Legal Knowledge and Information Systems
A. Wyner and G. Casini (Eds.)

doi:10.3233/978-1-61499-838-9-89

Passing the Brazilian OAB Exam: Data Preparation and Some Experiments[1]

Pedro DELFINO [a,b] Bruno CUCONATO [b] Edward Hermann HAEUSLER [c]
Alexandre RADEMAKER [b,d]
[a] *FGV Direito Rio, Rio de Janeiro, Brazil*
[b] *Applied Mathematics School of FGV, Rio de Janeiro, Brazil*
[c] *Departamento de Informática, PUC-Rio, Brazil*
[d] *IBM Research, Brazil*

Abstract. In Brazil, all legal professionals must demonstrate their knowledge of the law and its application by passing the OAB exams, the national Bar exams. This article describes the construction of a new data set and some preliminary experiments on it, treating the problem of finding the justification for the answers to questions. The results provide a baseline performance measure against which to evaluate future improvements. We discuss the reasons to the poor performance and propose next steps.

Keywords. OAB, bar exam, question-answering, justification, logic

1. Introduction

The "Ordem dos Advogados do Brasil" (OAB) is the professional body of lawyers in Brazil. Among other responsibilities, the institution 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). Only those who have been approved on this exam are allowed to work as practising attorneys in the country. In this way, the it is similar to the US Bar Exam. Thus, the OAB exam provides an excellent benchmark for the performance of a system attempting to reason about the law.

This paper reports the construction of the data set and some preliminary experiments. We obtained the official data from previous exams and their answer keys from `http://oab.fgv.br/`. As our first contribution, we collected the PDF files, extracted and cleaned up the text from them producing machine-readable data (Section 2). [2]

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, i.e., which sections of which norms provide support

[1] The authors would like to thank João Alberto de Oliveira Lima for introducing us to the LexML resources, and Peter Bryant for his careful review of the article. The extend version of this article is available in `http://arademaker.github.io`

[2] All data files are freely available at `http://github.com/own-pt/oab-exams`.

for the answer. Since such a system is far from our current capabilities, we started with a simpler task. In [4] the authors report a textual entailment study on US Bar Exams. 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, such as preprocessing natural language questions to make them more intelligible to the computer or restricting the legal domain. That is the reason we focused in the Ethics section of OAB Exams, one which is governed by only a few legal norms.

We have conducted three experiments in question answering (Section 3). To be able to provide the right justification for each question, we needed to have the text of the laws available. This is a particular challenge in the legal domain, as normative instruments are not readily available in a uniform format, suitable for being consumed by a computer program. Fortunately, using resources provided by the LexML Brasil project, we were able to collect and convert to XML format all the normative documents we needed (Section 2).

2. The data: OAB exams and norms

Before 2010, OAB exams were regional, only in 2010 were the exams nationally unified. In order to be approved, candidates need to be approved in two stages. The first phase consists of multiple choice questions and the second phase involves free-text questions. The first phase has 80 multiple choice questions and each question has 4 options. Candidates need at least a 50% performance.

Every year, there are 3 applications of the exam in the country. Concerning the exams statistics, the first phase is responsible for eliminating the majority of the candidates. Historically, the exam has a global 80% failure rate. Since 2012, the exams have revealed a pattern for which areas of Law the examination board focuses on and in which order the questions appear on the exam. Traditionally, the first 10 questions are about Ethics.

In the context of the OAB exam, Ethics means questions about the rights, the duties and the responsibilities of the lawyer. This is the simplest part of the exam with respect to the legal foundation of the questions. Almost all the questions on Ethics are based on the Brazilian Federal Law 8906 from 1994, which is a relatively short (89 articles) and well designed normative document. A minor part of the questions on Ethics is related to two other norms: (i) “Regulamento Geral da OAB” (OAB General Regulation, 169 articles) and (ii) “Código de Ética da OAB” (OAB ethics code, 66 articles). These two norms are neither legislative nor executive norms. Indeed, they are norms created by OAB itself. OAB’s prerogative to do so is assured by the Law 8906.

We obtained the exams files in PDF format and we converted them to text using Apache Tika [3]. The final data comprises 22 exams totaling 1820 questions. 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 Bar Exams and described by [4]. 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?”. Some questions are in a negative form, asking the examinee to select the wrong option or providing a statement in the negative form

[3] https://tika.apache.org/.

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.

We sampled 30 questions on Ethics for analysis (from the 210 questions in all exams) and one of the authors manually identified the articles in the laws that justify the answer, creating our golden data set. The key finding was that, usually, one article on the Law 8906 was enough to justify the answer to the questions (15 questions). Less often, the justification was not in the Law 8906, but rather in OAB Regulation (3 questions), or on the OAB Ethics Code (8 questions). Three questions were justified by two articles in Law 8906, and another in jurisprudence from the Superior Court of Justice about an article from the Law 8906.

For the experiments, we also needed the norms in a format that preserved the original internal structure, i.e., the sections, articles, and paragraphs. The LexML [2] is a joint initiative of the Civil Law legal system countries seeking to establish open standards for the interchange, identification and structuring of legal information. The Brazilian LexML project has developed a XML schema called “LexML Brasil” and it maintains a public repository at `https://github.com/lexml` with one useful tool for our project, the parser of legal documents. The software receives as input a DOCX file and outputs it in XML file, according the LexML schema.

3. The Experiments

We borrow ideas from [7] to construct a similar experiments that run as follows: one collects the legal norms and preprocesses them performing tasks such as converting text to lower case, eliminating punctuation and numbers and, optionally, removing stop-words. After that, the articles of the norms are represented as TF-IDF vectors in a Vector Space Model (VSM) [1].

A base graph is then created, with a node for each article of a norm and no edges. When provided a question-answer pair, our system preprocesses the question statement and the alternatives in the same way as it does to the articles in the base graph. It turns them into TF-IDF vectors using IDF values from the document corpus.[4] The statement node is connected to every article node, and each article node is then connected to every alternative node, creating a connected digraph.

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. The intuition behind such a method is that the more similar two nodes are, the lesser is the distance between them; as a document that answers a given query is presupposed similar to the question, it makes sense to retrieve the article in the shortest path between the statement and the alternative as a justification for the answer.

In our first experiment we had an ambitious objective: we had our system receive a question statement and its multiple alternatives, and we wanted it to retrieve the right

[4] This means that if a term occurs in the question statement or alternative but not on the legal norm corpus, its IDF value is 0.

answer along with its justification in the legal norm. When given the question and its alternatives, the system would add them to the base graph composed by the respective legal norm's articles. The system would return the shortest paths between the question statement and its alternatives, and the presumed justification would be the article connecting the statement and the closest alternative. The system's performance against this task was not impressive: although it chose the correct alternative 10 times, it only provided the correct justification for 8 of these.

Analyzing the system's output paints a more nuanced picture, however. In some cases, the system would find the correct justification article for the correct answer, but would pick as its putative answer another (incorrect) item, because it had a shorter path. Other times, it would not be capable of deciding between two (or more) answer items, as they all had a shortest path of the same distance. The following exam question is a sample case where this statistical approach to question answering is defective:

> The young adults Rodrigo (30-year-old), and Bibiana (35-year-old), who are properly enrolled in an OAB section [. . .] Considering the situation described, choose the correct alternative: A) Only **Bibiana** meets the eligibility criteria for the roles. B) Only **Rodrigo** meets the eligibility criteria for the roles. [. . .]

As one can see, these two options differ by only one word (the names of the fictional lawyers), and both are unlikely to be in the text of the legal norms, which means that they do not affect the calculation of similarity. A similar situation arises when one answer item makes a statement and another item denies this statement. 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.

As our first experiment demonstrated that our simple system could not reliably pick the correct answer among four alternatives, we turned our attention to shallow question answering (SQA), where our system would only have to provide the correct legal basis for the answer provided along with the question. In our second experiment, we built separate base graphs from each of the three norms. For each question in our golden set, we added its statement and its correct answer to the base graph created from the norm which contains the article that justifies it. The sole task of our system, in this case, is to the determine which article from the provided norm justifies the answer. In this simpler form, performance was not bad: the system retrieved the correct article in 21 out of 30 question-answer pairs.

In our third experiment, we tried to see if our system could provide the correct article from the appropriate legal norm without us telling it which norm it should consider. Following this idea, we have taken the articles from all norms and built a single base graph. 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, and then calculated the shortest path between them to retrieve the system's putative justification to the question-answer pair. The system now had to retrieve the correct article among articles from all norms – which, being in the same legal domain, had similar wordings and topics – therefore increasing the difficulty of the task. Despite this, its performance did not plunge: it scored the right article in 18 out of the 30 question-answer pairs.

4. A possible logic-based approach

One of the key observations that emerge from the results in Section 3 is the importance of logical reasoning for our final goal of constructing a system to pass the OAB exam with a full understanding of the questions and laws. For the future, we aim to investigate how to enrich the data with lexical information and syntactic dependencies as an intermediary step toward a semantic representation of the questions and laws statements. Nevertheless, we have to decide what should be an adequate logic language to represent laws and the deep semantics from the text statements. Since the adequacy of a logic language can be evaluated even before a procedure to obtain logic expressions from natural language texts is developed, we present some preliminary discussion about one possible logic.

In [5] we discuss how Kelsen's [6] pure theory of law points out a framework that takes into account the legal knowledge forming a collection of individual, legally valid statements. Thus, each legally valid statement may be seen as an inhabitant among the many individual laws of the represented legal system. The natural precedence existing between individual legal statements can be taken as a pre-order relation on the legal statements. The legal principle that rules the stability of the law implies that the precedence of individual laws preserves properties (decisions, conditions of applicability, adequate fora, etc) regarding them. In the presence of this natural precedence order between legally valid statements, the intuitionistic interpretation of subsumption between concepts A and B ($A \sqsubseteq B$) reflects more adequately the structure of existing legal systems than its classical interpretation counterpart.

To illustrate the use of *i*ALC for reasoning over the OAB exams questions, let us consider the translated question and its correct alternative:

> Three friends graduated in a Law School in the same class: Luana, Leonardo, and Bruno. Luana, 35 years old, was already a manager in a bank when she graduated. Leonardo, 30 years, is mayor of the municipality of Pontal. Bruno, 28 years old, is a military policeman in the same municipality. The three want to practice law in the private sector. Considering the incompatibilities and impediments to practice, please select the correct answer. [...] C) The three graduates, Luana, Leonardo, and Bruno, have functions incompatible with legal practice. They are therefore prohibited from exercising private practice. (CORRECT) [...]

The justification of the answer to this question is obtained in the Law 8906, article 28. [5] The relevant fragments of this article, translated into English, are:

> Legal practice is incompatible, even for self-defense, with the following activities: I - head of the Executive and members of the Bureau of the Legislative Branch and their legal substitutes; [...] V - occupants of positions or functions linked directly or indirectly the police activity of any nature; [...] VIII - occupiers of management positions in financial institutions, including private ones. [...]

In *i*ALC, the Law 8906 is formalized as a concept defined as the intersection of the concepts from its articles, that is, $Law_{8906} \equiv Art_1 \sqcap \ldots \sqcap Art_{28} \sqcap \ldots \sqcap Art_{87}$. Article 28 in turn is also further formalized as the intersection of the concepts from its paragraphs, $Art_{28} \equiv P_1 \sqcap P_2 \ldots$. The paragraph VIII is formalized by the two

[5] The complete text can be found at `http://bit.ly/29gZc83`

concepts $Lawyer \sqsubseteq \neg Financial$ and $Financial \sqsubseteq \neg Lawyer$. Paragraph V is formalized by $Lawyer \sqsubseteq \neg Police$ and $Police \sqsubseteq Lawyer$. Finally, paragraph I by $Lawyer \sqsubseteq \neg ChiefCouncil$ and $ChiefCouncil \sqsubseteq \neg Lawyer$. The *Lawyer* concept can be read as the set of valid legal statements (VLS) about lawyers. That is, each concept can be thought as the set of VLSs where it *holds*. From the statements of the question, we have the hypotheses $lual\colon Laywer$ (Luana acts as lawyer), $leoal\colon Lawyer$ and $bal\colon Lawyer$. Using the deductive system for *i*ALC [5], we can prove that Luana, Bruno and Leonardo can not act as lawyers.

$$\dfrac{\dfrac{\dfrac{lual\colon Police \qquad Police \sqsubseteq \neg Lawyer}{lual\colon \neg Lawyer} \qquad [lual\colon Lawyer]}{lual\colon \bot}}{\neg(lual\colon Lawyer)}$$

5. Conclusion and Future Works

We 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 presented some preliminary experiments with the goal of constructing a system to pass in the OAB exam. We obtained reasonable results considering the simplicity of the methods employed. For the next steps, we can construct the TF-IDF vectors using lemmas of the words, possibly increasing the similarities. We can also add edges between articles, considering that 10% of our golden set includes more than one article as justification. We also plan to use the OpenWordnet-PT [3], properly expanded with terms of the legal domain.

Finally, 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. The final objective is to obtain formal representations, encoded in *i*ALC or another variant, from the texts ready for formal reasoning.

References

[1] D Manning Christopher, Raghavan Prabhakar, and Schütze Hinrich. Introduction to information retrieval. *An Introduction To Information Retrieval*, 151:177, 2008.

[2] João Alberto de Oliveira Lima and Fernando Ciciliati. LexML brasil: versão 1.0. available at `http://projeto.lexml.gov.br/`, December 2008.

[3] Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. OpenWordNet-PT: An Open Brazilian WordNet for Reasoning. In *Proceedings of 24th International Conference on Computational Linguistics*, COLING (Demo Paper), 2012.

[4] Biralatei Fawei, Adam Z Wyner, and Jeff Pan. Passing a USA national bar exam: a first corpus for experimentation. In *Language Resources and Evaluation*, pages 3373–3378, 2016.

[5] Edward Hermann Haeusler, Valéria de Paiva, and Alexandre Rademaker. Intuitionistic logic and legal ontologies. In *Proc. JURIX 2010*, pages 155–158. IOS Press, 2010.

[6] Hans Kelsen. *General theory of norms*. Oxford Univ. Press, USA, 1991.

[7] Alfredo Monroy, Hiram Calvo, and Alexander Gelbukh. 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, 2008.