Bottom-up enrichment of top-down ontologies through annotation

Cristian Cardellino, Milagro Teruel, Raphaël Gazzotti, Laura Alonso Alemany, Serena Villata, Catherine Faron Zucker

Bottom-up enrichment of top-down ontologies through annotation

Cristian Cardellino[1], Milagro Teruel[1], Raphael Gazzotti[2], Laura Alonso Alemany[1], Serena Villata[2], and Catherine Faron-Zucker[1]

[1] University of Cordoba, Argentina,
crscardellino@gmail.com, milagro.teruel@gmail.com,
alemany@famaf.unc.edu.ar
[2] Université Côte d'Azur, CNRS, Inria, I3S,
{firstname.lastname}@unice.fr

Abstract. We present a methodology to enhance domain-specific ontologies by (i) addressing a manual annotation of texts with the concepts in the domain ontology, *(ii)* matching the annotated concepts with the closest YAGO-Wikipedia concept available, and *(iii)* using concepts from other ontologies that cover complementary domains. This method reduces the difficulty of aligning ontologies, because annotators are asked to associate two labels from different inventories to a concrete example, which requires a simple judgment. In a second phase, those correspondences are consolidated into a proper alignment. The resulting alignment is a partial connection between diverse ontologies, and also a strong connection to Linked Open Data. By aligning these ontologies, we are increasing the ontological coverage for texts in that domain. Moreover, by aligning domain ontologies to the Wikipedia (via YAGO), we can obtain manually annotated examples for some of the concepts, effectively populating the ontology with examples. We present two applications of this process in the legal domain. First, we annotate sentences of the European Court of Human Rights with the LKIF ontology, at the same time matching them with the YAGO ontology. Second, we annotate a corpus of customer questions and answers from an insurance web page with the P&C ontology for the insurance domain, matching it with the YAGO ontology and complementing it with a financial ontology.

1 Introduction and Motivation

Ontologies are the main mechanism for domain-specific knowledge representation as they allow for an exhaustive characterization of the domain of interest. However, their manual creation and maintenance is a very time-consuming and challenging task: domain-specific information needs to be created by domain experts to capture their full semantics.

A number of upper ontologies have been developed for the legal domain [22, 17, 5, 12, 14, 9, 2], but they only deal with higher-level abstract concepts and do not include concrete entities that are organized by those concepts. Other ontologies have been developed to improve the performance of applications on legal texts, and thus are aimed to ontologize entities in those texts. These ontologies suffer from dealing with too much detail, which makes it difficult to find adequate abstractions.

In this paper we propose a methodology to bridge the gap between these two approaches by using more abstract ontologies to annotate concrete entities occurring in texts. In the process of annotation, abstract ontologies provide generalizations for concrete concepts, and concrete concepts populate abstract ontologies, which makes them useful for tasks like information retrieval, question answering or information extraction.

Another problem of legal-domain ontologies is their limited coverage. Indeed, many of the existing ontologies cover only subdomains of the legal domain. We address this problem by annotating entities with two or more ontologies, using as a backoff a general-domain ontology, YAGO [20]. As a result, the ontologies that are used for annotation end up being aligned. Alignment of ontologies is a very challenging task, because they represent very subtle distinctions. The process of finding semantically equivalent concepts in two different conceptualizations of the same domain is very difficult for humans, even if they are adequately trained. The annotation task alleviates this difficulty by making decisions more concrete. In this task, human experts detect mentions of the relevant concepts in naturally occurring text and assign them to a concept of each of the ontologies to be aligned, which is much more natural for the annotators.

Our aim is not to develop a reference ontology, but to focus on a useful, working mapping, based on naturally occurring examples, that will enrich ontologies and make them more useful for practical use of the ontologies in wide-coverage Information Extraction (IE) tasks. Using the YAGO ontology has the additional benefit of obtaining examples of mentions of those concepts in the text of Wikipedia, which can be then used to train an automatic analyzer.

We describe this method by applying it to two sub-domains of the legal scenario, namely court judgments and insurance customer services, and we target two different applications, i.e., Named Entity Recognition and Classification (NERC), and Question Classification (QC).

The rest of the paper is organized as follows: in the following Section, we discuss some relevant work on the development of legal ontologies. Then, we outline our methodology. In Section 4, we describe the resources that have been used for annotation: the guidelines and annotation interface. Then we describe how we have instantiated this methodology to enrich the LKIF ontology, and how this has been applied to obtain a legal Named Entity Recognizer and Classifier. We also outline an ongoing application for the insurance domain. Conclusions end the paper.

2 Related work

A number of upper ontologies have been developed for the legal domain [22, 17, 5, 12, 14, 9, 2]. These ontologies are built following a top-down approach, designed from a very formal perspective for automated reasoning tasks. In general, each of them cover only a limited part of the legal domain, and from a particular perspective. They describe interactions between abstract concepts that *make* the legal domain, but do not reach down to include the detailed list of concrete entities that are organized by those concepts. As such, they are of little use for a big bulk of tasks that can be addressed by less sophisticated NLP techniques with a big impact on everyday work for law pro-

fessionals, like concept-oriented Information Retrieval (IR), advancing in the way of Information Extraction.

On the other hand, the bottom-up approach to ontology construction builds upon the actual concepts that are found in naturally occurring texts, and attempt to organize them. There are organizing efforts that are not ontological in nature, but more of a thesaurus-like organization, like all those related to the Eurovoc Thesaurus[3] or IATE[4]. In the same line, various projects and initiatives are aimed to align legal vocabularies from different countries, also without an ontological organization (LISE[5], Legivoc[23]).

Nevertheless, more recently some ontologies of the legal domain have been built bottom-up [11, 21], mainly oriented to applications like information retrieval. With this goal in mind, they need to account for the actual terms that occur in legal documents.

As a backoff ontology to ensure coverage and connection to Linked Open Data, we use YAGO [20], is a knowledge base automatically extracted from Wikipedia, WordNet, and GeoNames, and linked to the DBpedia ontology[6] and to the SUMO ontology[7]. It represents knowledge of more than 10 million entities, and contains more than 120 million facts about these entities, tagged with their confidence. This information was manually evaluated to be above 95% accurate. It also bridges the gap between an upper level ontology, and concrete real world entities.

To develop legal ontologies, a number of frameworks and methodologies have been proposed. The most comparable to our own is that of Ajani et al. 2016 [1], a method, schema and tool intended to help law practitioners organize European terminology from different domains, countries and languages at different levels of abstractions. It is a bottom-up approach and combines it with a schema that plays a role comparable to the role of ontologies in our annotation approach. The multi-lingual aspect presents problems comparable to our multi-ontology approach.

One of the by-product of our method is that the domain ontology gets populated as a result of its alignment to YAGO and the automatic transfer of Wikipedia examples through YAGO. Some approaches address the problem of legal ontology population. Bruckschen and colleagues [6] describe an ontology population approach to legal data running Named Entity Recognition over a corpus of legal and normative documents for privacy. Lenci *et al.* [18] report an ontology learning system that applies NLP and Machine Learning methods to extract terms and relations from free text to identify the classes of the ontology and hyponymy relations. Boella and colleagues [15, 4] identify and extract norm elements in European Directives using dependency parsing and semantic role labeling. The experimental system takes advantage of the way the Eunomos system [3] they developed present norms in a structured format. This approach focuses on how to extract prescriptions (i.e., norms) and other concepts (e.g., reason, power, obligation, nested norms) from legislation, and how to automate ontology construction. Similarly, they [4] propose an approach that provides POS tags and syntactic relations as input of a SVM to classify textual instances to be associated to legal concepts.

[3] http://eurovoc.europa.eu
[4] http://iate.europa.eu
[5] http://www.lise-termservices.eu
[6] http://wiki.dbpedia.org/
[7] http://www.adampease.org/OP/

3 Outline of the approach

Schematically, the annotation-based alignment process is as follows.
Given a target domain,

1. gather a corpus of text documents representative of the domain, and one or more ontologies specific for that domain
2. manually identify entities in the text
3. tag each entity with either
   (a) the most specific concept in the domain ontology, if it exists, or
   (b) the most specific concept from another domain ontology, or
   (c) the most specific concept in YAGO or Wikipedia.
4. find the most specific concept in YAGO or, if the concept is not in YAGO, in Wikipedia. Take into account that the most specific concept may be **the actual entity**.

After the annotation process, we revise the resulting mappings to check that the resulting alignments are sound and resolve some problems. In case the YAGO node that was assigned has a granularity that is too fine for the concept assigned from the domain-specific ontology, we establish the mapping between that concept and the most adequate ancestor of the selected YAGO node, as can be seen in the following example. When some equivalent concept has been found, we establish the alignment using the OWL primitives `equivalentClass` and `subClassOf`. Relations are not aligned, but only classes.

Example 1.

domain-specific

The [Court]$_{PublicBody}$ is not convinced by the reasoning of the [combined divisions of the Court of Cassation]$_{PublicBody}$, because it was not indicated in the [judgment]$_{Decision}$ that [Eitim-Sen]$_{LegalPerson}$ had carried out [illegal activities]$_{Crime}$ capable of undermining the unity of the [Republic of Turkey]$_{LegalPerson}$.

YAGO

The [Court]$_{\text{wordnet trial court 108336490}}$ is not convinced by the reasoning of the [combined divisions of the Court of Cassation]$_{\text{wordnet trial court 108336490}}$, because it was not indicated in the [judgment]$_{\text{wordnet judgment 101187810}}$ that [Eitim-Sen]$_{\text{wordnet union 108233056}}$ had carried out [illegal activities]$_{\text{wordnet illegality 104810327}}$ capable of undermining the unity of the [Republic of Turkey]$_{person}$.

We also find semantic areas that are not covered by the current domain-specific ontology, and that may need to be complemented by other domain-specific ontologies. These areas are identified because the annotator manually introduced a concept that was not available in the ontology, either in the domain-specific ontology or in YAGO. In that case, we look for complementary ontologies or make a point to have them developed in the future.

By doing this, named entities are associated to concepts from both the domain ontology and Wikipedia, and thus an alignment is effectively established between both. This alignment allows to transfer properties from one ontology to the other, like the relations of the nodes, leading to a relevant result for inference and reasoning tasks.

Being of importance for NLP applications like Named Entity Recognition and Classification or Information Extraction, this mapping also provides the domain ontology with manually annotated examples from the Wikipedia. Wikipedia provides a fair amount of naturally occurring text where some (though not all) entity mentions are manually tagged and linked to an ontology, i.e., the DBpedia ontology [13]. We consider as tagged entities the spans of text that are an anchor for a hyperlink whose URI is one of the entities that have been mapped through the annotation process.

4 Annotation of texts

The process of text annotation requires extensive support to provide consistency among annotators and reproducibility of the results. To achieve that, we developed guidelines for annotators and an annotation interface.

4.1 Guidelines

The guidelines were roughly based on the LDC guidelines for annotation of Named Entities [8], but adapted to the annotation of legal concepts. Slightly different versions of the guidelines were developed for the different corpora, to address specific needs.

Concepts in legal ontologies do not have the same semantics as your prototypical Named Entity, but a comparable textual representation in text, as can be seen in the following example[8]:

Example 41 *The* [*Court*]$_{PublicBody}$ *is not convinced by the reasoning of the* [*combined divisions of the Court of Cassation*[*, because it was not indicated in the* [*judgment*]$_{Decision}$ *that* [*Eitim-Sen*]$_{LegalPerson}$ *had carried out* [*illegal activities*]$_{Crime}$ *capable of undermining the unity of the* [*Republic of Turkey*]$_{LegalPerson}$.

In guidelines we defined which parts of the documents to tag, leaving out the most formulaic and content-poor parts.

We provide guidelines to determine the textual representation of concepts, that is, how they span in text. We establish that:

- Articles and determiners are not tagged as part of the concept.

  Example 42 *Lastly, the* [*applicant*] *pointed out that the* [*United Nations Human Rights Committee*] *had already found...*

- Concepts are not embedded unless they cannot be separated. If a complex syntactical structure contains two concepts that can be textually separated, they are tagged as separate concepts. If they are not textually separable, then the syntactical head is tagged, and the depending concept is included in the span but not tagged separately.

  Example 43 [*assurance individuelle scolaire*]$_{insurance}$ *de* [*John Smith*]$_{Person}$.

[8] The "Crime" concept is not present in the original LKIF, because it does not cover the subdomain of Procedural Law, but it has been added to annotate the judgments of the ECHR.

– Proper names are always tagged, even if they do not represent a legal concept, because they are helpful for the final application and thus included in the manual annotation process.

  Example 44 *Lastly, the applicant pointed out that the* [*United Nations Human Rights Committee*] *had already found a violation by* [*Spain*] *...*

– Nominalizations of legal actions are tagged, including non-tensed verbal forms.

  Example 45 *Lastly, the applicant pointed out that the United Nations Human Rights Committee had already found a* [*violation*] *by Spain on grounds of* [*discrimination*]*, which was proof that* [*discrimination*] *against immigrant black women was a structural problem in the country.*

Non-legal named entities (places, people, dates) may or may not be tagged depending on the application. Tensed verbs indicating actions that are concepts of the ontology may or may not be tagged, depending on the final application.

4.2 Annotation interface

To carry out annotation, we adapted an annotation interface for NERC from `https://github.com/mayhewsw/ner-annotation`, and the resulting code code is available at `https://github.com/MIREL-UNC/ner-annotation`. The process of annotation with this interface is as follows:

1. Upload a number of documents to be annotated with the ontology.
2. Load the concepts in the domain-specific ontology.
3. Annotate.
   (a) When the annotator finds an entity in the text, she selects the first word and identifies the span of the entity.
   (b) The entity is assigned a label from the domain-specific ontology, which is chosen from a drop-down menu that contains all the concepts in the ontology, as can be seen in Figure 1. This label is the most concrete concept for that entity in the ontology.
   (c) Then, it is assigned the adequate concept in the YAGO ontology, which is the exact canonical name of the entity that is mentioned. Concepts that are used for the first time to annotate are manually written in the box for the labels, and from then on they are available for further uses in the drop-down menu. For instance, as visualized in Figure 1, the entity "Government" in the text is annotated with the LKIF class `Public Body` and the Wikipedia URI `https://en.wikipedia.org/wiki/Government_of_Spain`, since the exact entity could not be found in YAGO.
   (d) If an entity of interest cannot be property labelled with the concepts in the domain ontology or with a YAGO URI, the annotator looks for that concept in Wikipedia. The new label is manually written in the text box for the corresponding label, and it is available from then on in the drop-down menu.

Fig. 1. The annotation of the entity *Government* in the domain-specific ontology LKIF and the Wikipedia.

5 Application to LKIF

As a first use case, we applied the proposed methodology to an upper ontology of the legal domain, the well-known LKIF ontology [14], over the judgments of the European Court of Human Rights. The LKIF ontology is not specific of the domain of judicial procedures, but it is a reference ontology of the legal domain, so we chose it as a first proof of concept.

5.1 Domain ontology and corpus

The LKIF core legal ontology [14] is an abstract ontology describing a core of basic legal concepts developed within the EU-funded Estrella Project. It consists of various modules with high-level concepts, and then three modules with law-specific concepts, with a total of 69 law-specific classes. It covers many areas of the law, but it is not populated with concrete real-world entities.

The HUDOC (Human Rights Documentation)[9] provides access to the case-law of the European Court of Human Rights (Grand Chamber, Chamber and Committee judgments and decisions, communicated cases, advisory opinions and legal summaries from the Case-Law Information Note), the European Commission of Human Rights (decisions and reports) and the Committee of Ministers (resolutions).

We annotated excerpts from 5 judgments of the ECHR, obtained from the Court website[10] and totalling 19,000 words. We identified 1,500 entities, totalling 3,650 words. There were 4 different annotators, and three judgments were annotated by at least 2

[9] `hudoc.echr.coe.int`
[10] `hudoc.echr.coe.int`

annotators independently, to assess inter-annotator agreement using Cohen's kappa coefficient [7]. The agreement between judges ranged from $\kappa = .4$ to $\kappa = .61$. Most of the disagreement between annotators was found for the recognition of concepts, not for their classification. We are working on developing the guidelines to enhance consistency among annotators. We will also apply automatic pre-processing and post-edition to annotated texts, in order to spot and correct errors.

5.2 Resulting mapping

After annotation, the mapping between concepts of LKIF and concepts of YAGO was revised and consolidated as explained in Section 3. Out of a total of 69 classes in the selected portion of the LKIF ontology, 30 could be mapped to a YAGO node, either as children or as equivalent classes. Two YAGO classes were mapped as parent of an LKIF class, although these we are not exploiting in this approach. 55% of the classes of LKIF could not be mapped to a YAGO node, because they were too abstract (i.e., *Normatively Qualified*), there was no corresponding YAGO node circumscribed to the legal domain (i.e., *Mandate*), there was no specific YAGO node (i.e., *Mandatory Precedent*), or the YAGO concept was overlapping but not roughly equivalent (as for "*agreement*" or "*liability*"). The resulting alignment is available online at `https://dl.dropboxusercontent.com/u/15116330/maply_v1.ttl`.

Seen from the YAGO side, 47 classes were mapped to a LKIF class, with a total of 358 classes considering their children, and a total of 174,913 entities. We retrieved 4'5 million occurrences of these entities within the Wikipedia text. However, not all of these classes were equally populated with mentions. The number of mentions per class is highly skewed, with only half of YAGO classes having any mention whatsoever within the Wikipedia text. Of these 122 populated YAGO classes, only 50 were heavily populated, with more than 10,000 mentions, and 11 had less than 100 mentions. When it comes to particular entities, more than half of the entities had less than 10 mentions in text, only 15% had more than 100 and only 2% had more than 1000.

Moreover, the subdomain of Procedural Law, which is obviously present within the judgments of the ECHR, is not represented in LKIF. Those concepts are currently annotated with YAGO labels only. We will complement this with an ontology of procedural law.

5.3 Learnign a NERC for the legal domain

Through the connection between LKIF and the Wikipedia through YAGO, we obtained material to train a Named Entity Recognizer and Classifier for the legal domain. We downloaded a XML dump of the English Wikipedia[11] from March 2016, and we processed it via the WikiExtractor [19] to remove all the XML tags and Wikipedia markdown tags, but leaving the links. We extracted all those articles that contained a link to an entity of YAGO that belongs to our mapped ontology. We considered as tagged entities the spans of text that are an anchor for a hyperlink whose URI is one of the mapped entities. We obtained a total of 4,5 million mentions, corresponding to 102,000

[11] `https://dumps.wikimedia.org/`

unique entities. Then, we extracted sentences that contained at least one mention of a named entity.

We consider the problem of Named Entity Recognition and Classification as a word-based representation, i.e., each word represents a training instance. Then, words within the anchor span belong to the I class (**I**nside a Named Entity), others to the O class (**O**utside a Named Entity). The O class made more than 90% of the instances. This imbalance in the classes results largely biased the classifiers, so we randomly subsampled non-named entity words to make them at most 50% of the corpus. The resulting corpus consists of 21 million words, with words belonging to the O-class already subsampled.

Using the corpus obtained from the Wikipedia, we trained a neural network classifier for Named Entity Recognition and Classification. The objective of this classifier is to identify in naturally occurring text mentions the Named Entities belonging to the classes of the ontology, and classify them in the corresponding class, at different levels of granularity. Note that we do not consider here the URI level, which is treated qualitatively differently by the Named Entity Linking approach.

We represented examples with a subset of the features proposed by Finkel *et al.* [10] for the Stanford Parser CRF-model. For each instance (i.e., each word), we used: current word, current word PoS-tag, all the n-grams ($1 \leq n \leq 6$) of characters forming the prefixes and suffixes of the word, the previous and next word, the bag of words (up to 4) at left and right, the tags of the surrounding sequence with a symmetric window of 2 words, and the occurrence of a word in a full or part of a gazetteer. We applied feature selection with Variance Threshold, filtering out all features with variance less than 2e-4, reducing the amount of features to 11997.

Alternatively, we also trained the classifier with the same approach, but using the examples of the manual annotation of the judgments of the ECHR, which are fewer. We evaluated the classifier with these two different trainings both in the Wikipedia and the judgments of the ECHR. Results can be seen in Table 1.

approach	accuracy	precision	recall	F1
test on Wikipedia, trained on Wikipedia	.95	.76	.64	.69
test on ECHR, trained on Wikipedia	.89	.16	.08	.08
test on ECHR, trained on ECHR	**.95**	.76	.76	.75

Table 1. Results for Named Entity Recognition and Classification on the test portion of the Wikipedia corpus or the ECHR, trained with Wikipedia examples or with the annotations for the ECHR. Accuracy figures take into consideration the majority class of non-NEs, but precision and recall are an average of all classes (macro-average) except the majority class of non-NEs.

We can see that the results are very good, but that the approach is very sensitive to domain change. Indeed, when the classifier is trained with the Wikipedia and tested on the ECHR, the performance drops dramatically, specially in recall.

6 Application to the insurance domain

As a second proof of concept, we applied this methodology to the insurance domain. This second proof of concept is still ongoing.

As a reference ontology for this domain we used the Property And Casualty Information Models, Version 1.0[12][16], developed by the Insurance Working Group of the Object Modelling Group (OMG)[13]. It is focused mainly on the regulated USA Property and Casualty insurance industry for both Personal and Commercial lines. The ceded reinsurance view is included; but, the reinsurer view is not. The WG initial submission focused on the data and models needed to support New Business, Policy Administration, and Claims.

The corpus to be annotated are questions and answers that customers of the French branch of a big insurance group ask to customer service through the webpage and by mail. They are in French, user-generated and cover different topics. The goal of annotation in this case is to improve automated question answering, and eventually developing a conversational bot for this domain.

The guidelines for annotation differ from the guidelines developed for the annotation of the corpus of ECHR in that tensed verbs are annotated as concepts. However, since their syntactical behaviour is very different from substantives, they are assigned a distinctive marker, so that they can be easily separated for experiments. Moreover, non-legal named entities, like dates, locations, amounts, etc. are also tagged.

We have found that the domain-specific ontology did not cover properly the domain of financial concepts. In current annotation, we are complementing it with the Financial Industry Business Ontology (FIBO) [14], again developed by the OMG group.

We are planning to apply the resulting annotation to improve question classification, first, by using the gold standard annotation, and, in a second phase, by training a specific NERC to identify legal concepts and applying it as a pre-process for question classification. To do that, we will apply the method described in Section 5.3.

7 Discussion and Future Developments

We have presented a methodology to enhance domain-specific ontologies of the legal domain. This enhancement consists in aligning them to a general-domain ontology, i.e., YAGO. The alignment is driven by examples of the concepts in naturally occurring texts, which facilitates the selection of the most adequate concept for the human annotator. After this first matching of concepts, the alignment is revised independently of the examples, applying abstraction where it is needed and identifying subdomains that are not covered and need to be complemented with another ontology. We have developed guidelines and a graphical annotation interface to aid this process.

We describe two applications of this methodology, in two different domains, namely the judgments of the Court, and questions and answers of an insurance company customer service, and having in mind two different target applications, concept recognition

[12] http://www.omg.org/spec/PC/1.0/

[13] http://www.omgwiki.org/pcwg/doku.php

[14] http://www.omg.org/spec/EDMC-FIBO/

and classification, and question classification. We show that the methodology applies satisfactorily in both cases.

Future work includes increasing the consistency of annotations, by improving the guidelines and applying automatic pre-processing and post-editing. We also plan to develop specific guidelines for the interrelation between domain ontologies, when more than one is used to annotate the same corpus..

References

1. Ajani, G., Boella, G., Caro, L.D., Robaldo, L., Humphreys, L., Praduroux, S., Rossi, P., Violato, A.: The european taxonomy syllabus: A multi-lingual, multi-level ontology framework to untangle the web of european legal terminology. Applied Ontology 11(4), 325–375 (2016), http://dx.doi.org/10.3233/AO-170174
2. Athan, T., Governatori, G., Palmirani, M., Paschke, A., Wyner, A.: LegalRuleML: Design principles and foundations. In: Wolfgang Faber and Adrian Pashke (ed.) The 11th Reasoning Web Summer School. pp. 151–188. Springer, Berlin, Germany (jul 2015)
3. Boella, G., Caro, L.D., Humphreys, L., Robaldo, L., Rossi, P., van der Torre, L.: Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. Artif. Intell. Law 24(3), 245–283 (2016), http://dx.doi.org/10.1007/s10506-016-9184-3
4. Boella, G., Caro, L.D., Ruggeri, A., Robaldo, L.: Learning from syntax generalizations for automatic semantic annotation. J. Intell. Inf. Syst. 43(2), 231–246 (2014), http://dx.doi.org/10.1007/s10844-014-0320-9
5. Breuker, J.: Constructing a legal core ontology: LRI-Core. In: Workshop on Ontologies and their Applications. Sao Luis, Maranhao, Brazil (2004)
6. Bruckschen, M., Northfleet, C., da Silva, D., Bridi, P., Granada, R., Vieira, R., Rao, P., Sander, T.: Named entity recognition in the legal domain for ontology population. In: 3rd Workshop on Semantic Processing of Legal Texts (SPLeT 2010) (2010)
7. Cohen, J.: A coefficient of agreement for nominal scales. Educational & Psycological Measure 20, 37–46 (1960)
8. Consortium, L.D.: Deft ere annotation guidelines: Entities v1.7. http://nlp.cs.rpi.edu/kbp/2014/ereentity.pdf (2014)
9. van Engers, T., Boer, A., Breuker, J., Valente, A., Winkels, R.: Ontologies in the legal domain. In: Digital Government, Integrated Series in Information Systems, vol. 17, chap. 13, pp. 233–261. Springer (2008)
10. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. pp. 363–370. ACL '05, Association for Computational Linguistics, Stroudsburg, PA, USA (2005), http://dx.doi.org/10.3115/1219840.1219885
11. Gangemi, A., Sagri, M.T., Tiscornia, D.: Metadata for content description in legal information. In: In Proc.s of LegOnt Workshop on Legal Ontologies (2003)
12. Gangemi, A., Sagri, M.T., Tiscornia, D.: A constructive framework for legal ontologies. Law and the Semantic Web pp. 97–124 (2005)
13. Hahm, Y., Park, J., Lim, K., Kim, Y., Hwang, D., Choi, K.S.: Named entity corpus construction using wikipedia and dbpedia ontology. In: Chair), N.C.C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland (may 2014)

14. Hoekstra, R., Breuker, J., Bello, M.D., Boer, A.: The lkif core ontology of basic legal concepts. In: Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007) (2007)
15. Humphreys, L., Boella, G., Robaldo, L., di Caro, L., Cupi, L., Ghanavati, S., Muthuri, R., van der Torre, L.: Classifying and extracting elements of norms for ontology population using semantic role labelling. In: Proceedings of the Workshop on Automated Detection, Extraction and Analysis of Semantic Information in Legal Texts (2015)
16. Kalou, K., Koutsomitropoulos, D.: Linking data in the insurance sector: a case study. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. pp. 320–329. Springer (2014)
17. van Kralingen, R.: A conceptual frame-based ontology for the law. In: Proceedings of the First International Workshop on Legal Ontologies. pp. 6–17 (1997)
18. Lenci, A., Montemagni, S., Pirrelli, V., Venturi, G.: Ontology learning from italian legal texts. In: Proceeding of the 2009 Conference on Law, ontologies and the Semantic Web: Channelling the Legal information Flood (2009)
19. of Pisa, M.U.: Wikiextractor. http://medialab.di.unipi.it/wiki/Wikipedia_Extractor (2015)
20. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: A core of semantic knowledge. In: Proceedings of the 16th International Conference on World Wide Web. pp. 697–706. WWW '07, ACM, New York, NY, USA (2007), http://doi.acm.org/10.1145/1242572.1242667
21. Tiscornia, D.: The lois project: Lexical ontologies for legal information sharing. In: Proceedings of of the V Legislative XML Workshop, European Press Academic Publishing, 2007 , http://www. e-p-a-p.com/dlib/9788883980466/art14.pdf. pp. 189–204
22. Valente, A.: Legal knowledge engineering: A modelling approach. Ph.D. thesis, University of Amsterdam (1995)
23. Vibert, H.J., Jouvelot, P., Pin, B.: Legivoc connecting laws in a changing world. Journal of Open Access to Law 1(1) (2013)