Compatibility Formalization Between PR-OWL and OWL

Rommel Novaes Carvalho, Kathryn Laskey, and Paulo Costa

Department of Systems Engineering & Operations Research
School of Information Technology and Engineering
George Mason University, Fairfax, VA 22030, USA
rommel.carvalho@gmail.com
{klaskey,pcosta}@gmu.edu
http://seor.gmu.edu

Abstract. As stated in [5], a major design goal for PR-OWL was to attain compatibility with OWL. However, this goal has been only partially achieved as yet, primarily due to several key issues not fully addressed in the original work. This paper describes several important issues of compatibility between PR-OWL and OWL, and suggests approaches to deal with them. To illustrate the issues and how they can be addressed, we use procurement fraud as an example application domain [2]. First, we describe the lack of mapping between PR-OWL random variables (RVs) and the concepts defined in OWL, and then show how this mapping can be done. Second, we describe PR-OWL's lack of compatibility with existing types already present in OWL, and then show how every type defined in PR-OWL can be directly mapped to concepts already present in OWL.

Key words: OWL, PR-OWL, MEBN, probabilistic ontology, semantic web, compatibility

1 Introduction

The Semantic Web (SW) is predicated upon radical notions of information sharing, which include [1]: (i) the Anyone can say Anything about Any topic (AAA) requirement; (ii) the open world assumption, i.e. there may exist more information of which we are not aware, and (iii) nonunique naming, meaning that different people can assign different names to the same concept. The Semantic Web (SW) differs from the document web in that it is intended to provide not only information sharing, but also knowledge synergy. We call such an environment characterized a Radical Information Sharing (RIS) environment. While the SW promises great power and flexibility, RIS environments present fundamental challenges, and can lead to chaos, disagreement and conflict.

The challenge facing SW architects is therefore to avoid the natural chaos to which RIS environments are prone, and move to a state characterized by information sharing, cooperation and collaboration. According to [1], one solution to this challenge lies in modeling. Modeling is the process of organizing information for community use. Modeling supports information sharing in four ways: (1) It provides a framework for human communication; (2) it provides a means for explaining conclusions; (3) it provides a basis for formalization and automation of reasoning; and (4) it provides a structure for managing varying viewpoints.

There is an immense variety of modeling approaches. Different approaches and processes are supported by different modeling languages. One of special interest to this research is the Web Ontology Language (OWL) [18, 9]. OWL was developed with the aim of enabling achievement of the full SW potential. According to [18] OWL is intended for use when the information contained in documents needs to be processed by applications, as opposed to situations in which the content need only be presented to humans. OWL can be used to explicitly and formally represent the meaning of terms in vocabularies and the relationships between those terms. This representation of terms and their interrelationships is called an ontology.

One of the first definitions of ontology in the context of the Semantic Web was given by Thomas Gruber [10].

> An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For Artificial Intelligence (AI) systems, what "exists" is that which can be represented. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.

In the past few years, as the Semantic Web community has developed standards and more complex use cases, the need for principled approaches for representing and reasoning under uncertainty has received increasing appreciation. As a consequence, the World Wide Web Consortium (W3C) created the Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG) in 2007 to identify requirements for reasoning with and representing uncertain information in the World Wide Web. The work of the URW3-XG provided an important beginning for characterizing the range of uncertainty that affects reasoning on the scale of the World Wide Web, and the issues to be considered in designing a standard representation of that uncertainty. However, the work to date likely falls short of what would be needed to charter an effort to develop that representation. A candidate representation for uncertainty reasoning in the semantic web is Probabilistic OWL (PR-OWL) [5], an OWL upper ontology for representing probabilistic ontologies based on Multi-Entity Bayesian Networks (MEBN) [15].

As stated in [5], a major design goal for PR-OWL was to attain compatibility with OWL. However, this goal has been only partially achieved as yet, due to several key issues not fully addressed in the original work. First, there is no mapping in PR-OWL to properties of OWL. Second, although PR-OWL has the concept of meta-entities, which allows the definition of complex types, it lacks compatibility with existing types already present in OWL.

These problems have been noted in the literature [20]:

> PR-OWL does not provide a proper integration of the formalism of MEBN and the logical basis of OWL on the meta level. More specifically, as the connection between a statement in PR-OWL and a statement in OWL is not formalized, it is unclear how to perform the integration of ontologies that contain statements of both formalisms.

This paper is structured as follows. Section 2 briefly describes PR-OWL and its underlying logic, MEBN. Section 3 presents PR-OWL's lack of mapping to OWL and our suggested solution to the problem. Section 4 presents the lack of compatibility between types in OWL and PR-OWL and describes how they could be integrated. Finally, Section 5 presents the conclusion of the proposed extension to PR-OWL language.

2 PR-OWL and MEBN Logic

Ontologies are becoming increasingly popular as a means to ensure formal semantic support for knowledge sharing [3, 4, 7, 8, 13, 22]. Representing and reasoning with uncertainty is becoming recognized as an essential capability in many domains. The naïve approach of simply annotating ontologies with numerical probabilities is inadequate, because it cannot capture complex relational probabilistic dependencies. More expressive representation formalisms are needed [16].

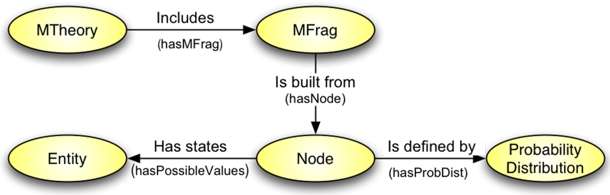

Fig. 1. PR-OWL main concepts.

Probabilistic Ontologies [5, 6] have been proposed as a more expressive formalism for representing knowledge in domains characterized by uncertainty. The PR-OWL probabilistic ontology language [5, 6] has its logical basis in Multi-Entity Bayesian Networks (MEBN), an extension of Bayesian networks (BNs) to achieve first-order expressive power [14, 15]. MEBN represents knowledge as a collection of MEBN Fragments (MFrags), which are organized into MEBN Theories (MTheories). Figure 1 presents the main concepts needed to define an MTheory in PR-OWL. In the diagram, the ellipses represent the general classes, while the arcs represent the main relationships among the classes.

An MFrag contains random variables (RVs) and a fragment graph representing dependencies among these RVs. An MFrag is a template for a fragment of a

Bayesian network. It is instantiated by binding its arguments to domain entity identifiers to create instances of its RVs. There are three kinds of RV: context, resident and input. Context RVs represent conditions that must be satisfied for the distributions represented in the MFrag to apply. Input nodes represent RVs that may influence the distributions defined in the MFrag, but whose distributions are defined in other MFrags. Distributions for resident RV instances are defined in the MFrag. Distributions for resident RVs are defined by specifying local distributions conditioned on the values of the instances of their parents in the fragment graph.

A set of MFrags represents a joint distribution over instances of its random variables. MEBN provides a compact way to represent repeated structure in a BN. An advantage of MEBN is that there is no fixed limit on the number of RV instances, and the random variable instances can be dynamically instantiated as needed.

An MTheory is a set of MFrags that satisfies conditions of consistency ensuring the existence of a unique joint probability distribution over its random variable instances.

To apply an MTheory to reason about particular scenarios, one needs to provide the system with specific information about the individual entity instances involved in the scenario. Throughout the remainder of the paper, we use procurement fraud as an example application domain [2]. On receipt of information about a particular procurement scenario, Bayesian inference can be used both to answer specific questions of interest (e.g., how likely is it that a particular procurement is being directed to a specific enterprise?) and to refine the MTheory (e.g., each new investigation provides additional statistical data about relevant indicators for a given category of fraud). Bayesian inference is used to perform both problem specific inference and learning in a sound, logically coherent manner (for more details see [15, 17]).

3 Mapping PR-OWL Random Variables to OWL Concepts

Suppose we have an OWL ontology for the public procurement domain. The ontology defines concepts such as procurement, winner of a procurement, members of a committee responsible for a procurement, etc. Figure 2 shows a light-weight ontology for this domain represented in Unified Modeling Language (UML) [21].

Now, imagine we want to define some uncertain relations about this domain, e.g. it is common to identify a front for an enterprise by looking at his/her income and the value of a procurement the enterprise he/she represents won, meaning, if the enterprise won a procurement of millions of dollars, but the responsible person for this enterprise makes less than 10 thousand dollars a year, it is likely that this person is a front. Figure 3 shows this probabilistic relation defined using PR-OWL in an open-source tool for probabilistic reasoning, UnBBayes.

As expected, we would need to ensure some conditions were met in order to make assertions about this probabilistic relationship. One of these conditions

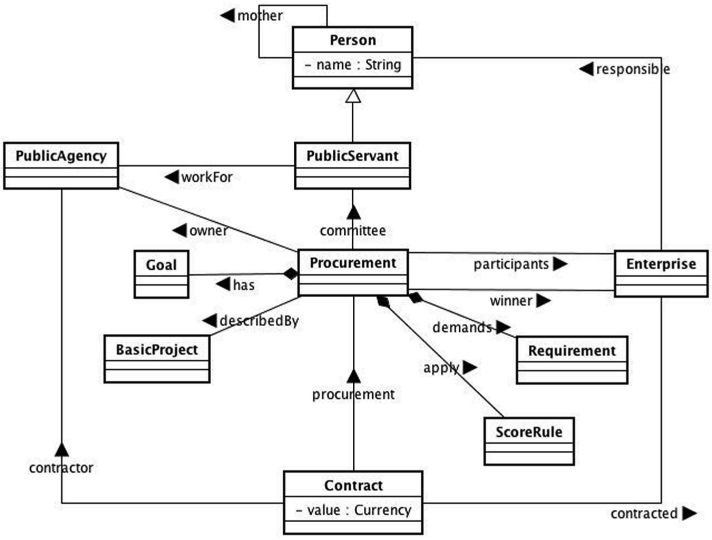

Fig. 2. A class diagram for the procurement domain.

is that the person we are trying to determine as a possible front has to be responsible for the enterprise we are analyzing.

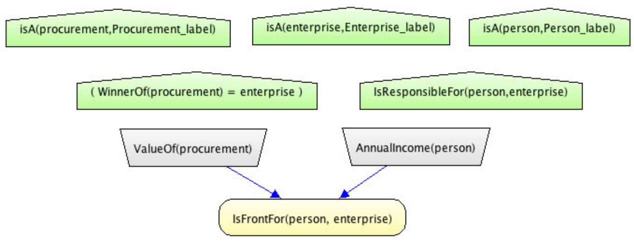

Fig. 3. Front of an Enterprise MFrag.

It is natural to think that the data we have about this domain would be associated with the ontological markups defined in OWL. In other words, our database would have instances of persons and enterprises, and these instances would be linked to their semantic meaning defined in the OWL ontology.

Accessing this information should be trivial once the definitions in the ontology were made available and permission was granted to retrieve data from the database. However, this can only be achieved by developing a link between PR-OWL random variables (RVs) and the concepts defined in OWL. In its current state, though, the relations defined in PR-OWL are not formally linked to the relevant concepts in the OWL procurement ontology. That is, the relation

IsResponsibleFor should be linked to the OWL concepts representing persons and enterprises.

From this simple example, it is clear that every probabilistic definition involving a concept must keep a reference to its semantic definition. In other words, full compatibility with OWL requires modifications to PR-OWL that guarantee the preservation of OWL's semantics.

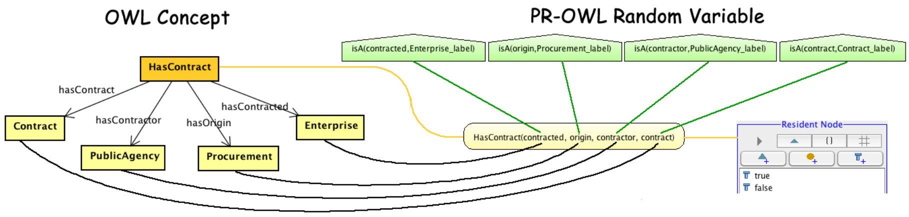

Fig. 4. Ternary relation mapping between OWL and PR-OWL.

Figure 4 shows a suggested approach to map concepts in OWL to random variables in PR-OWL. In this case, the relation is represented by a class (see [11] for details on how to define n−ary relations in OWL) named *HasContract* which represents a 4-ary relation that relates a contract that has a public agency as a contractor, has its origin in a procurement, and has an enterprise contracted. This relation is mapped as a predicate because in this example it is possible to have more than one contracted enterprise for the same contract.

The idea is to keep a reference to the main relation (OWL concept) when creating its probabilistic definition (PR-OWL random variable). In this case, the random variable *HasContract* is a function that defines the probabilistic characteristics of the concept *HasContract*, in this case a class that has a role of a relation as explained above. As it can be seen on Figure 4 the range of the random variable is a boolean.

However, it is not enough to map a PR-OWL random variable to an OWL concept whose probabilistic characteristics are being defined. It is also necessary to map the arguments of the random variable to their respective classes or data types in OWL. In this example we have that: the argument *contract* is mapped to the class *Contract*, which is the range of the property *hasContract*; the argument *contracted* is mapped to the class *Enterprise*, which is the range of the property *hasContracted*; the argument *origin* is mapped to the class *Procurement*, which is the range of the property *hasOrigin*; and the the argument *contractor* is mapped to the class *PublicAgency*, which is the range of the property *hasContractor*.

Finally, in First Order Logic (FOL) the range of any n-ary predicate is a boolean. However, due to the lack of n-ary relations in OWL, this predicate was modeled in PR-OWL as a class. Therefore, it has no defined range. In fact, the only possible value of a class is an instantiation. So there is one last mapping to be done.

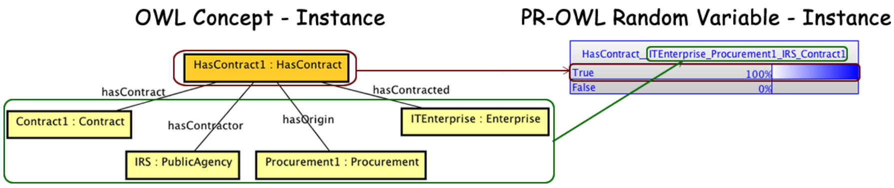

Fig. 5. Ternary relation instance mapping between OWL and PR-OWL.

We need to map the existence of an instance to a random variable with value true. This is a one−to−one mapping. I.e., there is only one RV that describes the uncertainty of an OWL instance and there is only one OWL instance that describes the semantics of a RV. If there is no such instance, then the value of the RV is either false (if we are assuming a closed world) or unknown (if we are assuming an open world). The fact is that we need to have an extra parameter in our RV to state to which instance of the class *HasContract* it is related to.

So, once the random variable is instantiated as a node in a Bayesian network for a specific situation, it is necessary to maintain the mapping we had in our PR-OWL random variable. Figure 5 shows a suggestion for how to perform this mapping. In this example, as we have an OWL assertion that *HasContract1* is a predicate that states that the *Contract1* contract had origin in *Procurement1*, has *IRS* as its contractor, and has contracted *ITEnterprise*, we have the PR-OWL counterpart stating that the node *HasContract__ITEnterprise_-Procurement1_IRS_Contract1* has the state *True* with probability 100%. The actual mapping between the OWL instance and the node is kept because the node is in fact an instance of the random variable defined in PR-OWL, which in turn is an OWL class. As the PR-OWL random variable has the mapping, so does its instance.

The mapping described in this section provides the basis for a formal definition of consistency between a PR-OWL probabilistic ontology and an OWL ontology, in which rules in the OWL ontology correspond to probability one assertions in the PR-OWL ontology. A formal notion of consistency can lead to development of consistency checking algorithms.

4 Extending PR-OWL to Use OWL's Types

One of the main concerns when developing OWL [12] was to keep the same semantics of its predecessors, RDF and XML, which meant reusing all the concepts already defined in those languages, including primitive types, such as string, boolean, decimal, etc. On the other hand, PR-OWL does not make use of the primitive types used in OWL. For instance, PR-OWL defines *Boolean* as an individual of the class *MetaEntity*, as shown in Figure 6, but does not keep any relation to the boolean type used in OWL.

If we wanted to define a continuous random variable for the annual income of a person in PR-OWL, we would need to define the real numbers, even though

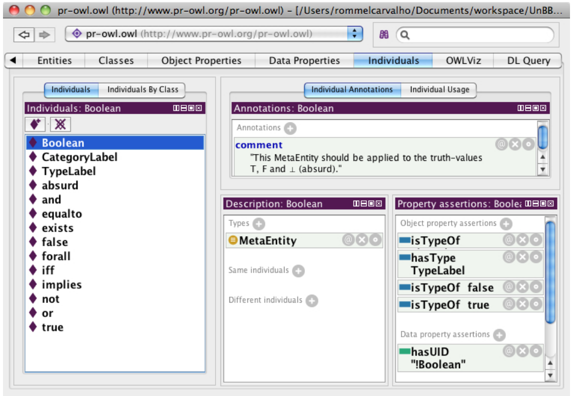

Fig. 6. Boolean individual defined in PR-OWL.

they are already defined in OWL. Moreover, concepts that use this primitive type in OWL would not be understood in PR-OWL, in other words, they lack compatibility as far as primitive types are concerned.

Figure 7 shows the different types of entities defined in PR-OWL. A possible approach to keep OWL's semantics is to avoid defining new types of entities and use what is already available in OWL. For instance, the class *ObjectEntity* can be substituted by the OWL class *Thing*, after all, according to [5] *ObjectEntity* aggregates the MEBN entities that are real world concepts of interest in a domain. They are akin to objects in Object-Oriented (OO) models and to frames in frame-based knowledge systems. In other words, they are nothing more than OWL classes.

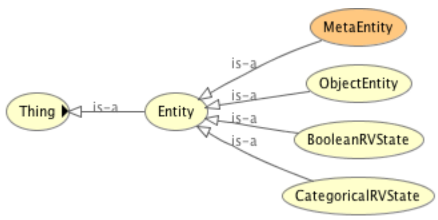

Fig. 7. The different types of entities defined in PR-OWL.

According to [5] *CategoricalRVState* is used to represent a list of mutually exclusive, collectively exhaustive states, which in turn are possible states of random variables, represented by nodes in PR-OWL. Therefore, it can be replaced by *DataOneOf* if it needs to enumerate data types or *ObjectOneOf* if it needs to enumerate objects. These concepts allow the enumeration of literals and individuals, respectively (see [19] for more details).

BooleanRVState can be replaced by the boolean data type present in OWL. Finally, the *MetaEntity* class, which includes all the entities that convey specific definitions about entities (e.g. typelabels that name the possible types of entities), can be eliminated since all other entities were replaced by a concept already present in OWL.

5 Conclusion

We described the main issues with PR-OWL probabilistic ontology language with respect to its compatibility with the OWL ontology language and presented possible approaches to deal with these issues.

The first issue described was the lack of mapping between PR-OWL random variables (RVs) and the concepts defined in OWL. In its current state, though, the relations defined in PR-OWL are not formally linked to concepts in OWL. We have shown through an example how this mapping can be done.

The second issue described was that PR-OWL does not make use of the primitive types used in OWL, as OWL did with respect to RDF and XML. For this reason, concepts already defined in one language must be redefined in the other. We have shown that every type defined in PR-OWL can be directly mapped to concepts already present in OWL without any loss of generality.

This paper has provided qualitative descriptions and examples of how to deal with these compatibility issues. We are currently working on formalizing these qualitative descriptions and on modifying PR-OWL's syntax and semantics to incorporate the approaches presented here.

Acknowledgments. The authors would like to thank the Brazilian Office of the Comptroller General (CGU) for their active support since 2008 and for providing the human resources necessary to conduct this research.

References

1. Dean Allemang and James A. Hendler. *Semantic web for the working ontologist.* Morgan Kaufmann, 2008.
2. Rommel Novaes Carvalho, Kathryn B. Laskey, Paulo C. G. Costa, Marcelo Ladeira, Laecio Lima Santos, and Shou Matsumoto. Probabilistic ontology and knowledge fusion for procurement fraud detection in brazil. In *Proceedings of the 5th Uncertainty Reasoning for the Semantic Web (URSW 2009) on the 8th International Semantic Web Conference (ISWC 2009)*, Chantilly, Virginia, USA, October 2009.
3. Huajun Chen and Zhaohui Wu. On Case-Based knowledge sharing in semantic web. In *Tools with Artificial Intelligence, IEEE International Conference on*, volume 0, page 200, Los Alamitos, CA, USA, 2003. IEEE Computer Society.
4. Huajun Chen, Zhaohui Wu, and Jiefeng Xu. KB-Grid: enabling knowledge sharing on the semantic web. In *Challenges of Large Applications in Distributed Environments, International Workshop on*, volume 0, page 70, Los Alamitos, CA, USA, 2003. IEEE Computer Society.

5. Paulo C. G Costa. *Bayesian Semantics for the Semantic Web*. PhD, George Mason University, July 2005. Brazilian Air Force.
6. Paulo Cesar Costa, Kathryn B. Laskey, and Kenneth J. Laskey. PR-OWL: a bayesian ontology language for the semantic web. In *Uncertainty Reasoning for the Semantic Web I: ISWC International Workshops, URSW 2005-2007, Revised Selected and Invited Papers*, pages 88–107. Springer-Verlag, 2008.
7. P.C.G. Costa, Kuo-Chu Chang, K.B. Laskey, and Rommel Novaes Carvalho. A multi-disciplinary approach to high level fusion in predictive situational awareness. In *Proceedings of the 12th International Conference on Information Fusion*, pages 248–255, Seattle, Washington, USA, July 2009.
8. A.-S. Dadzie, R. Bhagdev, A. Chakravarthy, S. Chapman, J. Iria, V. Lanfranchi, J. Magalhes, D. Petrelli, and F. Ciravegna. Applying semantic web technologies to knowledge sharing in aerospace engineering. *Journal of Intelligent Manufacturing*, 20(5):611–623, 2008.
9. W3C OWL Working Group. OWL 2 web ontology language document overview. http://www.w3.org/TR/2009/PR-owl2-overview-20090922/, September 2009.
10. Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, 43(5-6):907–928, 1995.
11. Patrick Hayes and Alan Rector. Defining n-ary relations on the semantic web. http://www.w3.org/TR/swbp-n-aryRelations/, 2006.
12. Ian Horrocks, Peter F Patel-Schneider, and Frank Van Harmelen. From SHIQ and RDF to OWL: the making of a web ontology language. *JOURNAL OF WEB SEMANTICS*, 1:2003, 2003.
13. Nicholas J. Kings and John Davies. Semantic web for knowledge sharing. In *Semantic Knowledge Management*, pages 103–111. 2009.
14. Kathryn B. Laskey and Paulo C. G. Costa. Of starships and klingons: Bayesian logic for the 23rd century. In *Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*, Arlington, Virginia, USA, 2005. AUAI Press.
15. Kathryn Blackmond Laskey. MEBN: a language for first-order bayesian knowledge bases. *Artif. Intell.*, 172(2-3):140–178, 2008.
16. K.B. Laskey, P. Costa, and T. Janssen. Probabilistic ontologies for knowledge fusion. In *Information Fusion, 2008 11th International Conference on*, pages 1–8, 2008.
17. Suzanne Mahoney and Kathryn B. Laskey. Constructing situation specific belief networks. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence (UAI-98)*, San Francisco, CA, 1998. Morgan Kaufmann.
18. Deborah L. McGuinness and Frank Van Harmelen. OWL web ontology language overview. http://www.w3.org/TR/owl-features/, February 2004.
19. Boris Motik, Peter F. Patel-Schneider, and Bijan Parsia. OWL 2 web ontology language structural specification and Functional-Style syntax. http://www.w3.org/TR/owl2-syntax/, October 2009.
20. Livia Predoiu and Heiner Stuckenschmidt. Probabilistic extensions of semantic web languages - a survey. In *The Semantic Web for Knowledge and Data Management: Technologies and Practices.* Idea Group Inc, 2008.
21. James Rumbaugh, Ivar Jacobson, and Grady Booch. *The Unified Modeling Language Reference Manual.* Addison-Wesley Professional, 1999.
22. Galina V Veres, Trung Dong Huynh, Mark S Nixon, Paul R Smart, and Nigel R Shadbolt. The military knowledge information fusion via semantic web technologies. Technical report, 2006.