Towards Checking Laws' Consistency through Ontology Design: The Case of Brazilian Vehicles' Laws

Fred Freitas[1], Zacharias Candeias Jr[2] and Heiner Stuckenschmidt[3]

[1] Federal University of Pernambuco, Informatics Center, Recife, Brazil, fred@cin.ufpe.br
[2] Information Technology Agency for the State of Pernambuco, Recife, Brazil,
zacharias.candeias@ati.pe.gov.br
[3] University of Mannheim, Germany, heiner@informatik.uni-mannheim.de

Abstract

Official documents, and particularly legal ones like law codes, often contain ambiguities and/or inconsistencies, due to linguistic problems like polysemy, as well as ontological problems like underspecification, disagreements and/or false agreements. Such problems can be identified by formalizing the terminology of a domain in terms of an ontology. We show this phenomenon in a particular domain, the definition of different classes of vehicles. Defining accurately these different vehicle types shed light on some of these semantic deficiencies present in two Brazilian legal codes responsible for defining vehicles' categories in an unambiguous manner for many purposes, e.g. tax calculations, and, more importantly, to make e-government systems interoperate while taking laws into account in a Semantic Web scenario. In this work, we define a framework linking the linguistic and conceptual problems to semantic deficiencies and show how these deficiencies were identified during the vehicles' ontology construction.

Keywords: Ontology-based analysis of texts, Semantic deficiencies, Vehicles, E-government, Law, Ontology engineering

1   Introduction

It is quite common, even routine, that people have different accounts about a text they have read or even a dialogue they have had. The ambiguities, inconsistencies and other semantic problems arise in readings or talks due to a number of reasons, which are studied in depth by many different branches of knowledge, ranging from human sciences like Philosophy and Linguistics to exact sciences like Mathematical Logics and Artificial Intelligence [20]. The main reason for these misinterpretations on human communication, however, seems to lie in the very nature of natural languages: they emerged as a result of a long – indeed, never lasting - constructive communication process. The process *per se* is rather a solution than a problem, nonetheless, the different interpretations of terms, phrases, intentions and meanings conveyed in written or spoken communication do cause harmful misunderstandings. Those are due to the fact that natural languages are not the outcome from a formal process. While formal languages are composed of formal syntaxes - that regulate expressions which are valid in the employed language - related to accurate semantics - with which interlocutors and readers can clearly disambiguate among expressions' different interpretations -, in natural languages two interlocutors expect to share a same interpretation about single phrases and entire texts, without relying on any sort of formal semantics, understood by both of them. Their expectations, unfortunately, are not always fulfilled.

In order to avoid the same problem to happen in a higher scale in computer communication, semantic technologies, and particularly the ones related to the Semantic Web and its ontologies [1], have proven useful for many government related applications and prototypes, such as service configuration and automatic service connection among many others. This is possible because the Semantic Web is based on ontologies, which, in practical words, stands for a detailed conceptualization of a domain and its concepts, relations, constraints and axioms. These specifications must be defined in an unambiguous manner using formal logic. On the other hand, official documents, and particularly legal ones like law codes, often contain semantic deficiencies that are not realized by their authors. The most common among them are ambiguities and inconsistencies ought to linguistic problems like polysemy (i.e., one word with multiple senses), as well as underspecifications and inconsistencies. In this work, we sketch a detailed framework of these deficiencies and show how they were identified during ontology construction. These deficiencies are certainly a source of integration problems and confusion during their usage, when the intended meanings can differ depending upon the stakeholder.

During the ontology development of a domain as simple as vehicles, we have witnessed such phenomena in our case study, in the domain of vehicles. The necessity of defining an ontology with the different vehicle types in detail for classification and checking purposes shed light on some of these deficiencies present in two Brazilian legal codes.

This finding can open up new possibilities in the usage of semantic technologies, as guides to check whether official documents are ontologically and logically correct, by not containing ambiguities, underspecifications or inconsistencies.

As a result, this article aims at the following objectives:

- To introduce a framework of possible semantic deficiencies that can be encountered in texts. Such framework would help identify potential misunderstandings in law codes, and lead both to improvements in the law, by making explicit the different interpretations that a law can assume, for instance.

- To present a case study in which the process of ontology construction clearly indicates the existence of some of the semantic deficiencies mentioned in the framework, as well as the adopted solutions to deal with them.

- To propose an architecture of law checkers based on ontologies that could make for the semantic deficiencies found on laws, also discussing the possible functions for such tool.

In order to tackle these objectives, we organized this article as follows: Section 2 describes semantic technologies and the Semantic Web techniques, together with its uses and benefits to electronic government; section 3 discusses which linguistic and conceptual problems lead to semantic deficiencies and why. The semantic deficiencies' framework is conceptualized in this section; section 4 details the domain of vehicles according the two Brazilian official documents (legal codes) used to construct the ontology. It also describes the ontology itself, with its classes, relations and axioms; section 5 illustrates the types of deficiencies with examples found out in these two legal documents throughout the process of ontology construction, with a special focus on a specific semantic deficiency type – inconsistencies -, which could be very common in law codes; section 6 envisions what types of checking could be tested against official documents. A discussion of how these tests could be carried out in a methodical fashion is also conveyed. Additionally, it brings about the issue of the roles ontologies can play in checking laws' correctness; section 7 brings related work on e-government law applications and compares them with the proposed approach, in terms of knowledge representation and other gains. We also comment and compare our resulting ontology against other ontologies available on the Web about the same domain. Finally, section 8 concludes the article and envisions future work.

Towards Checking Laws' Consistency through Ontology Design: The Case of
Brazilian Vehicles' Laws

Fred Freitas
Zacharias Candeias Jr.
Heiner Stuckenschmidt

2   Semantic Web and E-government

The key idea that leveraged the conception of the Semantic Web was the fact the World Wide Web is mainly aimed at human information consumption, rather than at its use by the software [2]. That fact is evidenced by the low performances displayed by search engines (like Google) in finding information in the presence of polysemy (i.e., a word with multiple meanings). For example, in a Google query, one cannot state a query that refers to the word 'Robin' as a surname. Consequently, the keyword limitations of these pages hamper the safe processing of web information by software agents and programs.

According to Tim Bernes-Lee [4], one of the responsible for the establishment of the World Wide Web, ontologies are the central components and motivation of the Semantic Web, a Web in which the software "understands" and processes data from Web pages. Ontologies supply means of defining the context surrounding these pages and the meaning of the jargon used on them, which is defined in a formal way, using a specific form of logic, description logic [1], [3].

How can ontologies provide context for Web pages? In one of the most cited definitions of ontologies, Gruber states that "*an ontology is an explicit specification of a conceptualization*" [1]. Therefore, ontologies comprise a body of formally represented knowledge that can be machine processed, so that two pieces of software can communicate and interoperate, using such definitions as a shared vocabulary, hence enabling direct business-to-business operations [24]. In practical words, ontologies encompass definitions of concepts, properties, relations, constraints, axioms, processes and events about a certain domain or universe of discourse. If a Web page instantiate a body of definitions about the domain that this page refers to, the software or agent which handles the page can make use of a precise, clear, formal semantics that defined *a priori* knowledge related to the pieces of information present in the page.

Due to the lack of ambiguity, for example, semantic technologies can be employed in information integration and system interoperability of e-government services which can be free of misinterpretations, particularly if the description logic OWL (Ontology Web Language) [24], which is recommended by the W3C as one of the standards for the Semantic Web, is used for defining the terminology employed in the specific domain. A practical example should make things clearer and is presented below.

Comte and Léclére [6] report on a semantic reasoning framework for the French social welfare e-government system. The problem for the French was providing a unified vocabulary and its corresponding meanings to which the welfare systems should comply with, as a first step towards shared interoperability among e-government systems. Thus, an ontology of social welfare services for family household service professions was one of the resources defined for this goal. This ontology contains a definition of what is a house hold professional, in the following form:

$$HouseHoldProfessional \equiv Person \sqcap hasValue(lives,?x) \sqcap \exists\, workFor.\,(\,Person \sqcap hasValue(lives,?x)\,) \sqcap$$
$$\forall\, workFor.\,(\,Person \sqcap hasValue(lives,?x)\,) \tag{1}$$

This definition means that such a professional is:

- a person (i.e., can only be filled with instances of the class $Person$), and (meant by the symbol $\sqcap$)

- who lives somewhere, in a certain address $?x$ - this is defined in the predicate $hasValue(lives,?x)$ -, and

- who works for *some* (meant by the symbol $\exists$) person who lives in the same address $?x$, and

- who works *only* (meant by the symbol $\forall$) for persons who live in the same address as she ($?x$).

Note that this definition is unambiguous, as it does not give rise to any false interpretation of household in this domain, and is treated by systems as so, given that they are endowed with a description logic reasoner such as RACER [17]. If an object or database instance fulfills this definition, it is automatically classified by the reasoner as an instance of the class $HouseHoldProfessional$. This will enable indexing electronic resources and representing the content of document-based resources in a semantic annotation base, according to the semantic web jargon. Hence, a query to this knowledge base will allow finding relevant information from different media, and a way to access it.
This trend of semantic information processing, and particularly the use of the Semantic Web as a means of knowledge sharing and system interoperability, is already provoking ample changes in e-government practice over the Internet. The benefits are various.

Since Semantic Web technologies allow for b2b (business-to-business) integration, their counterparts for electronic governments, g2g (government-to-government) and b2g (business-to-government) can analogously be performed by e-government public services in a more powerful, safe and fast way. An example was illustrated during the EU OntoGov project, considered strategic for the EU community. When a citizen changes her address in some

European countries, she needs to perform a sequence of bureaucratic steps: announce the move, register for the new address, unregister for the old one, update many companies (light and water suppliers, for instance), inform the post, and some other more. In the project, Stojanovic et allii [29] presented a solution in which the citizen only needs to communicates once with the government, and all of these steps are accomplished automatically. A first process informs some of the others using automatically generated electronic documents, annotated with semantic contents. Some of the others continue the communication chain until the workflow defined for that operation is finished. Their prototype evidenced that the workflow of some e-government services find good room for improvement using semantic technologies based on ontologies.

Nevertheless, this is not the only facet where ontologies promise to change drastically the current reality of information handling for e-government. Other areas such as knowledge management, tutoring and geographic systems, among many other application areas for governments, are also starting to take advantage of ontologies to improve their systems' abilities.

We may remark, however, that some challenges must still be overcome to the adoption of e-government solutions based on semantic technologies in a larger scale. The availability and construction cost of ontologies and the heterogeneity of ontologies about a same subject [11] are usually enlisted as some of the key barriers inherent to the technology that may hamper a wider use of Semantic Web-based solutions in e-government. Other noteworthy issues are linked to government and e-government practices themselves. For instance, services to be integrated must be available and the integration of them is often problematic, due to the different perspectives, incompliant conceptualizations among other typical problems related to knowledge heterogeneity. The government institutions need also to formalize their workflow and document exchange, before integration takes place. As for the citizen side, the offered services must be easy to find and use; since a government as a rule offers a huge variety of services, good internal search engines, based on natural language processing methods that are able to distinguish among the several meanings of some words in the government context are essential. For instance, the word 'incentive' can refer to financial deductions in the payment of some taxes or the promotion of some public event. Such enhanced search tools must assist citizens in the task of finding the desired services and already in use. The Italian government in the site of the Presidency of the Ministers Council (Site 3) deploys such a tool.

In this article, we aim at stressing a relevant role to be played by ontologies in law information systems: the clarification and disambiguation of semantic deficiencies present on law texts. To this end, in the next section, we enlist and describe types of misunderstandings in texts caused by linguistic or conceptual problems.

3   On Semantic Deficiencies in Textual Definitions

Natural languages give rise to a long list of distinct ambiguity types. A representative set of these ambiguity types, not meant to be exhaustively comprehensive, is listed and briefly explained in the next subsection.

3.1   Linguistic Ambiguities

A first sort of ambiguity relates to word senses, once words can have multiple meanings, e.g., the word "foot" can assume two different interpretations at least; in the expressions "foot of a mountain" and "foot of a man". This linguistic phenomenon known as Polysemy, hampers human communication. A phonetic counterpart of Polysemy is called Homonymy [10]. It happens when two different words spell or sound the same, like the word "read", that means the present or past tense of the verb with the same name, according to the intonation of the speaker. In both cases, Polysemy and Homonymy, the "linguistic unit" linked to the problem is a word.

Other ambiguity types are related to referents in a phrase, e.g. in the phrase "John helped Carl; he was very kind", in which the noun "he" refers to "John" and not to "Carl". This figure of speech, in which an instance of an expression refers to another instance in the same text, has phrases as linguistic units. Ambivalences in phases are classified in Linguistics according two types:

- Anaphoras [9], in which referents come before, as in the latter example, or are supposed to be known. There are two main categories of Anaphoras:

  o The Endophoric Anaphoras [18], in which the referent is within the text being analyzed this referent has been previously identified and mentioned (like in the previous example);

  o The Exophoric Anaphoras [18], in which the referent is outside the text. For instance, the expression "our president" often found in news reports, can only be understood according to the country the report was written at. In this case, contextual knowledge of the interlocutors is needed for understanding the specific meaning the anaphoras refer to.

- Cataphoras [18], in which the referents come after its references, like in the phrase "If you like them, there are cookies in this shelf". Note that Cataphoras can only be Endophoric, as the referents have to come after its mentioning.

In the next subsection, we enlist and describe types of misunderstandings in texts caused by these linguistic problems as well as the conceptual ones.

3.2 Semantic Deficiencies and Their Origins

The main consequence of the three types of (problematic) linguistic constructions and its subtypes enlisted above, on their turn, is generating communication misunderstandings, read or written. According to linguists, there are too many types of possible ambiguities. Poesio et alii [25], e.g., mentions mereological anaphoras, in which the ambiguous referent is a part of another referent; this part is identifiable only after some posterior utterance about certain particular functionalities of it appear. Other problems relate to the conceptualizations that (might) support the ideas and context in a text. Summarizing these problems, we have outlined a framework of semantic deficiencies in texts that is depicted in figure 1. This subsection is devoted to explain and discuss them.
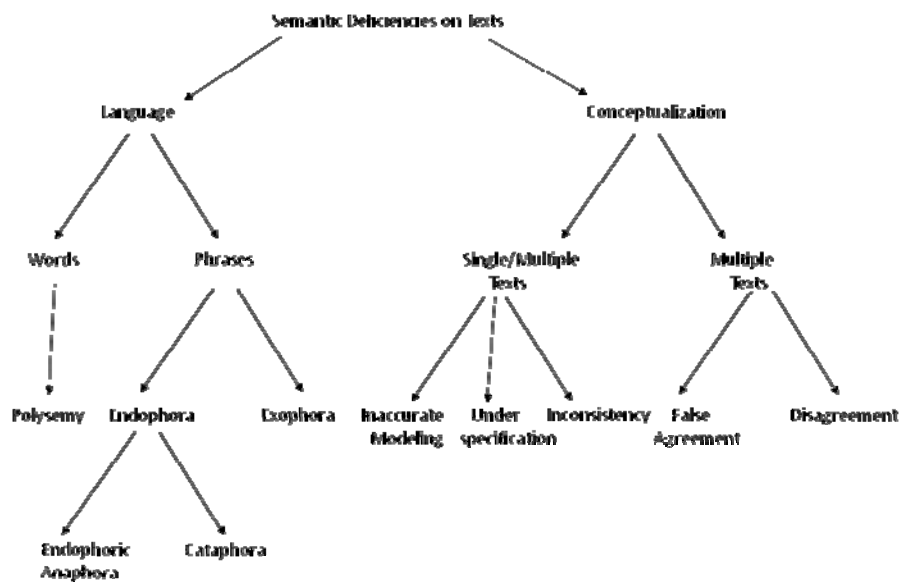

Figure 1: A framework of semantic deficiencies in texts

A quite common type of ambiguity relates to *word or phrase ambivalences*, in which the intended meaning of a word can correspond to more than one possible interpretation. This ambiguity type is usually linked to (and can be caused by) Polysemy, as well as Endophoric and Exophoric references.

Although problematic, linguistic problems like distinct word senses and anaphoras are not hard to be identified; nonetheless, there are many other ways of generating semantic deficiencies, which are not linked to the inaccurate usage of language, but to the *representations* of the domain conveyed in texts, or, in more technical terms, in their *conceptualizations*. In other words, the problem relies not on the words or phrases, but on the poor or different, contradictory conceptualizations related to them. The representations then cannot grasp the actual state of affairs of the objects and their corresponding relationships, abstractions and qualifications in the universe of discourse being represented, what characterizes an *inaccurate modeling*. This kind of conceptualizations are said to lack *ontological engagement.*

A particularly common type of such conceptual discrepancies is *Underspecification*. It stands for a special case of ambiguity in which a description of a certain class of objects in the world is too broad, and, therefore, incomplete. In other words, a description of a prototypical object of the class - which is supposed to work as a "rule of thumb" to check whether a given object should be recognized as an instance of the described class or not - is not specific enough to rule out false members. For instance, stating that a motorcycle is a vehicle with two wheels does not exclude bicycles to be classified as such. Underspecifications thus occur because precise details are lacking, like in the example just presented.

On the one hand, underspecification, at one level, is a necessary and even benign phenomenon in human communication, because phrases can refer only to their focus, thus leaving out many details that are supposed to belong to the shared conceptualizations of two interlocutors (that can be a reader and a writer as well), or *default knowledge*, in the jargon of Knowledge Representation. On the other hand, wrong interpretations can be brought out just because relevant details on qualities, features, premises, roles, and / or many other delicate aspects that bring

116

Fred Freitas
Zacharias Candeias Jr.
Heiner Stuckenschmidt

meaning to the descriptions are not explicit, or are either falsifiable, ambivalent or disagree at some point. This latter is the most emblematic and subtle subtype of underspecification, the *False Agreement* Problem. This term, coined in Informatics by Guarino [15], denotes a situation in which two interlocutors or authors refer the very same object or concept and think they agree on their definition of it, but indeed the two get different accounts of it, i.e., employ slightly distinct conceptualizations of it. For instance, two developers agree that Agent such be a subclass of Person. Nevertheless, one of them thinks that in his system, the concept "Agent" can be a subclass of software too, so in his system not all agents are persons.

The usage of natural language enforces the appearance of such mistakes, since interlocutors and / or authors do not explicit and make agreements about their inner beliefs about each concept used in a communication or document. Therefore, the disagreement is hidden, and none of the interlocutors is aware of it, until the bad consequences appear. In the example mentioned, a good conceptualization would consider that agents should be divided in two types, human agents and software agents, because "Agent" is a role while "Person" is a rigid class that instead conveys the meaning that any of its instances will ever be so while they exist, and roles do not exhibit this feature [17]. Had the two interlocutors made explicit a *disagreement*, i.e. one sticking to the subsumption "Person"–"Agent" and the other not, at least the differences in their conceptualizations would be clear, and integration could be accomplished by mappings, for example, without further problems.

The false agreement problem is deeply rooted in Philosophy. The first mention of it (although not under this name) is due to Aristotle; he realized that sophists' strategy during the debates was just to take advantage of the different accounts of concepts in order to build their falsifiable arguments [2]. Currently, False Agreements threaten design in general in the exact sciences (including the software industry) by leading developers to wrong interpretations of procedures' descriptions and design plans, with potentially disastrous consequences. In database definitions, for example, subtle distinctions in conceptualizations that grounded schemata frequently hinder heterogeneous database integration.

The opposite of underspecification, overspecification, can also lead to another serious conceptual problem, *inconsistencies*. It takes place when descriptions (including the default assumptions) either contain or entail definitions that contradict each other or describe classes to which no element is able to fit. One example can be found in subsection 4.4.

Such problems can be clarified and solved by formalizing the terminology of a domain in terms of an ontology, depending upon the suitability of the knowledge representation formalism employed. Ontologies specified with semantic networks and frames languages, are adequate for disambiguating the language level of this framework (phrases and words). Ontologies written in Description Logic languages, like OWL (Ontology Web Language) [24] or in First-Order Logic are endowed with unambiguous semantics, and thus are able to circumscribe classes precisely as sets of axioms that control not only which objects it can be constituted of, but also the containment relations among the classes. Therefore, they can solve all the semantic deficiencies mentioned above, but cannot clarify and solve all the cases of False Agreements, because, to do so, one needs to rely on top or foundational ontologies, like Sowa's [27] and the Unified Foundational Ontology (UFO) [16], that offer high level predications that can be assigned to characterize the inherent features of the classes. Foundational ontologies constrain the definitions in the new ontology being created so as to prevent any kind of tricky absurd construct to come out from the basic definitions, like the different partonomical relations and their inner features (like transitivity, which occurs in some of these partonomical relations). While foundational ontologies usually need to be expressed in complex, expressive modal or higher order formalisms, for the ontology being created does not always require this high expressiveness. Table 1 presents the adequacy of formalisms against the solvable semantic deficiencies.

In the next section, we delve into an actual case study in the field of laws, the definition of vehicle types in Brazil. As it can be seen, the process of clarifying the terms and their actual meanings evidenced some semantic deficiencies in the laws codes that were hidden behind bad texts and their conceptualizations.

Table 1: Adequacy of formalisms to clarifying / solving semantic deficiencies

Knowledge Representation Formalism	Solvable Semantic Deficiencies
Semantic Networks, Frames	Level of Language
Description Logic, First Order Logic	Levels of Language and Conceptualization, but not all types of False Agreements
Modal, Intensional Modal Logics, Higher-Order Logic	Levels of Language and Conceptualization, including all types of False Agreements

4   The Vehicles' Domain and Resulting Ontology

Though acceptable and, in most of the times, manageable on daily life, small misunderstandings in texts are potentially dangerous not only in the exact sciences, but also plagues the field of Law. It is so commonly accepted in many countries that laws are ambiguous or form an inconsistent set of regulations, in which some laws contradict others, that a deluge of trials depend upon personal law interpretation by judges, attorneys, courts and juries, thus hampering justice to be efficient. Indeed, our claim is that a profitable field of ontology application could be to clarify,

Fred Freitas
Zacharias Candeias Jr.
Heiner Stuckenschmidt

disambiguate and correct semantic deficiencies contained in law cases. We have experienced a case study in that direction that we will describe in the following.

4.1   The Vehicles' Ontology

Although the State Government of Pernambuco is one of the states that offers most advanced e-government online services in Brazil through its portal (Site 2), such as tourist information, crimes' and vehicles' registry, only now Semantic Web technologies are being used, with a twofold purpose: First, to ensure that the interoperability among database schemata and the document exchange between systems will take place without risk of misinterpretations or mistakes. The second reason lies on the reuse possibilities: most laws in Brazil are federal; therefore, the definitions generated in this effort are likely to become standard and can be reused in many other states of the country. It can even be reused in other contexts, such as private organizations that build applications compliant to the current laws.

Having this scenario as a background, we have developed a vehicles' ontology in order to serve many e-government purposes, being document exchange the main one. It was written in OWL, since we will need to classify instances of vehicles for taxes assignment, for example. The construction of this artifact was built in conformance with the following sources:

- Brazilian Traffic (law) Code (CTB – available online at Site 4) that states the Brazilian traffic laws in general terms, so that specific details have to be defined elsewhere;

- Resolutions and deliberations from Contran (Conselho Nacional de Trânsito, , the National Traffic Code) [7], [8] which make for the details missing in CTB;

- Sites of the State and Federal Traffic Departments – in Brazil, Traffic Departments are the public institutions empowered to enforce traffic laws (Sites 1 and 4),

- Many existing database schemas and systems' manuals, from the State Traffic Department of Pernambuco (Site 1),

- Interviews with analysts responsible for the systems and databases.

The Methontology methodology [13] was followed very closely, such that all documents required by it were produced throughout the ontology construction process, including the terms' glossary and the concepts dictionary.

4.2   Definitions

As a guiding principle, we decide to stick to the law categories and their definitions during the vehicles' ontology design. So, according to the laws, vehicles are classified in the dimensions of User category, Traction, Function and Type. Although the classes under these dimensions are mentioned in the codes, most of their definitions are not present, but only some relevant features. Therefore, a hard task was to appropriately portray these classes into axioms in a way that they can correctly infer to which class a given described vehicle belongs.

Figure 2 displays the most relevant relationships between the main class Vehicle and its main features (such as propulsion type, driver position (this information piece is valuable to distinguish between motorcycles, in which drivers go mounted, and motor scooters, in which drivers go seated), driving control device, function, owner type, etc) and components. The specialization of these features and specific constraints on vehicles accurately defines the different vehicle types that constitute the ontology backbone.

Just to have a flavor of how complex these axioms turned out to be, these are the definitions, in Portuguese and English, that corresponds to a Moped (*Ciclomotor*, in Portuguese), according to the Annex I of CTB, which contains "Concepts and Definitions" related to vehicles' categories:

"Ciclomotor – veículo de duas ou três rodas, provido de um motor de combustão interna, cuja cilindrada não exceda cinqüenta centímetros cúbicos [...] e cuja velocidade máxima [...] não exceda a cinqüenta quilômetros por hora."

Fred Freitas
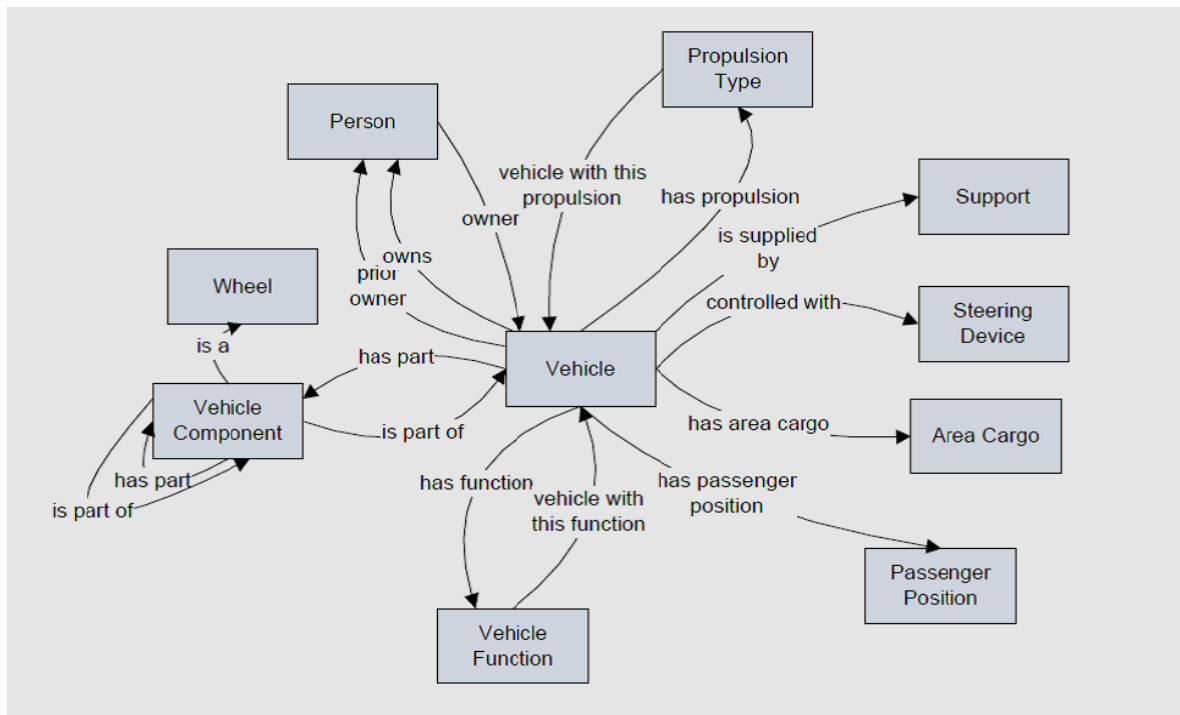Zacharias Candeias Jr.
Heiner Stuckenschmidt

Figure 2: The main classes of the vehicle ontology and their relationships

"Moped – vehicle with two or three wheels, endowed with in an internal combustion engine restricted with a maximum engine displacement of fifty cubic centimeters […] and whose maximum speed does not exceed fifty kilometers per hour."

In Description Logic, it reads as

Moped Vehicle ≡ (= 2 hasPart.Wheel ⊔ = 3 hasPart.Wheel) ⊓ =1 controledWith ⊓ ∀controledWith.Handlebar ⊓

$$\textit{hasValue (cc, UpTo50cc)} \sqcap \textit{hasValue (maxSpeed, UpTo50KmH)} \tag{2}$$

We have the following remarks about the definition above:

- In order to support these axioms, we had also to deploy support definitions in the form of classes like vehicle components, accessories, functions and user types, as well as 27 data and 18 object properties.

- The axiom above also states that Mopeds are vehicles driven by handlebars, although this was not stated in any law code. This is, however, the common understanding of the law by people, lawyers and judges *by default.*

- On the other hand, we dropped the reference about the internal combustion engine as a necessary and sufficient condition for Mopeds, since engines with the property *cc* can only be of internal combustion.

- It was necessary to introduce wheels as individuals, since there can be different types of wheels even in a same vehicle, like tractors,

- and also to define categories of speed and displacements, instead of working with continuous values, once OWL reasoners are not able to compute with numbers. It can just count instances, like the two or three wheels.

A common car had a tricky definition, too:

Automobile ≡ Vehicle ⊓ ∃ hasPropulsion.MotorizedTraction ⊓ >= 3 hasPart.Wheel ⊓ <=6 hasPart.Wheel ⊓

$$\textit{hasValue (passengerPosition, Seated)} \sqcap \textit{hasValue (MaximumNumberOfPassengers, Until9Passengers)} \sqcap$$
$$\neg \, (\exists \, \textit{hasPart.(RearCargoArea} \sqcup \textit{FifthWheel))} \tag{3}$$

All the conditions above stated are meant to differentiate cars from other vehicles: the position of passengers is seated (so tricycles are not cars), it can carry up to 9 passengers (this is in the law code, intended not to confuse with buses or vans), and containing neither rear cargo areas (like pickup trucks) nor fifth wheels (like tractor trucks).

In the next section, we look into the semantic deficiencies that were evidenced throughout ontology design process.

5 Semantic Deficiencies Regarding Vehicles Found in the Law Codes throughout the Process of Ontology Construction

As an introduction to the theme, we discuss next drawbacks of the law writing style. Law texts are written in a particular, convoluted style, so that most citizens face difficulties on understanding them at their complete extent. Therefore, assessing the consequences and actual possibilities of law action is limited by the law writing style.

5.1 Laws' Writing Style

On the one hand, law writing style is characterized by at least not possessing Cataphoras (see subsection 2.1). On the other hand, all the other semantic deficiencies can be found on them, which sometimes make the intended meaning hard to understand. For instance, Contran Resolution 291 [8] states in its Article 1 that:

"[...] Ao requerer a concessão do código específico de marca/modelo/versão e emissão do Certificado de [...] o interessado deve: [...]"

"[…] When requesting the concession of the specific brand/model/version code and the issuing of the […] Certificate, interested ones must: […]"

The term *interested ones* is a complicated anaphorical case: it constitutes an Exophoric Anaphora, as it refers to an external source (the vehicle owner), which is not in the text. To make things worse, the vehicle is only identified by means of a mereological anaphora that mentions the vehicle code without using the word vehicle.

While this type of reference poses difficult problems for automatic text processing, at least, this problem does not prevent people to understand the text meaning, since it is not ambiguous, contradictory or underspecified. This does not apply, unfortunately, to the next semantic deficiencies that we present on the following.

5.2 Ambiguities, False Agreement

For instance, the Portuguese term *Carroçaria* (or *Carroceria*) consists of a Polysemy in Brazil. A first denotation of this word stands for rear area cargo. This sense is employed in the aforementioned Resolution 291 in its Fifth Article:

"Art. 5° Em caso de [...] caminhão, com carroçaria aberta ou fechada, […]"

"Art. 5. In case of […] a [...] truck, with open or closed rear area cargo, […]"

In another word sense, *Carroçaria* stands for the vehicle body. Contran Deliberation number 64 [7], another document which disciplines weights and capacities registers in traction, cargo, collective and transport vehicles, uses *both* senses in its only five pages. In its item 3.1.3, it uses the first sense (area cargo):

"3.1.3 Veículo [...] que recebeu carroçaria ou implemento [...]"

"3.1.3 Vehicle [...] that received a rear area cargo or new implement [...]"

While in item 4.2.3, it refers to the second (vehicle body):

"4.2.3 – [...], a indicação [do peso] deverá ser afixada na parte externa da carroçaria na lateral dianteira."

"4.2.3 – [...], the [weight] indication must be shown on the right side of the vehicle body external part."

As (implicitly) stated in section 2.2, polysemy can lead to False Agreement. The example above can indeed cause a false agreement that could not be solved even relying on good modeling or on the OntoClean methodology [16], which checks if metaproperties of inherited classes (like unity, rigidity, etc) match. Both classes can fit to the given description, since it could be easily (and correctly) interpreted as:

"4.2.3 – [...], the [weight] must be shown on the right side of the rear area cargo external part."

5.3 Underspecification and Bad Modeling

This consists by far the most frequent semantic deficiency. For instance, motorcycles can be classified into two disjoint categories: passengers and cargo, but no further details are conveyed in the law codes. So, how should we make clear distinction between the two?

Description Logics fits quite well to the task of correctly determining the exact perimeter of each class, as well as the inheritance relations among them. One solution could be based on the fact that cargo motorcycles are employed for cargo transport. Defining cargo motorcycles on this way may bear little usefulness, though; see the definition below:

$$CargoMotorcycle \equiv Motorcycle \sqcap \exists hasFunction.CargoTransport \qquad (4)$$

This axiom states that for cargo motorcycles to be recognized as such, in systems that classify vehicles, iff there is an instance c of CargoTransport. a motorcycle instance m and an instance of object property hasFunction (m, c). Nevertheless, this modeling lacks ontological engagement: if any motorcycle is used for light cargo transport (e.g. transporting a single document), it would be classified as a cargo one! The intended meaning of cargo motorcycles for the traffic institutions is based more on vehicle's components as well as with its function (e.g. for taxes assignment). So a definition coherent with this stance is displayed below:

$$CargoMotorcycle \equiv Motorcycle \sqcap \exists hasPart.CargoArea \sqcap \exists hasFunction.CargoTransport \qquad (5)$$

5.4 Inconsistencies

An inconsistency needs at least two axioms. Therefore, they are sometimes hard to locate, since they can hide themselves in distinct law codes. Nonetheless, we have even found inconsistencies in a same official document. For instance, the Law number 9,602 of CTB, brings the following statement:

"Art. 96. Os veículos classificam-se [...] quanto à tração: a) automotor; b) elétrico; c) de propulsão humana; d) de tração animal"

"Art. 96. Vehicles classify according to their traction into [...]: a) automotor; b) electrical; c) human propulsion [like bikes]; d) animal traction"

Although not explicitly stated, the default interpretation of this law considers that these partitions (automotor, etc) disjoint. In order to discuss a specific inconsistency, we are particularly interested in the disjointness of the first pair of classes; for that reason, we state that:

$$ElectricalVehicle \sqsubseteq \neg AutomotorVehicle \qquad (6)$$

Electrical vehicles were not defined in the codes - they are vehicles moved by their electrical engines. Engines that are driven on roads are divided into two distinct classes according to the Brazilian law: electrical and internal combustion engines. The set of definitions below exhibits this state of affairs, assuring via a closure axiom that an electrical vehicle is and can only be moved by an electrical engine (7):

$$ElectricalVehicle \sqsubseteq \exists isMovedBy. ElectricalEngine \sqcap$$
$$\forall isMovedBy. ElectricalEngine \qquad (7)$$
$$Engine \equiv ElectricalEngine \sqcup InternalCombustionEngine \qquad (8)$$

Electrical and internal combustion engines are disjoint too:

$$ElectricalEngine \sqsubseteq \neg InternalCombustionEngine \qquad (9)$$

On the other hand, the Annex I of CTB brings details about automotor vehicles:

"Veículo Automotor - todo veículo a motor de propulsão que circule por seus próprios meios [...]. O termo compreende os veículos conectados a uma linha elétrica e que não circulam sobre trilhos (ônibus elétrico)."

"Automotor Vehicle - every vehicle with a propulsion engine that can move by its own means [...]. The term encompasses the vehicles connected to an electrical wire and which do not move over tracks ([like] electrical buses)."

Note that, on purpose, this definition does not accept electrical trains. However, this statement entails *lots* of inconsistencies. Vehicles that move by their own means have to possess an engine:

Fred Freitas
Zacharias Candeias Jr.
Heiner Stuckenschmidt

AutomotorVehicle ≡ Vehicle ⊓ ∃ isMovedBy. Engine   *(10)*

Unfortunately, this definition clashes with the one of electrical vehicle, in any interpretation. The inconsistency was caused by a conflict between the axioms that declares automotor and electrical vehicles disjoint as well as their engines (axioms 6 and 9 respectively) and the actual concept of automotor (axiom 10). They cannot simultaneously hold. The most reasonable interpretation, which we included in our ontology, is to drop the disjointness axiom 6. After this change, the system automatically subsumes electrical vehicles as a subclass of automotor.

The same problem applies to all the subclasses of electrical vehicles, like electrical bus and electrical trains. They were as a consequence all inconsistent. Moreover, another flaw in the law was explicitly portrayed in Article 96, which considers electrical trains as not being automotor vehicles, despite of moving due to electrical engines.

6   Envisaged Application

If in a domain as simple as vehicles such semantic flaws arise, they certainly occur even more often in more complex law branches, like commercial disputes, tax regulations, and family law, to cite but a few. In these law domains, opposite law interpretations are acceptable in many cases, since contradictions among different law codes abound. Portraying a correct modeling of these branches in thorough details is certainly to be conducted with all the care and requires a clear identification of the correct or intended law interpretations in a way that prevent contradictions among the codes that shall appear.

Our claim with the current work is that ontologies can be used in the field of law to improve the quality of law creation. Support tools that benefit from such ontologies could be used in parliaments, courts and attorneys' offices, provoking a deep behavior change in the way of handling, changing and working with laws. Another good consequence and usage of law ontologies defined in this way is bridging the gap between law and systems, particularly for e-government. By using a common representation, such systems have less risk of not following closely laws and regulations. Indeed, even public services can be available via Semantic Web Services to take on the job. Therefore, typical tasks accomplished by law ontology-based tools could comprise:

- Checking whether new laws conflict with older ones;

- Identifying law possible interpretations (as logic models);

- Providing support for choosing the correct interpretations to be included in a particular law article;

- Clearly showing the chain of laws to be changed in case a change is required;

- Simulating situations in order to check defense or accusation possibilities;

and many other tasks of the same flavor.

We devise a concrete application of the present work in the development of a semantic law checker prototype with such features. Such semantic law checker should make for more defying modeling cases, like stating the laws of traffic ontologically. In these scenarios, false agreements as pointed by Guarino [15] and ill-defined modeling - see many examples at -, that are not being dealt here, will probably take place often, as well as laws from the same or from different codes that interweave and conflict. Therefore, a reputed semantic law checker needs to employ top ontologies (like DOLCE [12], for instance) to assign relevant features of the ontologies' classes, and meta-property checking, using the OntoClean methodology [16].

Figure 3 displays the architecture of a semantic law checker, including the two types of ontologies necessary (law and top ontologies) and the meta-property checking componenet. As the user (attorneys, judges, law authorities) describes laws and juridical situations to be analyzed the system builds the models and warn in case of inconsistencies or alternative models. Naturally, the system should offer a suitable interface to users that converts the information they provide to the system. If the law ontologies are already built, it can also be used to check whether a legal action (process or law suit, for instance) is applicable. Such type of systems could clearly show the different law interpretations (or, in more formal words, logical words, models) and point out when they conflict or when a valid interpretation cannot come out from a set of laws.
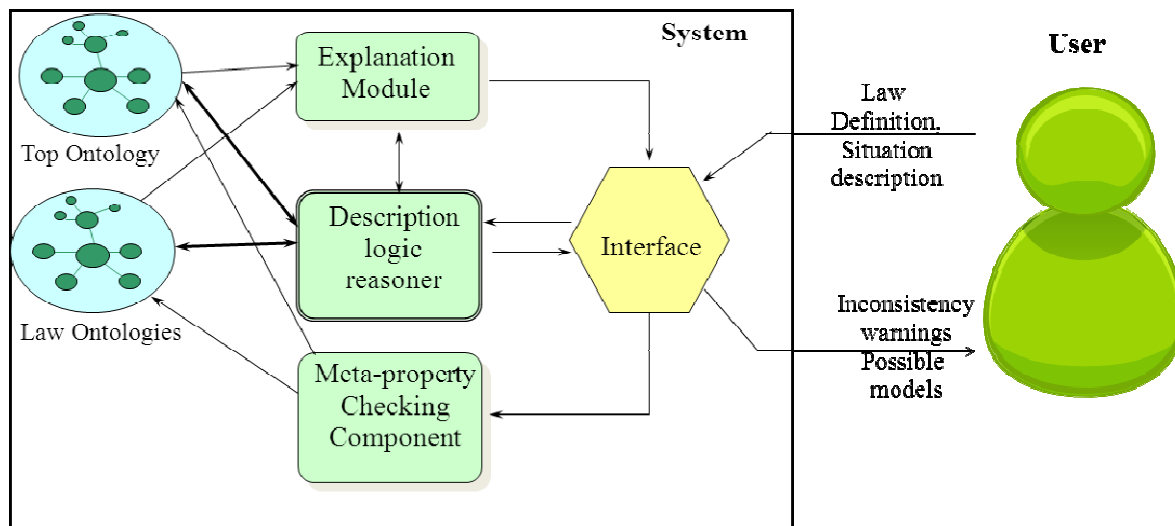
Fred Freitas
Zacharias Candeias Jr.
Heiner Stuckenschmidt

Figure 3: Architecture of a semantic law checker

7   Related Work

Law has attracted strong interest from ontology researchers since mid-90s, when ontologies started occupying its relevant place in the Artificial Intelligence mainstream. One of the first works to address the complexity of law modeling, which requires deep meta-knowledge, was the Core Ontology of Law by Valente and Breuker [30]. They grounded their work under the premise that any branch of law must make use of a common set of concepts like responsibility, normative knowledge, agents (in the law sense), regulations, etc. Breuker [5] has also shown the role that top ontologies can play as a basis for defining knowledge for a specific law subfield (like family law, for instance).

Many applications and prototypes take advantage of legal ontologies nowadays, e.g., systems that facilitate citizens to check whether disputes are feasible, by checking the law possibilities [28]. We will introduce in the next sub-section a brief review of already existent law checkers together with a comparison of such systems with our proposal in terms of knowledge representation and reuse. Next, we discuss ontologies of vehicles which were available in the web and a comparison with one ontology which has the same focus of ours, but with a different perspective.

7.1   Law Checkers

As for the usage of ontologies to validate laws, Van Engers et alii [32] have produced a system that employs legal ontologies to provide flexibility of maintenance to systems, including checking for semantic anomalies among law articles and regulations. It relies on a framework similar to ours. Serving mainly to their system, their framework of semantic errors is however not as general as ours, since it does not consider the conceptualization deficiencies (such as bad modeling and false agreements). Their system validates laws using UML/OCL specifications. The risk of this practice lies on the fact that UML bears no formal semantics. Good features of their tool also include the determination of possible decision paths for juridical decisions and the easy maintainability of the systems against law changes. Since these changes are rather frequent, the systems should change accordingly, and representing knowledge declaratively is the best solution to cope with the task.

Although a tool to test whether our laws are correct is future work for us yet (while they have implemented an efficient tool), we are using OWL, which is based on the well-defined Description Logic formalism, while Enger's system relies on rules. We identify at least two representational problems with rules. First, as any non-monotonic reasoning mechanism, it cannot be possible to take into account logical models, what would hamper checking classes' consistency and the checking of the meta-property compatibility of inheritance relations among classes in the flavor of OntoClean [16], unless they are explicitly stated as rules. Full-fledged reuse of law ontologies written in OWL and its respective reasoning is then limited and impaired by the impossibility of infer class subsumption, i.e., the automatic checking that a class is a subclass of another class.

Another good example of an ontology-based law system was produced by Hoekstra et al [19]. This system employs legal (mostly lightweight) ontologies that are used to search for contextual legal information in texts. Information pieces collected from these tests are used to select case frames. If a case frame is selected, then the system considers that legal action is applicable. The system seems to be very flexible mainly because it is not formal, once it is based on case frames and retrieved texts. However, the knowledge used by it can hardly be reused by other systems, a common drawback of non-formal representations.

Towards Checking Laws' Consistency through Ontology Design: The Case of Brazilian Vehicles' Laws

Fred Freitas
Zacharias Candeias Jr.
Heiner Stuckenschmidt

The formal system HARNESS [31] makes good usage of core law and law ontologies (e.g. library laws) stated in OWL, where these latter's classes contain many subclasses of the former's classes. The system is capable of recognizing when a law is not being followed in a deep and principle way, once, by doing that practice (inheriting and qualifying law types), laws have elegant hierarchies.

7.2 Comparison with other Vehicles' Ontologies

As for ontologies on vehicles, we could find many of them in the Web, most of them with the distinct focus and perspectives from ours, although this theme is present in many description logic examples and exercises. For instance, Lukibanov [22] shows ontologies that support design activities at DaimlerChrysler, whose focus are on car parts, connections, compatibility among the parts, etc. Another ontology [26] focus on the activities of controlling vehicles and war equipment automatically via a multi-agent system, containing classes such as obstacles, missions, tasks, etc.

Nevertheless, we have found out only one ontology focused on our subject, vehicles' categories, although in a different context, the domain of car sales (Site 5). Thus, in order to assess the (quantitative) quality of the resulting ontology, we have carried out a comparison of our Vehicle's ontology with that ontology. Numbers are shown in Table 2.

Table 2: Comparison between the ontology developed here and the car ontology (Site 5). The latter, by not containing axioms, resembles more a database, with its many datatype properties

	Vehicles' Ontology	Car Ontology
Named Classes		
Total	102	90
Primitive	39	89
Defined	63	1
Anonymous Classes' Restrictions		
Total	184	0
Existential	119	0
Universal	20	0
Cardinality	7	0
Value assignment	21	0
Properties		
Object	18	13
Datatype	27	105
With specified domain	45	110
With specified range	18	13

These numbers tell us that the Car ontology resembles a database more than an ontology, once it encompasses a high number of datatype properties while just few object properties. The datatype properties are intended to characterize the different types of vehicles under negotiation. Beyond this deficiency, the cars' ontology does not contain a single axiom, giving the impression that its classes are artificially organized in a hierarchy. This organization is naturally not logically supported by the classes' definitions as we did in our ontology, therefore the organization of such hierarchy cannot be logically justified. Moreover, for not presenting axioms, a reasoning classifier cannot take advantage of it, and infer which classes are subclasses of other departing from their definitions in the description logic terminology.

8 Future Work and Conclusions

With this work, we tried to make the case for applying Semantic Web technologies for checking the consistency of official documents, particularly law codes. To serve this purpose, we have outlined a framework of semantic deficiencies that can occur in texts. These deficiencies are divided into two main classes, linguistic and conceptual flaws. We conceived it to be as general as possible; although we restrict our discussion to law and official legal documents, the framework can be useful for any field, like medicine and informatics.

Governments should pay more attention to the neglected issue of inconsistent laws, a problem that could be minored by formalizing them into ontologies. The consistency checking of laws is a key process that can clarify some introduced vaguenesses and contradictions in the laws, and therefore improve the work of institutions, with its implications to their e-government solutions. In our case study, the Brazilian law codes are responsible for defining vehicles' categories in a presumed unambiguous manner, to be used for many purposes, e.g. tax calculations. In spite of that, we found inconsistent and misleading law definitions that allow for different interpretations.

Van Engers et allii [33] argue that '*the legal context in which government administrative processes take place ask for a specific information architecture and knowledge infrastructure'*. According to these authors, e-government workflow solutions must include knowledge descriptions, and particularly law knowledge, because it permeates most government procedures. For that reason, the practice of checking law consistency can be viewed as a key step for e-government, once governments' good functioning depends upon correct laws. Moreover, such checking could assure that the services' functioning may not conflict in behaviour.

This work motivates us to go further in that research track. Our hypothesis for future work include a larger and more complex and challenging case study with more convoluted law codes, in which plenty of ontological mistakes and inaccuracies are supposed to arise. After carrying this step out, the design and development of a prototype law checker system for traffic issues could be produced. Such a system can serve to lawyers, judgers and law makers in order to analyze the consistency of law codes.

Acknowledgments

Websites List

Site 1: Governo de Pernambuco. DETRAN-PE. Departamento Estadual de Trânsito de Pernambuco
http://www.detran.pe.gov.br

Site 2: Governo de Pernambuco. Portal Pernambuco
http://www2.pe.gov.br/web

Site 3: Governo Italiano. Presidenza del Consiglio di Ministri
http://www.governo.it

Site 4: Ministério das Cidades, Código de Trânsito Brasileiro e Legislação Complementar
http://www.senado.gov.br/web/codigos/transito/httoc.htm

Site 5: Umberto Straccia site
http://gaia.isti.cnr.it/~straccia

References

[1]   G. Antoniou and F. van Harmelen, A Semantic Web Primer, 2nd Edition. The MIT Press, 2008.
[2]   Aristotle, Aristotle: On Sophistical Refutations On Coming-to-be and Passing away, Harvard University Press, 1965.
[3]   F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, The Description Logic Handbook: Theory, Implementation, Applications. Cambridge University Press, Cambridge, UK, 2003.
[4]   T. Berners-Lee, J. Hendler, and O. Lassila, The Semantic Web, Scientific American Magazine, vol. 284, no. 5, pp. 34-43, 2001.
[5]   J. Breuker, Constructing a Legal Ontology: LRI-Core, in Proceedings of WONTO-2004, Workshop on Ontologies and Their Applications. LivroRápido, Recife, Brazil, 2004.
[6]   F. Comte and M. Leclére, A Semantical Reasoning Framework for eGovernment of the French Social Welfare System, The Semantic Web meets eGovernment, 2006 AAAI Spring Symposium Serie, Stanford University, California, USA, March 27-29, 2006.
[7]   Conselho Nacional de Trânsito - CONTRAN, Deliberação Nº 64, de 30.05. 2008. [Online]. Available: http://www.denatran.gov.br/download/Deliberacoes/DELIBERACAO_CONTRAN_64_08.pdf.
[8]   Conselho Nacional de Trânsito - CONTRAN, Resolução Nº 291, de 30.05. 2008 [Online]. Available: http://www.denatran.gov.br/download/Resolucoes/RESOLUCAO_CONTRAN_291.pdf.
[9]   David Crystal, A first dictionary of linguistics and phonetics. Boulder, CO: Westview Press, 1980, pp 25-26.
[10] J. Frazier and K. Rayner, Taking on semantic commitments: Processing multiple meanings vs. multiple senses, Journal of Memory and Language, vol. 29, no. 2, pp. 181-200, 1990.
[11] F. Freitas, H. Stuckenschmidt, and N. Noy, Ontology Issues and Applications: Guest Editorial, Journal of the Brazilian Computer Society, vol. 11, no. 2, pp. 5-16, 2005.
[12] A. Gangemi, N., Guarino, C. Masolo, and A. Oltramari, Sweetening WORDNET with DOLCE, AI Magazine, vol. 24, no. 3, pp. 13-24, 2003.
[13] A. Goméz-Pérez, M. Fernández-López, and O. Corcho, Ontological Engineering: Principles, Methods, Tools and Languages, in Ontologies for Software Engineering and Software Technology. Berlin: Springer-Verlag, 2006, pp. 1-48.
[14] N. Guarino and C. Welty, Evaluating ontological decisions with OntoClean, Commucations of the ACM, vol. 45, no. 2, 2002, pp. 61-65.

[15] N. Guarino, Formal Ontology in Information Systems, in Proceedings of the 1st International Conference, Trento, Italy, June 6-8, 1998, pp. 3-15.
[16] G. Guizzardi, Ontological Foundations for Structural Conceptual Models. PhD thesis, University of Twente, Netherlands, 2005.
[17] V. Haarslev and R. Möller, RACER system description, in Proceedings of the International Joint Conference on Automated Reasoning (IJCAR), June 18-23, 2001, pp. 701-705.
[18] R.Hartmann and F.C. Stork, Dictionary of language and linguistics. London: Applied Science, 1972, p. 193.
[19] R. Hoekstra, BestPortal: Lessons learned in lightweight semantic access to court proceedings, in Proceedings of the 22nd International Conference on Legal Knowledge and Information Systems. Rotterdam, The Netherlands. Jurix, 2009, pp. 69-78.
[20] D. Jurafsky and J. Martin, Speech and Language Processing. Prentice-Hall, 2000.
[21] LinguaLinks Library, Version 5.0, CD-ROM, SIL International, 2003.
[22] O. Lukibanov, Use of Ontologies to Support Design Activities at DaimlerChrysler, presented at the 8th International Protégé Conference, Madrid, Spain - July 18-21, 2005.
[23] Ministério das Cidades, Conselho Nacional de Trânsito, Departamento Nacional de Trânsito. 2009. [Online]. Available: http://www.denatran.gov.br/contran.htm.
[24] P. Patel-Schneider, P. Hayes, and I. Horrocks. (2004, February) OWL - Web Ontology Language Semantics and Abstract Syntax W3C Recommendation. [Online]. Available: http://www.w3.org/tr/2004/rec-owl-semantics-20040210.
[25] M. Poesio, P. Sturt, R. Artstein, and R. Filik, Underspecification and anaphora: Theoretical issues and preliminary evidence. Discourse processes, vol. 42, no. 2, pp. 157-175, 2006.
[26] C. Schlenoff, R. Washington, and T. Barbera, An Intelligent Ground Vehicle Ontology for Multi-Agent System Integration, in Proceedings of the International Conference on Integration of Knowledge Intensive Multi-Agent Systems, Waltham, USA, 2005, pp. 169-174.
[27] J. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations. USA: Brooks Cole, 2000.
[28] W. Steenbergen, Rationalizing Dispute Resolution, From best alternative to the most likely one, in Proceedings of the 3rd ODR Workshop, Brussels, December 8-10, 2005.
[29] N. Stojanovic and D. Apostolou, Change management in e-government: OntoGov case study, Electronic Government, vol. 3, no. 1, pp. 74-92. 2006.
[30] A. Valente and J. Breuker, Towards Principled Core Ontologies, in Proceedings of the KAW-96, Banff, Canada, 1996.
[31] S. Van de Ven, R. Hoekstra, J. Breuker, L. Wortel, and A. El-Ali, Judging Amy: Automated legal assessment using OWL 2, in Proceedings of OWL: Experiences and Directions, Karlsruhe, Germany, October 2008, pp. 1-12.
[32] T. Van Engers and E. Glassée, Facilitating the Legislation Process Using a Shared Conceptual Model, IEEE Intelligent Systems, vol. 16, no. 1, pp. 50-58, 2001.
[33] T. Van Engers, R. Winkels, and P. Kordelaar, The know and the flow, in Proceedings of the 5th International Symposium on the Management of Industrial and Corporate Knowledge, Colombus, Ohio, United States, 2008.