

# Using OpenWordnet-PT for Question Answering on Legal Domain\*

**Pedro Delfino**

FGV Direito Rio and EMap/FGV

**Bruno Cuconato**

EMAp/FGV

**Guilherme Paulino-Passos**

COPPE/UFRJ and IBM Research

**Gerson Zaverucha**

COPPE/UFRJ

**Alexandre Rademaker**

IBM Research and EMap/FGV

## Abstract

In order to practice a legal profession in Brazil, law graduates must be approved in the OAB national unified bar exam. For their topic coverage and national reach, the OAB exams provide an excellent benchmark for the performance of legal information systems, as it provides objective metrics and are challenging even for humans, as only 20% of its candidates are approved. After constructing a new data set on the exams and doing shallow experiments on it, we now employ the OpenWordnet-PT to verify whether using word senses and relations we can improve previous results. We discuss the results, possible future ideas and the additions to the OpenWordnet-PT that we made.

## 1 Introduction

Automatic analysis of legal content offers opportunities for improving the effectiveness of legal actors, transparency of the system and, ultimately, the welfare of the public. As law is practiced with language itself, linguistic approaches are invaluable. This focus on language and higher demand for precision created by a technical domain makes it natural to try to grow upon and evaluate the performance of a lexical-semantic resource, such as wordnets, in this area.

One task for legal technology is question answering: an automatic way of determining the right answer to a question presented in natural language form (Mitkov, 2005). An ideal legal question answering system would take a question in natural language and a corpus of all legal documents in a given jurisdiction, and would return both a correct answer and its legal foundation (answer justification), i.e., which sections

---

The authors would like to thank João Alberto de Oliveira Lima for introducing us to the LexML resources.

(or articles) of which norms provide support for the answer. Considering lack of knowledge about facts, incompleteness, inconsistency or disagreements between sources of law, an ideal system would generate each possible answer with corresponding arguments, explanations and confidence value. Since such a system is still far from our current capabilities, as the results of recent evaluation tasks such as ResPubliQA (Peñas et al., 2010) has shown, we started with a simpler task.

In Brazil, even after graduating from Law school, it is required that one is approved in the OAB exam in order to practice a legal profession. The “Ordem dos Advogados do Brasil” (Order of Attorneys of Brazil, OAB) is the professional body of lawyers in Brazil. The first stage of the exam is a multiple-choice test. We are interested in investigating the performance of simple methods in answering this test correctly, and providing justifications for its answers. We measure the impact of the usage of an open lexical resource such as wordnet, and also promote its expansion into the legal domain by demand. In particular, we use FreeLing (Carreras et al., 2004b) for linguistic analysis, and evaluate specially the usage of the word sense disambiguation (WSD) module (Padró et al., 2010), which in Portuguese, uses openWordnet-PT (de Paiva et al., 2012) (OWN-PT). We find that the system does not improve considerably over the performance of our previous effort (Delfino et al., 2017); however, this might be because of missing concepts and relations in OWN-PT, which in turn render some of Freeling’s processing inaccurate.

In Section 2 we present the data-set we created and made available for experimentation. In Section 3 we discuss our previous experiments with the data-set, while in Section 4 we describe the tools and resources we employed for our current experiment: Freeling, OWN-PT, and the word sense disambiguation algorithm UKB (Agirre and Soroa, 2009). In Section 5 we describe the meth-

ods used in our experiments and then discuss its results in Section 6. Finally, we conclude and debate future works in Section 7.

## 2 The OAB Exams data set

Among other responsibilities, OAB is responsible for the regulation of the legal profession in the Brazilian jurisdiction. One of the key ways of regulating the legal practice is through the “Exame Unificado da OAB” (unified bar examination), required for enrolling at OAB, which is mandatory to practice law.

In order to be approved in the OAB exam, candidates need to be approved in two stages. The first phase consists of multiple choice questions, while the second phase involves free-text questions. Since 2012, the first phase has 80 multiple choice questions and each question has 4 alternatives. Candidates are asked to choose the correct alternative and in order to be approved, candidates need at least a 50% performance. Historically, the exam has had a global 80% failure rate, with the first stage being responsible for eliminating the majority of the candidates (Amorim and Tebechrani Neto, 2016).

Thus, the first stage of the OAB exams provides an excellent benchmark for the performance of a system attempting to reason about the law. That is, passing the OAB exam would signal that the system has acquired important aspects of legal knowledge, up to a level comparable to a human lawyer. In trying to build such a system, it was necessary to create the appropriate data sets, which includes not only the questions and answer keys in machine readable format, but also the legal literature involved (Delfino et al., 2017).

In previous work (Delfino et al., 2017), we have obtained the PDF files of the all the previous OAB exams, extracted their text, cleaned them up and made the data freely available in a public repository<sup>1</sup>.

Along with the 1820 questions (from 22 exams) in plain text and in XML, it contains a golden set of 30 questions which were manually analyzed and annotated with the answer keys’ legal basis, i.e., which articles from which norms justify the correct answer to the question. These 30 questions are on a single subject, legal ethics.

Since 2012, the exams have revealed a pattern for which areas of Law the examination board fo-

cuses on and in which order the questions appear on the exam. Traditionally, the first 10 questions are about legal ethics, that is, the rights, the duties and the responsibilities of the lawyer in regard to Brazilian law. We have chosen to provide a golden set on legal ethics because this subject area is the simplest part of the exam with respect to the legal foundations of the questions. It also has a high frequency rate, and the highest performance rate among candidates (65%) (Amorim and Tebechrani Neto, 2016).

The key finding from analysis done in our previous work is that, usually, only one article on federal law no. 8906 was enough to justify the answer to the legal ethics questions (15 questions). Less often, in four questions, the justification was in “Regulamento Geral da OAB” (OAB General Regulation), or on the “Código de Ética da OAB” (OAB Ethics Code, 7 questions). Three other questions were justified by two articles in law no. 8906 each, and one question only in case law from the Superior Court of Justice about an article from the law no. 8906. Federal law no. 8906 has 89 articles, while the OAB general regulation has 169 articles, and the OAB ethics code has 66 articles.

### 2.1 Brazilian law texts

Another critical component of our data set is Brazilian legal norms in machine-readable format. This resource is essential for employing legal knowledge in answering the exam questions.

For the experiments made on the golden set, we needed the three normative documents (see Section 2) in a machine readable format. Moreover, we needed the documents in a format that preserved the original internal structure of the documents, i.e., the sections, articles, and paragraphs.

In order to obtain this data, we employed a legal document parser,<sup>2</sup> provided by the LexML project (de Oliveira Lima and Ciciliati, 2008). The LexML is a joint initiative of the Civil Law legal system countries seeking to establish open standards for the interchange, identification and structuring of legislative and court information. The goal is to convergence the national standards to international standardization of some instruments, such as URN-LEX, the use of XML formatting standards and the exchange of its metadata.

<sup>1</sup><http://github.com/own-pt/oab-exams>

<sup>2</sup><https://github.com/lexml/lexml-parser-projeto-lei>

The LexML parser, still in beta, receives as input a DOCX<sup>3</sup> file with the norm and outputs it in XML format, using the tags and the structure following the conventions of the LexML schema (de Oliveira Lima and Ciciliati, 2008). We had to make minor modifications in the three documents before submitting them to the parser; the XML files produced and the modifications made are available in our repository.

### 3 The previous work

Question answering in legal domain is a hard problem. In the last ResPubliQA evaluation task, the only system that dealt with Portuguese texts, the Priberam system, has the worst performance among the competitors, obtaining only 0.56 in the C@1 score (Peñas et al., 2010).<sup>4</sup>

In (Fawei et al., 2016) the authors report a textual entailment study on the US Bar exam material. In the experiment, the authors treat the relationship between the question and the multiple-choice answers as a form of textual entailment. Answering a multiple choice legal exam is a more feasible challenge, although it is still a daunting project without restrictions on the input form. That is the reason we have chosen in (Delfino et al., 2017) to restrict the domain to a single section of the OAB exams: legal ethics, one which is governed by only a few legal norms. In (Delfino et al., 2017), we conducted 3 experiments in question answering (section 5). In the first experiment, they tried to find the right answer between the multiple-choice alternatives. The last 2 were in shallow question answering (SQA), a form of question answering where a system retrieves documents that justify the already provided answer. They have adapted the methodology described in (Monroy et al., 2008; Monroy et al., 2009) to answer multiple-choice exams instead of closed-ended answers.

A range of issues on the texts of the questions of the exams was identified. Many of the problems are similar to the ones found in the US bar exams and described by (Fawei et al., 2016). For instance, some questions do not contain an introductory paragraph defining a context situation for the question. Instead of that, they have only meta comments (e.g. “assume that...” and “which of the following alternative is correct?”) followed by the

<sup>3</sup>The Microsoft Word editor format, commonly used for Brazilian legal documents.

<sup>4</sup>We were not able to obtain the article describing the Priberam system.

choices. Some questions are in a negative form, asking the examinee to select the wrong option or providing a statement in the negative form such as “The collective security order **cannot** be filed by...”. Moreover, some questions explicitly mention the law under consideration, others do not. Many questions present a sentence fragment and ask for the best complement among the alternatives, also exposed as incomplete sentences.

Even in the presence of such problems, our results in this previous work were not bad, given our system’s simplicity. But our initial approach also had its shortcomings: it could not distinguish successfully between two almost identical alternatives which differed only by few words (such as an alternative and its negation), nor could it treat related words in an appropriate manner. The former problem may require deep linguistic processing of the texts for properly obtaining the meaning of the utterances, while the latter can be partly tackled by the use of lexical resource such as the OWN-PT, as is done in this paper.

### 4 Freeling, OpenWordnet-PT and Word Sense Disambiguation

Freeling is an open source language processing library developed at the TALP research center<sup>5</sup> (Carreras et al., 2004a; Padró and Stanilovsky, 2012). It has support for many languages, including English, Portuguese, among others. It implements modules for tokenization, sentence splitting, morphological analysis, part-of-speech tagging, word sense disambiguation, parsing and other tasks. FreeLing distribution includes linguistic data for the supported languages provided by many different projects and collaborators: morphological dictionaries, gazettes, lexical-semantic resources etc. Particularly, for Portuguese, its word sense disambiguation (WSD) module relies on OWN-PT, an open freely available wordnet for Portuguese (de Paiva et al., 2012).

FreeLing implements a pipeline-based approach. After tokenization, sentence split and the morphological analysis and part-of-speech tagging, the user can choose to execute the WSD module to search for senses in Wordnet matching the lemma and part-of-speech tag of each word or multi-word expression. Every possible sense is returned and may be weighted by the sense disambiguation module. The disambiguation is

<sup>5</sup><http://nlp.cs.upc.edu/freeling/>

an implementation of the UKB algorithm (Agirre and Soroa, 2009), an unsupervised graph-based method which uses Personalized PageRank to select the right sense of each word in a lexical database such as OWN-PT.

Before running our experiment, we did a preliminary survey on the coverage of OWN-PT for the OAB corpus – a proxy for the legal domain as a whole. In Princeton Wordnet (PWN) (Fellbaum, 1998), the synset [08441203-n, *law/jurisprudence: the collection of rules imposed by authority.*] is a general concept about law, and is linked to hundreds of synsets via the `classifiesByTopic` relation. This suggests that PWN already covers (synset-wise) the relevant context, but it remained to be investigated whether such synsets are properly translated in OWN-PT with the relevant words, and if the existent concepts indeed encompass notions used in the Brazilian legal context, as legal jargon can be language and cultural dependant.

In order to further evaluate the coverage of the legal domain in OWN-PT we have taken a simple approach: after running Freeling on our corpus, we have listed the most common words whose senses Freeling could not find. We then proceeded to add them to OWN-PT. Some synsets did not seem to exist yet, such as one for “cartório” (notary office).<sup>6</sup> Other synsets existed, but the word at hand was not included in it, as in [06532763-n, *nulidade: nullity*]. Other cases were those of relations that did not exist in OWN-PT; if present, these relations would improve the results of the UKB algorithm. One such relation that we included in OWN-PT was the nominalization (morphosemantic link) between [00664276-v, *comprovar: authenticate*] and [06855035-n, *comprovação: authentication*]. In the end, since we focused only on the possible improvements to our immediate purpose, we have added to OWN-PT two synsets, eight semantic and lexical relations, and 25 words.

After running our experiment (to be described in the next sections), we also reevaluated the legal domain coverage in OWN-PT. To do so we looked at the difference between the questions answered and justified correctly by our previous system (Delfino et al., 2017) and the present one. One observation is that even when the WSD was not

<sup>6</sup>We will make the data available as part of the OWN-PT distribution available at <http://wnpt.br/1cloud.com/wn/>.

done correctly, as when a Portuguese word that should be in the synset [06532095-n, *ato: legal act*] was assigned to the synset [00037396-n, *act: as in action*], these mistakes were consistent, so that terms in both legal norm and OAB question had been given the same senses. Surely, that is not the most desirable outcome, but at least does not impose a problem for our experiments.

The question below and the first article from law no. 8906 following it illustrate cases where Wordnet resources are helpful and a more shallow approach could fail. Even though article and question alternative are related, this relation is not captured by our previous algorithm, because it does not take into account anything but the equivalence of tokens. Using OWN-PT, we can exploit the relationship between the action (sign, “visar”) and the result of the action (signature, “visto”).

Constitutive acts and contracts of legal persons, in order to be registered regarding the legal practice statute, must: [...] C) contain the lawyer’s [...] **signature**. (17th ed. OAB exam, question 2)

§ 2º The constitutive acts and contracts of legal persons can only be registered in the competent bodies, under a penalty of invalidity, when **signed** by lawyers. (law no. 8906, article 1)

In the example above, however, OWN-PT was missing the words “visar” and “visto” in the appropriate synsets: [00996485-v, *sign, subscribe: mark with one’s signature*] and [06404582-n, *signature: your name written in your own handwriting*]. These missing senses, of course, had to be created before being properly linked by the morphosemantic link *result*.

During our evaluation, we also had to make some changes in the Freeling dictionary, some adjectives and their lemmas and part-of-speech tags were introduced. An important attribute of this approach is that it propagates. Extending the Wordnet and giving the right senses for some words can improve the classification of other words that were not changed directly due to correct part-of-speech tagging and adequate linking between senses, tasks which depend on neighboring words. The missing words, synsets and links in OWN-PT is both a problem and an opportunity: in order to make better use of OWN-PT for the task at

hand one must further extend it to the legal domain (Sagri et al., 2004).

## 5 Experiment Setup

The original idea for the experiment was inspired by (Monroy et al., 2008), and it runs as follows: one collects legal norms in a corpus and preprocesses them performing tasks such as converting text to lower case, eliminating punctuation and numbers and removing stop-words. After that, the articles of the norms are represented as Term Frequency - Inverse Document Frequency (TF-IDF) vectors in a Vector Space Model (VSM) (Manning et al., 2008). In (Delfino et al., 2017), we have adapted this method to deal with exam questions with multiple choice alternatives. In the present article, we relied on Freeling to incorporate more linguistic processing in our pipeline.

We use the Freeling tokenizer, sentence splitting, morphological analyzer (POS tagging and lemmatisation), and the WSD modules to assign OWN-PT synsets, with a weight value (normalized in order to sum 1), to each token or sequence of tokens. For an input text we thus have a list of key-value pairs  $(s, w)$  with a sense key and a weight value, in contrast to a simple list of tokens, as we had in the previous experiment.

The intuition behind TF-IDF is that the more similar two text fragments are, the lesser is the distance between them. As the articles of the norms are not lists of tokens anymore, we have adapted the TF-IDF definition to deal with the weights assigned to each synset, as Equation 1 shows.

$$\begin{aligned} \text{TFIDF}_{s,w,d} &= \text{TF}_{s,w,d} \text{IDF}_{s,w,D} & (1) \\ \text{TF}_{s,w,d} &= \frac{f_{s,w,d}}{\sum_{s' \in d} f_{s',w',d}} \\ \text{IDF}_{s,w,D} &= \log \left( \frac{|D|}{\sum_{d \in D} w^{\mathbb{1}(w < 1)} \mathbb{1}(s \in d)} \right) \end{aligned}$$

where  $f_{s,w}$  is the sum of each occurrence of sense  $s$  weighted by  $w$ . Here  $\mathbb{1}_X$  is the characteristic function for  $X$ : 1 if  $X$  is true and 0 otherwise. An intuitive explanation is that, for TF, we count the weighted occurrence as a “continuous occurrence”, instead of boolean, where the degree of occurrence is the weight of the sense. For IDF, if the sum in a document is higher than 1, then it counts as an occurrence, which is counted only once. Otherwise, it counts only according to the weight received.

A directed graph is then created, with a node for each article of the used norms. This is the base graph, used for answering all questions. When provided a question-answer pair, our system processes the question statement and the alternatives in the same way as it does to the articles in the base graph: turning them into a list of  $(s, w)$  pairs. It then turns them into TF-IDF vectors using IDF values from the document corpus.<sup>7</sup> The statement node is connected to every article node, and each article node is then connected to every alternative node. In this we differ once more from (Monroy et al., 2008), as we have no need for heuristic rules for splitting the questions.

The edges are given weights whose value is the inverse cosine similarity between the connected nodes’ TF-IDF vectors. The system then calculates the shortest path between question statement and answer item using Dijkstra’s algorithm, and returns the article that connects them as the answer justification. Unlike (Monroy et al., 2008) our graph structure does not allow for more than one node connecting statement and alternative, as we knew from previous analysis that questions were usually justified by only a single article. Figure 1 illustrates the types of graphs we construct for each question.

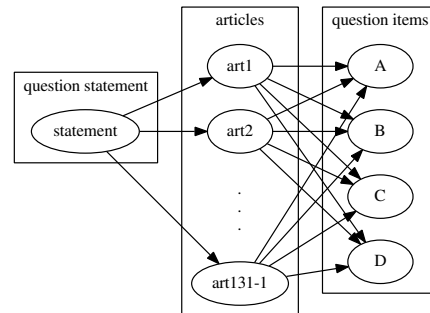


Figure 1: If a  $A$  is the number of article nodes, we then have  $5A$  edges (as we have one statement and four alternatives).

## 6 Results

Using the method described in section 5, we conducted two experiments. As we explained in section 2, our golden answer set was manually created

<sup>7</sup>This means that if a sense occurs in the question statement or alternative but not in the legal norm corpus, its IDF value is 0.

	QA	QA+J	J
word system	12	12	18
synset system	14	11	17

Table 1: Experiments results, number of right answers out of the 30 question-answer pairs from the golden data.

by one of the authors and it consists of 30 questions from eleven different editions of the OAB exam associated to the article on the appropriate norm that justify the answer of the question. Table 1 presents the results comparing the current system (“synset system”) to the previous system (Delfino et al., 2017) (“word system”).

Our first experiment aimed to evaluate the main task (QA): choosing the right answer at the multiple choice problem, given the questions and the laws (all three normative documents related to the legal ethics area). The performance of the synset system was of 14 questions, against 12 in the word system. If we require not only correct answer, but a correct justification as well, experiment (QA+J), the synset system achieves 11 correct answers, while the word system scores the same 12 correct answers.

In some cases, both systems would find the correct justification article for the correct answer, but would pick as their putative answer another (incorrect) item, because it had a shorter path. Other times, they would not be capable of deciding between two (or more) answer items, as they all had a shortest path of the same length. The following exam question is a sample case where this statistical approach to question answering is defective:

Concerning the expiration of punitive disciplinary infractions, choose the right alternative. [...] A) The punitive aim in regard to disciplinary infractions expires after **five** years [...] B) The punitive aim in regard to disciplinary infractions expires after **three** years [...]

(15th ed. OAB exam, question 4)

These two options differ by only one word (the number of years until expiration), and coincidentally both are in the text of the article which justifies the answer key. In the synset system, as “three” and “five” are both hyponyms of [13741022-n, digit: one of the elements that col-

lectively form a system of numeration], this difference shouldn’t interfere with WSD of the other words. This gives us almost the same distance between the question statement and these two answer choices, and the system is incapable of choosing between them. A similar situation arises when one answer item makes a statement and another item denies this statement:

[question statement] [...] A) does **not compel** him to pay the agreed upon legal fees. [...] B) does **compel** him to pay the agreed upon legal fees. [...]

(18th ed. OAB exam, question 1)

In a question like this a system can only systematically report a correct answer if it has a higher-level understanding of the texts at hand: no bag-of-words model will suffice.

Although results in the first and second experiments may be humble, we then considered shallow question answering. As our approach tries to find not only the correct answer, but to find through a justification, it’s reasonable to evaluate the ability to find the correct justification given the correct answer to the question. Therefore in our third experiment (J) the system’s task was to determine which article (considered every law it has seen) justified the (already given) answer to the sentence. For each question in our golden set, we again added its statement and correct answer as nodes connected to all article nodes in the graph (see Figure 1). The word system was able to find 18 while the synset system found 17.

The overall results are not very impressive, although they are not bad as well. Using part-of-speech tagging and word sense disambiguation in order to improve the use of TF-IDF does not solve important difficulties, such as compositional understanding, pragmatics, etc. Nevertheless, the contributions to OWN-PT can be seen as a benefit by itself and will be valuable in the future planned experiments. These contributions may also improve the synset system to the point that it outperforms the word system noticeably.

## 7 Conclusion and Future Works

We tested the coverage and improved OWN-PT with terms from the Legal Domain. We also presented a new data set with all Brazilian OAB exams and their answer keys jointly with three Brazilian norms in LexML format. Furthermore,

we also reported our findings in the course of constructing a system to pass in the OAB exams. We obtained reasonable results considering the simplicity of the methods employed and the limited golden data available.

For the next steps, many other ideas can be tested. The TF-IDF VSM approach was devised as a baseline for the next phases of the project. Even so, we can still explore variations on that approach with lemmas and edges between articles, considering that 10% of our golden set includes more than one article as justification. Moreover, such approach can be combined with other methods, following classical ideas such as (Hobbs, 1986), since it seems to be sufficient for solving many questions. In another direction, we need to increase the size of the golden set. Using crowdsourcing websites to obtain more justifications from humans or crawling data from websites dedicated to discussions about the OAB exams is likewise a possibility.

Many different proposals for encoding laws in a machine readable format are available. Why no single standard have been largely adopted yet? We aim to explore the best candidates for the remaining normative documents that we will need to cover all areas of the OAB exams. We can consider ideas used in the data preparation of the ResPubliQA editions (Peñas et al., 2010).

Other techniques for textual entailment could be used as well for the task of answering multiple choice questions. Given the legal information (such as statutes, regulations and case law) as background knowledge, inferring the correct answer would amount to selecting the item which is entailed by the question statement and background knowledge (in case of multiple entailed answers, the one with highest confidence). The results of the experiments presented here clearly show that we need ‘deep’ linguistic processing to capture the meaning of natural language utterances in representations suitable for performing inferences. That will require the use of a combination of linguistic and statistical processing methods, possibly using leveraging experiences from (Quaresma and Rodrigues, 2005). In (Delfino et al., 2017) we begin to explore the use of the logic called *i*ALC (de Paiva et al., 2010; Haeusler et al., 2010). *i*ALC can be used to represent legal knowledge and it may help in the next steps of our project.

We may also explore recent advances in statistical relational learning, specially combining probabilistic and logical methods for semantic tasks, such as done by (Beltagy, 2016; Beltagy et al., 2013). This approach uses syntactical parsing in order to construct a logical form, which is given probabilistic semantics, weighted by linguistic resources (e.g. Wordnet). Using probabilistic logics (such as Markov Logic Networks (Richardson and Domingos, 2006) and Probabilistic Soft Logic (Kimmig et al., 2012)) allows a semantic with clear support for vagueness and ambiguity, as well for a integrated use of lexical resources, hand-coded rules and information learned from the data itself. The base of this approach is general: logical forms could be encoded in different formalisms, such as *i*ALC or others intermediary semantic representation formats such as AMR (Banarescu et al., 2013), if suitable probabilistic semantics could be given.

## References

- Eneko Agirre and Aitor Soroa. 2009. Personalizing pagerank for word sense disambiguation. In Alex Lascarides, Claire Gardent, and Joakim Nivre, editors, *EACL 2009, 12th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, Athens, Greece, March 30 - April 3, 2009*, pages 33–41. The Association for Computer Linguistics.
- Flávia Alfenas Amorim and Gabriel Dib Tebechrani Neto. 2016. Exame da ordem em números. Technical report, Fundação Getulio Vargas and Conselho Federal da Ordem dos Advogados do Brasil. <http://hdl.handle.net/10438/18493>.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.
- Islam Beltagy, Cuong Chau, Gemma Boleda, Dan Grette, Katrin Erk, and Raymond J. Mooney. 2013. Montague meets markov: Deep semantics with probabilistic logical form. In Mona T. Diab, Timothy Baldwin, and Marco Baroni, editors, *Proceedings of the Second Joint Conference on Lexical and Computational Semantics, \*SEM 2013, June 13-14, 2013, Atlanta, Georgia, USA.*, pages 11–21. Association for Computational Linguistics.
- I. Beltagy. 2016. *Natural Language Semantics Using Probabilistic Logic*. Ph.D. thesis, Department

- of Computer Science, The University of Texas at Austin, December.
- Xavier Carreras, Isaac Chao, Lluís Padró, and Muntsa Padró. 2004a. Freeling: An open-source suite of language analyzers. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal*. European Language Resources Association.
- Xavier Carreras, Isaac Chao, Lluís Padró, and Muntsa Padró. 2004b. Freeling: An open-source suite of language analyzers. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*.
- João Alberto de Oliveira Lima and Fernando Ciciliati. 2008. LexML brasil: versão 1.0. available at <http://projeto.lexml.gov.br/>, December.
- Valeria de Paiva, Edward Hermann Hausler, and Alexandre Rademaker. 2010. Constructive description logic: Hybrid-style. In *Proc. HyLo'2010*.
- Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. 2012. OpenWordNet-PT: An open brazilian wordnet for reasoning. In Martin Kay and Christian Boitet, editors, *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Demonstration Papers, 8-15 December 2012, Mumbai, India*, pages 353–360. Indian Institute of Technology Bombay.
- Pedro Delfino, Bruno Cuconato, Edward Hermann Haeusler, and Alexandre Rademaker. 2017. Passing the brazilian oab exam: data preparation and some experiments. under review.
- Biralatei Fawei, Adam Z Wyner, and Jeff Pan. 2016. Passing a USA national bar exam: a first corpus for experimentation. In *Language Resources and Evaluation*, pages 3373–3378.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press.
- Edward Hermann Haeusler, Valeria de Paiva, and Alexandre Rademaker. 2010. Using intuitionistic logic as a basis for legal ontologies. In *Proceedings of the 4th Workshop on Legal Ontologies and Artificial Intelligence Techniques*, pages 69–76, Fiesole, Florence, Italy. European University Institute.
- Jerry R. Hobbs. 1986. Overview of the tacitus project. In *Strategic Computing - Natural Language Workshop: Proceedings of a Workshop Held at Marina del Rey, California, May 1-2, 1986*.
- Angelika Kimmig, Stephen Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2012. A short introduction to probabilistic soft logic. In *Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications*, pages 1–4.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. Introduction to information retrieval. *An Introduction To Information Retrieval*, 151:177.
- Ruslan Mitkov. 2005. *The Oxford handbook of computational linguistics*. Oxford University Press.
- Alfredo Monroy, Hiram Calvo, and Alexander Gelbukh. 2008. Using graphs for shallow question answering on legal documents. In *MICAI 2008: Advances in Artificial Intelligence: 7th Mexican International Conference on Artificial Intelligence, Atizapán de Zaragoza, Mexico, October 27-31, 2008 Proceedings*, pages 165–173. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Alfredo Monroy, Hiram Calvo, and Alexander Gelbukh. 2009. NLP for shallow question answering of legal documents using graphs. *Computational Linguistics and Intelligent Text Processing*, pages 498–508.
- Lluís Padró and Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*, pages 2473–2479. European Language Resources Association (ELRA).
- Lluís Padró, Samuel Reese, Eneko Agirre, and Aitor Soroa. 2010. Semantic services in freeling 2.1: Wordnet and ukb. In Pushpak Bhattacharyya, Christiane Fellbaum, and Piek Vossen, editors, *Principles, Construction, and Application of Multilingual Wordnets*, pages 99–105, Mumbai, India, February. Global Wordnet Conference 2010, Narosa Publishing House.
- Anselmo Peñas, Pamela Forner, Álvaro Rodrigo, Richard F. E. Sutcliffe, Corina Forascu, and Cristina Mota. 2010. Overview of respubliqa 2010: Question answering evaluation over european legislation. In *CLEF 2010 LABs and Workshops, Notebook Papers, 22-23 September 2010, Padua, Italy*.
- Paulo Quaresma and Irene Rodrigues. 2005. A question-answering system for portuguese juridical documents. In *10th international conference on Artificial intelligence and law*, pages 256–257. ACM.
- Matthew Richardson and Pedro Domingos. 2006. Markov logic networks. *Machine Learning*, 62(1):107–136, Feb.
- Maria Teresa Sagri, Daniela Tiscornia, and Francesca Bertagna. 2004. Jur-WordNet. In Petr Sojka, Karel Pala, Pavel Smrz, Christiane Fellbaum, and Piek Vossen, editors, *Global Wordnet Conference*.