Enhancing domain-specific ontologies at ease

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

HAL Id: hal-01626620
https://hal.science/hal-01626620

Submitted on 31 Oct 2017

Enhancing domain-specific ontologies at ease

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

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

Abstract. We present a methodology to enhance domain-specific ontologies by (i) manual annotation of texts with the concepts in the domain ontology, (ii) matching annotated concepts with the closest YAGO-Wikipedia concept and (iii) using concepts from other ontologies that cover complementary domains. This method reduces the difficulty of aligning ontologies, because the alignment is carried out within the scope of an example. 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 of 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 OMG 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.

In this paper we present a method to enhance ontologies through alignment to other ontologies. Alignment of ontologies is a very challenging task because it is very abstract. 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. We propose to alleviate this difficulty via the annotation 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. Making it concrete, the task is much more natural for the annotator. Our aim is not to develop a reference ontology, but to focus on a useful, working mapping, based on naturally occurring examples, that will allow for a practical use of the ontologies in wide-coverage IE tasks.

We show that, by aligning domain-specific ontologies and the general-purpose ontology YAGO, we have the additional benefit of obtaining examples of mentions of those concepts in the text of the Wikipedia, which can be then used to train an automatic analyzer.

YAGO [17] is a knowledge base automatically extracted from Wikipedia, WordNet, and GeoNames, and linked to the DBpedia ontology[3] and to the SUMO ontology[4]. 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.

---

[3] http://wiki.dbpedia.org/
[4] http://www.adampease.org/OP/

We describe this method as applied to two subdomains of the legal domain, Court judgments and insurance customer service, and two different applications, Named Entity Recognition and Classification and Question Classification.

The rest of the paper is organized as follows: in the following section we describe some relevant work on the development of legal ontologies. Then, we outline our approach, with the following section detailing the annotation process. Then we describe two applications of this approach, one to the legal domain and another to finance.

2   Relevant work

A special class of ontologies are the legal ones which specify legal concepts in a formal way, such that reasoning mechanisms can then be exploited over such information. Many legal ontologies have been proposed in the literature with different purposes and applied to different sub-domains, e.g., [1, 12, 2]. Legal ontologies need to specify carefully the legal concepts highlighting possible conflicts among them and further subtle issues of the legal domain, and second, such ontologies have little coverage, i.e., they have a small number of entities and only very few annotated legal corpora exist where entities can be gathered from.

There exist few of ontologies to represent the legal domain. The Language for Legal Discourse [15] is not properly an ontology, but it provides a formalization of many legal terms and definitions, trying to define legal concepts for formal reasoning. LRI-Core [4] is intended as a core ontology for law, but it contains very few legal concepts. However, it is thoroughly based on principles of cognitive science, and its top structure is the base of LKIF. The Core Legal Ontology [10] organizes legal concepts and relations on a commonsense basis inspired by DOLCE+ [9]. The LegalRuleML ontology [2] aim to represent machine-readable legal knowledge, with a particular attention to legal sources, time, defeasibility, and deontic operators. Moreover, general-purpose ontologies usually contain some representation of the legal domain, but legal concepts are either not explicitly delimited or very few, or both.

In the literature, only few approaches addressed the problem of legal ontology population. More precisely, Bruckschen and colleagues [5] describe an ontology population approach to legal data, whose experimental evaluation is run over a corpus of legal and normative documents for privacy. The goal of this research is to provide a resource that can help software industry project managers to calculate, understand and lower privacy risks in their projects. Ontology population is then obtained through the task of NER. Lenci *et al.* [14] report an experiment on an ontology learning system called T2K. They use NLP and Machine Learning methods to extract terms and relations from free text. The experimental evaluation is conducted on Italian legal texts, and it is able to identify the classes of the ontology, as well as many hyponymy relations. Related approaches to legal ontology population are presented by Boella and colleagues [13, 3]. The former discusses the results of the classification and extraction task of norm elements in European Directives using dependency parsing and semantic role labeling. The experimental system takes advantage of the way the Eunomos system 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 [3] 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.

While the approaches in [5, 14] tackle the issue of legal ontology population, they differentiate from our approach regarding many aspects. The main difference with all the above mentioned approaches is the generality of the approach we propose in this paper, that can be easily adapted to any legal ontology and that shows good performance. Moreover, the goal of our approach, i.e., Named Entity Recognition and Entity Linking, and the populated ontologies respectively, are different.

3 Outline of the approach

Schematically, the 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 and either
   (a) tag them with the most specific concept in the domain ontology, if it exists, or
   (b) tag them with the most specific concept from another domain ontology, or
   (c) tag them with the most specific concept in YAGO or the Wikipedia.
3. find the the most specific concept in YAGO or, if the concept is not in YAGO, in the Wikipedia. Take into account that the most specific concept may be **the actual entity**.

When some equivalent concept has been found, we establish the alignment using the OWL primitives `equivalentClass` and `subClassOf`. We align classes, not relations.

After annotation, 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, establish the mapping between that concept and the most adequate ancestor of the selected YAGO node, as can be seen in the following example.

Example 31

domain-specific

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

YAGO

The $[Court]_{wordnet\_trial\_court\_108336490}$ is not convinced by the reasoning of the $[combined\ divisions\ of\ the\ Court\ of\ Cassation]_{wordnet\_trial\_court\_108336490}$, because it was not indicated in the $[judgment]_{wordnet\_judgment\_101187810}$ that $[Eitim\text{-}Sen]_{wordnet\_union\_108233056}$ had carried out $[illegal\ activities]_{wordnet\_illegality\_104810327}$ 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 the Wikipedia, and thus a mapping is effectively established between both. This mapping allows to transfer properties from one ontology to the other, like relations of the nodes, which is relevant for inference and reasoning.

Relevant for NLP applications like Named Entity Recognition and Classification (NERC) 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 [11] ontology. 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 [7], but adapted to 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:

Example 41 *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}$5 *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.
– 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 42 $[assurance\ individuelle\ scolaire]_{insurance}$ *de* $[John\ Smith]_{Person}$.

– Proper names are always tagged, even if they do not represent a legal concept.

   Example 43 *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.*

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

   Example 44 *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`, the resulting 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 2, the entity "Convention" in the text is annotated with the LKIF class `wordnet_convention_106774316***` and the YAGO URI `European_Convention_on_Human_Rights`.

(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.

4. Visualization of the annotated legal document: as shown in Figure 3, the resulting annotation is visualized by highlighting the annotated entities, and the ontology used for each annotation by means of different colors.
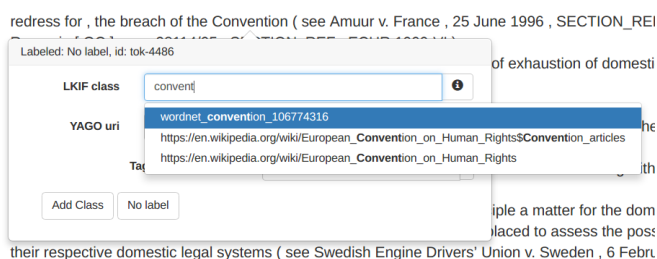

Fig. 1. The annotation of the entity *Convention* in LOAV.

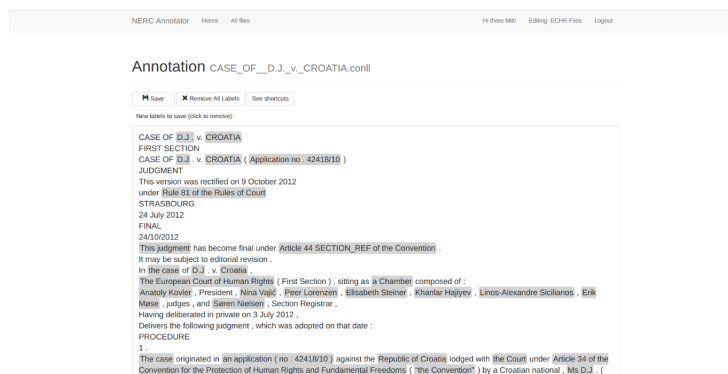**Fig. 2.** The annotation of the entity *Convention* in LOAV.

Fig. 3. Visualization of the resulting annotated text.

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 [12], to judgments of the European Court of Human Rights. This 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 [12] 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)[6] 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[7] 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 annotators independently, to assess inter-annotator agreement using Cohen's kappa coefficient [6]. 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

---

[6] `hudoc.echr.coe.int`

[7] `hudoc.echr.coe.int`

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[8] from March 2016, and we processed it via the WikiExtractor [16] 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 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 needs to be treated qualitatively differently by a Named Entity Linking approach.

We represented examples with a subset of the features proposed by Finkel *et al.* [8] 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.

---

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

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 undergoing.

As a reference ontology for this domain we used the Property And Casualty Information Models, Version 1.0[9], developed by the Insurance Working Group of the Object Modelling Group (OMG)[10]. 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 ask to customer service ("*Espace Client*") through the webpage of French branch of Allianz Insurance Group. 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) [11], again developed by the OMG group.

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

---

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

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

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

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, YAGO. 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, to two different domains, judgments of the Court and questions and answers of an insurance company customer service, and with two different target applications, concept recognition 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., Lesmo, L., Martin, M., Mazzei, A., Radicioni, D.P., Rossi, P.: Legal taxonomy syllabus version 2.0. IDT p. 9 (2009)
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., 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
4. Breuker, J., Hoekstra, R.: Epistemology and ontology in core ontologies: Folaw and lri-core, two core ontologies for law. In: In Proceedings of the EKAW04 Workshop on Core Ontologies in Ontology Engineering. pp. 15–27. Northamptonshire, UK (2004)
5. 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)
6. Cohen, J.: A coefficient of agreement for nominal scales. Educational & Psycological Measure 20, 37–46 (1960)
7. Consortium, L.D.: Deft ere annotation guidelines: Entities v1.7. http://nlp.cs.rpi.edu/kbp/2014/ereentity.pdf (2014)
8. 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
9. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with DOLCE. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, 13th International Conference, EKAW 2002, Siguenza, Spain, October 1-4, 2002, Proceedings. Lecture Notes in Computer Science, vol. 2473, pp. 166–181. Springer (2002), https://doi.org/10.1007/3-540-45810-7_18
10. Gangemi, A., Sagri, M.T., Tiscornia, D.: A constructive framework for legal ontologies. Law and the Semantic Web pp. 97–124 (2005)
11. 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)

12. 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)

13. 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)

14. 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)

15. McCarty, L.T.: A language for legal discourse i. basic features. In: Proceedings of the 2Nd International Conference on Artificial Intelligence and Law. pp. 180–189. ICAIL '89, ACM, New York, NY, USA (1989), http://doi.acm.org/10.1145/74014.74037

16. of Pisa, M.U.: Wikiextractor. http://medialab.di.unipi.it/wiki/Wikipedia_Extractor (2015)

17. 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