December 2018

Utilizing iALC to Formalize the Brazilian OAB Exam

Bernardo ALKMIM [a,1] Alexandre RADEMAKER [b,c] and Edward HAEUSLER [a]
[a] *Departamento de Informática, PUC-Rio, Brazil*
[b] *Applied Mathematics School of FGV, Rio de Janeiro, Brazil*
[c] *IBM Research, Brazil*

Abstract. In Brazil, the national Bar exams for legal professionals are called the OAB exams, which indicate aptitude for practice of the law. The logic iALC is an intuitionistic description logic which was created to deal with the world of laws. In this article we convey some experiments on the use of logical deductions as a kind of intermediate data structure to help in the task of explaining "juridicial sentences" on the basis legal systems in "Civil Law" states. The deductions are concretely developed in the iALC description logic for representing legal systems and reasoning inside them. We then discuss what can be extrapolated from that which was found and propose next steps.

Keywords. OAB, description logic, intuitionistic logic, formalization, Brazilian law

1. Introduction

The OAB (*Ordem dos Advogados do Brasil*, the Order of Attorneys of Brazil) Exam can be considered an excellent choice for analyzing the performance of a system with the goal to reason in the legal world. Similarly to the US Bar Exam, it evaluates whether the candidates have the aptitude to practice the law.

In order to reason on legal texts and on the questions of the exam, apart from the more obvious necessity for natural language processing (NLP), one also needs a way to express deduction. From this rises the need for the usage of logic.

Representing law in logic is not a new concept. There have been many approaches at formalizing the concept of law and legal statements, the most known being Deontic Logic, in which laws are propositions. One major problem with Deontic Logic is how it deals with Contrary-to-Duty paradoxes, as explained in [4]. It also faces Jørgensen's dilemma (stated in [7]), since norms are imperative statements.[2] Hansen et al. (in [5]) present and explain some problems with it.

In this article, we propose the usage of a Description Logic, iALC, based on the canonical ALC, whose characteristics will be explained in Section 3, for playing the deductive component of Law formalization.

[1] The authors gratefully acknowledge financial support from CAPES.

[2] The dilemma deals with arguments on imperative statements, and one of its results is that, if one allows the statements to have validity, one cannot treat them as mere propositions.

In Section 2 we will explain in further detail the structure of the OAB Exam, and how we adapt its questions in order to make the deductions. We explain more about iALC (Section 3), focusing on its practical usage. Finally, in section 4, we reach the main goal of this article, which is to show how the questions in the exam and the Brazilian laws are formalized in iALC.

2. The OAB Exam

The OAB Exam evaluates the aptitude of the candidates to practice law in Brazil, having been unified through all national territory since 2010.[3] It occurs three times per year, and is divided in two parts: multiple choice questions, and free-text questions (which will be ignored for the sake of this research due to its higher complexity and scope when one considers a NLP perspective).

Due to being renowned nationally and requiring candidates to make deductions according to their interpretation of the law, the OAB Exam forms a very interesting benchmark for the testing of a logic that deals with the world of laws. Another characteristic in its favor is the fact that there is **public** access to previous exams, and for which we already know the correct answers. We have, now, hundreds of questions in our disposal. This allows us to have always a goal in mind when utilizing iALC on the Exam.

But, even though it has a large base of usable questions from previous exams for our peruse, there is still a lot of *data curatorship* problems in both how the questions and the Brazilian laws themselves are formatted. Delfino et al. (in [2]) show ways to deal with these problems, as well as where is stored the database in which they are being dealt with.

3. iALC

Description Logics (DL) were already proposed for semantic analysis of natural language utterances in [1]. In [2] we have used the DL iALC [4] as the basis of a solver/reasoner to the questions in the exam. The iALC is based on the canonical ALC, and was by conception made to deal with law systems. In iALC, valid legal statements (VLS) are not propositions, but individuals in a legal ontology, i. e. one law cannot be true or false, it just either exists or not in a concept. The propositions we consider are some of the *concepts* of the ontology (i. e. concepts of the form $a : C$ or $C \sqsubseteq D$, for a nominal and C, D concepts), which represent the legal systems that can hold different kinds of laws. Also, VLSs have a precedence relationship, derived by the hierarchy of individual laws of Kelsenian jurisprudence, the basis for iALC.

Thus, one can make a concept *Attorney* to represent every VLS for attorneys, for instance. Then, the (legal) individuals for which this concept is valid will be interpreted semantically as the worlds in which *Attorney* holds. These individuals must not be viewed directly as real, physical attorneys, though.

In practice, legal individuals can be viewed as specific combinations of laws that end up representing an artificially created legal being. For instance, we can suppose that *John Doe* has passed the OAB Exam, and now is an attorney. So, there is a legal document

[3] More on the exam (in Portuguese): `https://www.oab.org.br/servicos/examedeordem`.

which says exactly that, and we can represent this as $j_a : Attorney$ (j_a representing our artificial abstraction over the fact that *John Doe* is a lawyer). But, as a Brazilian citizen, he has a birth certificate as well. Supposing we have a concept representing every legal statement valid for Brazilians, say, *BrCitizen*, we can also conclude that we can have $j_b : BrCitizen$ (j_b being a **different** legal statement, representing the fact that he is Brazilian). Now the precedence rules for legal individuals of iALC start to appear explicitly. Since, according to [4], from the intuitionistic aspect of the logic, a Kripke model in it is a Heyting algebra, every pair of worlds has a finite *meet* (say, j_c), related by *precedence* of laws to the others, in which are valid both the concepts *Attorney* and *BrCitizen*. In other words, given j_a and j_b, there is always a world j_c such that $j_c \preceq j_a$ and $j_c \preceq j_b$.

Note that there is no world (legal individual) that represents fully the physical individual *John Doe*, since all that we deal with are laws and legal statements. From this, we can conclude that, conceptually, we never deal with people nor objects from the real world, only with legal statements over them. This is an extremely vital point, especially when dealing with the OAB Exam, whose questions tend to always contextualize with (real) individuals, and our first instinct is to deal directly with the individuals met in those questions instead of abstractions over them. This will be better explained with the examples of Section 4.

iALC was tailored specifically to avoid contrary-to-duty (CTD) paradoxes, which happen all the time when dealing with Deontic Logic (instead of Description Logic) since it views laws as propositions. These paradoxes are avoided by utilizing an intuitionistic approach to the logic, instead of a traditional (classical) one. Intuitionistic logic is constructive by nature [6]. When, for instance, a formula A is said to be true, it means that there is a *proof* of A, i.e. A can be *shown*, *constructed*. From that, the semantics of the negation ($\neg$) change from those of classical logic. Intuitionistically, $\neg A$ being true can be understood as there being no way to construct a proof of A (or that, by assuming A to be true, one is led to a contradiction). With this, we have (among others) that $\neg\neg A$ being true does not imply in A being true, which is a valid implication in classical logic. In fact, intuitionistic logic is weaker than classical logic (see [6]), but it allows for non-conflicting existence of otherwise thought-to-be contradictory logical formulas in Description Logic, particularly.

This *paracomplete*[4] aspect of the negation in intuitionistic logic allows iALC to deal with different legal systems that, despite being related, have different sets of valid logical formulas. To say, for instance, that a VLS i is in a concept $\neg A$ (notated by $i : \neg A$) actually means that there cannot be legal support from i for any other VLS in A, and not that i is in some kind of complementary legal system of A. This, combined with the precedence of laws, allows iALC to more faithfully interpret the subjunction between different legal concepts ($A \sqsubseteq B$).

[4] A logic is said to be paracomplete if it does not accept the law of the excluded middle (i. e. $A \vee \neg A$) as a tautology. This law is equivalent to the already mentioned $\neg\neg A \rightarrow A$. Please note that the usage of *law* in this footnote differs from that of the rest of the article.

4. Utilizing iALC

The first example shows a more artisanal (dare we say, more naïve) way of formalizing and reasoning in order to answer the question itself, and should be seen as a stepping stone to the second example. It showcases a couple of different ways of viewing the alternatives to the question at hand, serving as an introduction to the reader to the inner workings of iALC in practice.

In the second example we show another way of formalizing the usage of iALC, making more explicit the legal individualization of the question itself via Kripke semantics like the *John Doe* example in Section 3. Its formalization is to be viewed as an evolution of the previous example, as a means of showing the improvements that have been made an can be done, at least when considering finding a more procedural and general way of formalization.

In both the examples we aim to show just the logical modeling and reasoning of the problem, so the previous parts of processing the question, finding the law(s) which is(are) related, marking the correct option, and selecting the related common knowledge (as a *TBox*) are assumed. In general, our goal is to construct deduction trees for the correct answers and counter-models for the wrong ones, but depending on the question, both approaches are possible for correct and incorrect answers.

4.1. First Example: Brands

In this example we make use of a natural deduction system for iALC [3], based on the sequent calculus introduced in [4]. The question is a (rough) translation of Question 50 from the OAB Exam of 2012, in the area of Business Law. The subject is *branding*, which is related to Law 9279.[5] As shown in [2], this law can be defined as the intersections of the concepts from the articles, i.e. $Law_{9279} = Art_1 \sqcap Art_2 \sqcap ... \sqcap Art_{244}$. The same happens to the articles with their respective paragraphs. The question is as follows:

> Regarding brands, which option is the correct one?
>
> A) Brand of high renown is a synonym of notoriously known brand.
> B) The period of validity of the register of a brand is 5 years, renewable by equal and successive periods of time.
> C:CORRECT) The assignment of brand registration request is allowed, if the transferee meets the legal requirements.
> D) The brand of product or service is the one utilized to identify products or services from members of a set entity.

Option `A` is related to articles 125 and 126, which can be formalized as $HiRe \sqsubseteq BrBrand$ (*a brand to be considered of high renown has to be registered in Brazil*) and $NoKn \sqsubseteq (BrBrand \sqcup \neg BrBrand)$ (*a brand can be considered notoriously known having been registered in Brazil or not*), respectively. Our intent is to show that these concepts are, in fact, not synonyms. One can go a few ways in order to do so, either by generating a counter-model, or a deduction tree (it cannot be a proof since we still have assumptions).

We want to show that a brand can be notoriously known and not of high renown (which constitutes as a counterexample to option `A`). We then assume a non-Brazilian

[5] Original text in `http://www.planalto.gov.br/ccivil\_03/Leis/L9279.htm`.

and notoriously known brand. Then, $b : \neg BrBrand$ and $b : NoKn$ are our premises, $\neg BrBrand \sqsubseteq \neg HiRe$ comes from the counter-positive of what is in article 125, and our goal is to reach $b : (\neg HiRe \sqcap NoKn)$, thus separating the concepts of *of high renown* and *notoriously known*. Here is an example of a deduction tree:

$$\dfrac{\dfrac{b : \neg BrBrand \quad \neg BrBrand \sqsubseteq \neg HiRe}{b : \neg HiRe} \qquad b : NoKn}{b : (\neg HiRe \sqcap NoKn)}$$

An alternative approach would involve generating a counter-model, which would involve generating different worlds related by the *logical precedence* between laws presented before. There would be a world with our premises and assumptions, and that world would precede a world in which there would be $\models NoKn$ and $\not\models HiRe$ coexisting (among others, which do not matter for the sake of the counter-model).

Option `B` is related to article 133, which can be formalized as $p : Ten$, for a register r of a brand (*the period of validity of the register of a brand is 10 (ten) years, renewable by equal and successive periods of time*). To us this clearly contradicts what is written in `B` ($p : Five$, for a period p) since 10 is a different number than 5. In iALC, Ten and $Five$ will represent the concepts whose individuals have validity for 10 and 5 years, respectively.

Here, we see an instance where it is needed to utilize a *TBox* to express this knowledge of the world, formalized as the disjunction of the concepts, i.e. $Ten \sqsubseteq \neg Five$ and $Five \sqsubseteq \neg Ten$, since despite it not being directly related to the law or the question, it can be considered common knowledge (i.e. something that is renewed every five years clearly is not renewed every ten years and vice-versa).

We want to show that it is not the case that a register r of a brand lasts for 5 years: $\neg (r : Five)$. Our premises are $r : Ten$ comes from article 133, and $Ten \sqsubseteq \neg Five$ comes from our *TBox* knowledge. Here is the deduction tree:

$$\dfrac{\dfrac{\dfrac{r : Ten \quad Ten \sqsubseteq \neg Five}{r : \neg Five} \qquad [r : Five]^1}{r : \bot}}{\neg (r : Five)}\ 1$$

Were we to go for a counter-modelling approach, we would create a model with a world holding our assumptions and premises, which would precede a world with $\models Ten$ and $\not\models Five$ coexisting.

Option `C` is related to article 134, and it is the correct option. It states almost word for word what is written in the article, thus making the proof a simple tautology. Since it is correct, it would be impossible for a counter-model to be created.

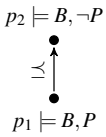
Article 134 (or option `C` itself) can be formalized as $LegalRequire \sqsubseteq BrRegRequest$, and leads to the following deduction tree:

$$\dfrac{}{LegalRequire \sqsubseteq BrRegRequest}\ Tautology$$

Option `D` is related to the first paragraph of article 123, which states that *brand of product or service is the one that distinguishes a product or service from another identical one, from a different origin*. It can be formalized as having $p_1 : (B \sqcap P)$ and $p_2 : (B' \sqcap P)$ implying in $p_1 \not\trianglelefteq p_2$, B and B' being brands (or, more accurately, the concept representing VLSs for brands B and B', respectively), p_1 and p_2, products, and P any legal statement holding when related to a product.

The question can be formalized as having $p_1 : (B \sqcap P)$ and $p_2 : (B \sqcap \neg P)$ implying in $p_1 \not\preceq p_2$. Notice how it does not utilize B', since it deals only with the same entity.

The generation of the proof for this question is trickier, and the deduction tree would be a bit larger than the space we can utilize here, thus making us having to build up a counter-model. The key part here is to show that, with our premises, it would be impossible to differ one product of a brand from another product from another brand, even if they both have the same defining properties. It would suffice for us to show that a world p_1 in which we have B and P (i. e. $p_1 \models B, P$ or $p_1 : B$ and $p_1 : P$) would precede a world p_2 in which we would have $p_2 \models B$ and $p_2 \not\models P$ coexisting, making it explicit in the model that the law cannot discern between the products, since one precedes the other.



4.2. Second Example: Deise

For this question, we attempted a more procedural way of interpreting and formalizing the question and the law. When formalizing the law, we always deal only with concepts in a more general way (i. e. without individuals), since the legal text is, by nature, generic. This was the same as before. But now, for the questions and answers we look for the syntactic chunks, and each of those will lead to a formula of the form $subject : Object$ (with $subject$ being a legal individual over a legal fact usually representing the syntactic subject of a sentence, and $Object$ being the concept of the VLSs regarding the object of the sentence). With that in hands, our focus will be to find the *relevant* chunks, and utilizing them to create counter examples (via the generation of *meets* of individuals) in order to find the correct answers to the questions. For the sake of this example, we will consider only the already relevant syntactic chunks.

This question is a (rough) translation of Question 6 from the OAB Exam of 2015, in the area of Ethics. The subject is related to Law 8906.[6] As before, this law can be defined as the intersections of the concepts from the articles, i.e. $Law_{8906} = Art_1 \sqcap Art_2 \sqcap ....$ The same happens to the articles with their respective paragraphs. The question is as follows (main syntactic chunks after a potential text processing step are underlined):

> Deise is an Attorney and *[she]* was elected State Representative. Due to her skills, she was then elected to be part of the Directory of the Legislative Assembly of her state. She willed to conciliate this work with her law practice. According to the Legal Profession Bylaws:
>
> A) Deise's parliamentary activity is incompatible with the *[her]* practice of the law.
> B : CORRECT) Her participation in the Directory is incompatible with the *[her]* practice of the law.
> C) *[her]* Being part of the Directory of the Legislative Assembly is compatible with the *[her]* practice of the law.
> D) Deise's parliamentary activity at the Directory can be compatible with the *[her]* practice

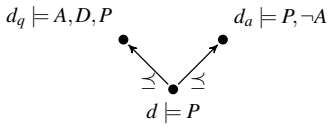
[6] Original text in `http://www.planalto.gov.br/ccivil_03/Leis/L8906.htm`.

of the law on behalf of those in need.

All the options are contemplated by Article 28. The law states that *The practice of law is incompatible with the following activities: I - head of the Executive Power and members of the board of Legislative Power and their legal substitutes...*, and can be formalized as $Attorney \sqsubseteq \neg Directory$ and $Directory \sqsubseteq \neg Attorney$,[7] meaning that those individuals for which *Attorney* is valid cannot precede individuals for which *Directory* is valid, and vice-versa.

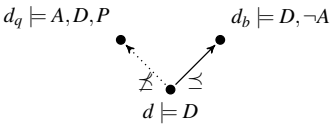
We can formalize the question preamble of the question as $d_q : Attorney$, $d_q : Parliamentarian$, and $d_q : Directory$, d_q representing our (partial) legal knowledge about *Deise*.

Option `A` can be formalized as $d_a : \neg Attorney$ and $d_a : Parliamentarian$. This leads to a *meet* d of d_q and d_a which has *Parliamentarian* as a valid concept. For the sake of readability, in the shown models the concepts will be reduced to their initials.



Since d precedes a world which has *Attorney* as a valid concept, we can conclude that it is possible to be a parliamentarian and an attorney, thus making option `A` false.

Option `B` is the correct answer to the question, and is the only one which cannot produce a counter model. It can be formalized as $d_b : Directory$ and $d_b : \neg Attorney$ (alternatively, the negation can be on *Directory* and the reasoning will be the same). A *meet* between d_b and d_q would have *Directory* as a valid concept, preceding the valid *Attorney* in d_q, clearly contradicting what is stated in Law 8906.



Option `C` can be formalized as $d_c : Directory$ and $d_c : Attorney$. It is a direct negation of option `B`, and one can verify that is is false by a simple deduction tree (notice that all our assumptions are part of either the law, the question, or option `C` itself):

$$\dfrac{\dfrac{d_c : D \quad D \sqsubseteq \neg A}{d_c : \neg A} \qquad d_c : A}{d_c : \bot}$$

Since one option was the direct negation of the other, there is no need to generate a counter model for the second one, since it is clear that it is not the right answer. Also, it is clear that in cases such as this one, once two options can be viewed as stating clearly opposite statements, the right answer **has** to be one of them. This is an indicator for us to

[7] The concept *Directory* is valid in all VLSs about members of directories of legislative assemblies and the concept *Attorney* is valid in all VLSs about attorneys.

first formalize all options at once and then deal with them individually, instead of trying to reason before formalizing everything.

Option `D` can be formalized as $d_d : Directory$ and $d_d : Attorney$, since there is no exception to the legal text as to the part that refers to *on behalf of those in need*. The justification is the same as the one for Option `C`.

5. Conclusion

In this article we started by explaining the basic structure of the OAB Exam, in order to familiarize the reader with the later sections. Then, we highlighted the more practical applicability of the Intuitionistic Description Logic iALC, as well as shown our initial successful attempts at utilizing it to formalize multiple choice questions of the OAB Exam. The formalization and deduction techniques are being refined, and we intend to make them as general and algorithmic as possible, for, in the future, start planning for an automatic reasoner for iALC (which currently has no specific reasoning tools available for use) in order to utilize on the Exam. As our methodology for formalization starts to take shape, we will evaluate which existing tools for DLs in general may be used as a framework or base for our reasoner.

The reasoner will be only part of a greater project, though. As the reader may have noticed, we make a lot of assumptions about the *TBox* and assume that the NLP part is already successfully done. Our current concerns involve creating each of these parts in a coherent way in order to make them as cohesive as possible with one another.

Finally, we would like to state our awareness of the possible size of this project, but we are confident that as time goes by and our expertise of the subject evolves, our efforts will bear good fruits in the near future, maybe even paving the way for a related approach not involving necessarily legal texts.

References

[1] Cleo Condoravdi, Dick Crouch, Valeria de Paiva, Reinhard Stolle, and Daniel G. Bobrow. Entailment, intensionality and text understanding. In *Proceedings of the HLT-NAACL 2003 Workshop on Text Meaning - Volume 9*, HLT-NAACL-TEXTMEANING '03, pages 38–45, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.

[2] Pedro Delfino, Bruno Cuconato, Edward Hermann Haeusler, and Alexandre Rademaker. Passing the brazilian oab exam: Data. *Legal Knowledge and Information Systems*, page 89, 2017.

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

[4] Edward Hermann Haeusler and Alexandre Rademaker. On how kelsenian jurisprudence and intuitionistic logic help to avoid contrary-to-duty paradoxes in legal ontologies. *Proc. Journal of Applied Non-Classical Logics*, 2016.

[5] Jörg Hansen, Gabriella Pigozzi, and Leendert van der Torre. Ten philosophical problems in deontic logic. In Guido Boella, Leon van der Torre, and Harko Verhagen, editors, *Normative Multi-agent Systems*, number 07122 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2007. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[6] A. Heyting. *Intuitionism: An Introduction*. North-Holland Publishing, Amsterdam, Netherlands, 3rd edition, 1956.

[7] Jørgen Jørgensen. Imperatives and logic. *Erkenntnis*, page 7: 288–96, 1937–38.