl.* **16**(4), 192–199 (2009)
39. Silberstein, M.: *Formalizing Natural Languages: The NooJ Approach*. Wiley, Hoboken (2016)
40. ADALA Morocco legal and judicial Portal. <https://adala.justice.gov.ma/FR/Home.aspx>. Accessed 12 Sept 2020

41. Mesfar, S.: Named entity recognition for Arabic using syntactic grammars. In: Kedad, Z., Lammari, N., Métais, E., Meziane, F., Rezgui, Y. (eds.) NLDB 2007. LNCS, vol. 4592, pp. 305–316. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73351-5_27
42. Black, W.: Introducing the Arabic WordNet project. In: Sojka, Choi, Fellbaum, Vossen (eds.) Proceedings of the third International WordNet Conference (2006)
43. JAWS. <https://github.com/jaytaylor/jaws>. Accessed 12 Sept 2020