

Building Probabilistic Ontologies Based on Meta-Model PODM

Hlel Emna^(✉), Jamoussi Salma, Turki Mohamed,
and Ben Hamadou Abdelmajid

Miracl Laboratory, Technology Center of Sfax BP 242,
3021 Sakiet Ezzit, Sfax, Tunisia
emnahlel@gmail.com, med.turki@gmail.com,
{salma.jamoussi, abdelmajid.behamadou}@isimsf.rnu.tn

Abstract. In (Hlel et al. 2016), we have presented an extension of OWL2 meta-model, called Probabilistic Ontology Definition Meta-model (PODM), for representing the fundamental elements of probabilistic ontologies (POs). Indeed, we have presented a list of new probabilistic components which allow representing the probabilistic basic elements of a domain of interest. PODM can be used by users for creating probabilistic ontologies of complex domains. In this article, we will present how we can construct probabilistic ontologies based on this meta-model.

Keywords: Probabilistic ontology · Meta-model · OWL2 · PODM · Probabilistic assertion · Uncertainty

1 Introduction

In literature, various researchers address the requirements to model the uncertain and probabilistic knowledge in the semantic web. The authors of (Predoiu and Stuckenschmidt 2010) have described some areas where probabilistic information plays a role in the context of the semantic web such as representation of uncertain information, ontology learning, etc. None of the existing ontologies languages such as RDF/RDFS, SHOE, OWL provide a means for representing this knowledge. Different probabilistic approaches for extending these languages, especially OWL, with the ability to support uncertainty are explored in literature (Yang 2007; Salvatore 2015). However, currently there is no established foundation or no standard for doing so. Moreover, these works have not focused on the proposal of a meta-model for defining the fundamental components of probabilistic ontologies which allow representing the probabilistic knowledge.

On the other hand, there are various meta-models in the literature for defining the components of classical ontologies (which allow representing the deterministic (classical) knowledge) like W3C OWL2 meta-model (ODM: Ontology Definition Meta-model for OWL2) (Motik et al. 2012). However, the components of POs (which allow representing the probabilistic and uncertain knowledge) are not taken into account. In (Hlel et al. 2016), we have presented an extension of OWL2 meta-model, called PODM, for representing the fundamental elements of probabilistic ontologies. Indeed, we have presented a list of new probabilistic components which allow

representing the probabilistic basic elements of a domain of interest. PODM can be used by users for creating POs. In this article, we will present how we can construct a probabilistic ontology of a particular domain based on this meta-model (PODM).

This article is organized as follows. In Sect. 2, we begin with a description of some related works. Next, we introduce our Probabilistic Ontology Definition Meta-model (PODM). In Sect. 4, we present a list of new probabilistic elements for supporting the uncertainty to Assertion. Then, we present our method for constructing probabilistic ontologies based on PODM. Finally, we finish with a conclusion and perspectives.

2 Related Work

The uncertainty is a ubiquitous aspect of most real world problems. It exists in almost every aspects of ontology engineering (Ding and Peng 2004). Today, there is a very interesting requirement to develop formalisms of knowledge representation allowing to deal with uncertainty. Various researchers address the need to model the probabilistic and uncertain information in the semantic web. The authors of (Predoiu and Stuckenschmidt 2010) describe five areas where probabilistic information plays a role in the context of the Semantic Web: Ontology Learning, Ontology Mapping Usage for Information Integration, Representing inherently uncertain Information, Ontology Matching and Document Classification. Despite most researchers have focused on the representation of uncertainty in ontologies; however, currently there is not established foundation or standard for doing so. In literature, there are different approaches for modeling the uncertainty in ontologies. These approaches present various extensions of Description logics and extensions of the web semantic languages for representing uncertain and probabilistic knowledge (Ding 2005; Yang and Calmet 2005; Fabio et al. 2011). None of the existing semantic web languages such as RDF/RDFS, SHOE and OWL provide a means for representing uncertain and probabilistic knowledge of real world domains. Different probabilistic approaches for extending these languages, especially OWL, with the ability to support uncertainty are explored in literature. Indeed, several Bayesian-based approaches to model uncertainty in ontologies have been proposed: BayesOWL (Ding and Peng 2004), OntoBayes (Yang 2007) and PR-OWL (Costa and Laskey 2006). BayesOWL (Ding and Peng 2004) is one proposal to represent the uncertainty in OWL ontologies through Bayesian network (BN) (Ben Mrad et al. 2015; Finn 1996). Probabilistic OWL (PR-OWL) (Costa and Laskey 2006) is a probabilistic ontology approach that is implemented on the basis of first-order logic. It is a probabilistic extension which enables OWL ontologies to represent MEBNs (Multi-Entity Bayesian Networks) (Laskey 2008). It provides a number of new OWL constructs for constructing POs probabilistic ontologies. OntoBayes (Yang 2007) is an ontology-driven uncertainty model, which integrates Bayesian network into OWL ontologies for preserving their advantages. It was developed as an extension which enables OWL ontologies to represent BNs. Indeed, the authors of (Yang 2007) have proposed an upper ontology, called Ontology OntoBayes, for representing random variables of Bayesian network, dependencies between them and probabilities associated to these variables. The representation of probabilistic knowledge in BayesOWL is performed via additional language markups, which can be simply viewed as an upper ontology (Yang 2007).

Description Logics (DLs) (Baader et al. 2003) are a family of ontological knowledge representation languages. They represent knowledge in terms of objects, concepts, and roles. To encode uncertainty, probabilistic description logics must be contemplated. The literature contains a number of proposals for probabilistic description logics (Giugno and Lukasiewicz 2002; Fabio et al. 2011). P-SHOQ is a probabilistic description logic (Giugno and Lukasiewicz 2002), extension of the DL-SHOQ (Horrocks and Sattler 2001). It adds to the syntax for SHOQ a list of conditional constraints that are defined as expressions $P(D|C) [l, u]$ with C, D are classes and $[l, u]$ is an interval between 0 and 1. These constraints can be used to represent different kinds of probabilistic knowledge, for example $(D|\{o\})[l; u]$ means “*o is an instance of the concept D with a probability in [l; u]*”. CRALC (Fabio et al. 2011) is a probabilistic description logic, extension of the DL-ALC (Schmidt-Schauss and Smolka 1991). It retains all constructors offered by ALC (conjunction, disjunction, etc.) by adding probabilistic inclusion such that $P(C | D) = \alpha$ or $P(r) = \beta$, with C and D are two concepts and r is a role.

In literature, various works have been proposed for representing POs. However, currently there is no established foundation or no standard for doing so. Moreover, these works have not focused on the proposal of a meta-model for defining POs by specifying the new probabilistic components of ontology, which allow representing the uncertainty. We think that future standard OWL2 versions should be extended in a way to allow the creation of the POs. In (Hlel et al. 2016), we have presented an extension of OWL2 meta-model, called PODM, for representing the fundamental elements of probabilistic ontologies. Indeed, we have presented a list of new probabilistic components which allow representing the probabilistic basic elements of a domain of interest. To our knowledge, this work is the first one to propose a meta-model which provides support for defining POs. So, PODM can be used by users for creating POs. In the following, we present this meta-model.

3 PODM: Probabilistic Ontology Definition Meta-Model

In (Hlel et al. 2016), we have presented an extension of OWL2 meta-model, called PODM, for representing the fundamental elements of POs. Indeed, we have presented a list of new components which allow representing the probabilistic basic elements of a domain of interest like *Probabilistic Individual*, *Probabilistic Class*, etc. (see Fig. 1). In the following, we present these components.

Probabilistic Individual. The attribution of data or objects to the corresponding concept (or class) may be uncertain. For example, “*Tom*” is an instance of class “*Animal*” with a probability equal to 0.6 and it is an instance of class “*Person*” with a probability equal to 0.4. This type of instance is called probabilistic or uncertain individual. It is associated with a probabilistic value expressing the belonging degree of an instance to a corresponding concept. Similarity to (Motik et al. 2012), we can distinguish two kinds of probabilistic individual: probabilistic named individual (identified with URI) and probabilistic anonymous individual.

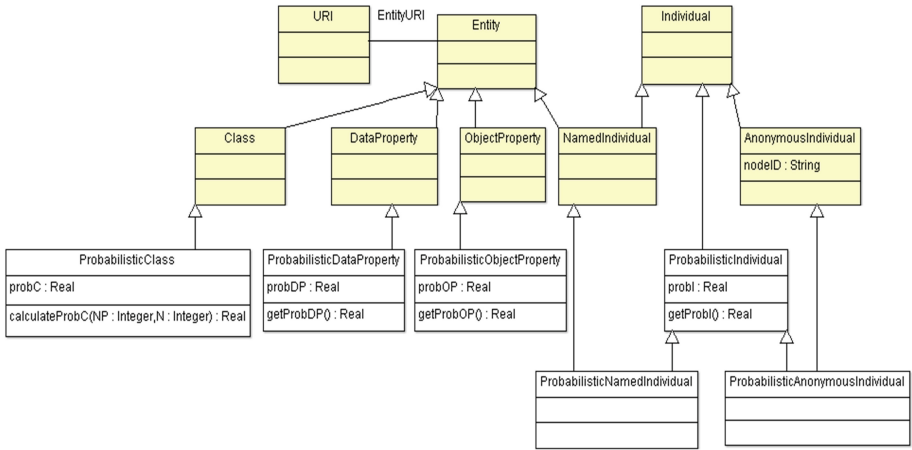


Fig. 1. Probabilistic Ontology Definition Meta-model (PODM). (Note that the classes with color white represent the new probabilistic components and the classes with color yellow represent the classical components of ODM of OWL2.)

Probabilistic class. The classes (or concepts) of OWL ontology describe a collection of objects for a particular domain. If this collection includes one or more probabilistic instances then the type of this class becomes a probabilistic class. Let C be a class of an OWL ontology and $I = \{I_1, \dots, I_i, \dots, I_n\}$ be a list of instances of this concept. In OWL ontology, we can distinguish two types of concepts: if all elements of I are classical instances then C is a classical concept and if I contains at least one probabilistic instance then C is a probabilistic concept. Assuming that C is a probabilistic concept, N is the total number of instances of this concept and NP is the number of probabilistic instances of C . This concept is attached with a probabilistic value $ProbV$ which expresses the uncertainty (Hlel et al. 2016):

$$ProbV = NP/N \in [0, 1] \tag{1}$$

Probabilistic Data Property. Generally, the extraction of knowledge in an automatic way provides us uncertain and undetermined knowledge, because the knowledge extracted by using automatic or semi-automatic systems is uncertain and probabilistic. For example, the extraction of hobby for each person can be realized automatically or semi-automatically from social networks (Facebook, Twitter, etc.). The result of this task is a list of uncertain and probabilistic knowledge (list of hobbies for each person). In an ontology, the concept “*Person*” can be used to model the set of persons. The data property “*name*” can be used to represent the name for each person. The data property “*hobby*” can be used to model the hobbies for each person. The first property is a precise element of this ontology. However, the second property is a probabilistic element of this ontology (probabilistic data property). It is attached with probabilistic value that expresses the degree of certitude of this knowledge. For example, the hobby

of “John” (instance of *Person*) is “music” (value of the probabilistic data property “hobby”) with a degree equal to 0.5.

Probabilistic Object Property. In the real world, it is often the case that the relationships between resources hold probabilistically. For example, “Imagery” (Theme) is connected to “Data-Mining” (Theme) with a probability of 0.7. In PO, this relation (“be-connected”) is considered as probabilistic object property which is associated with a probability. R is probabilistic object property between two instances if and only if it represents a probabilistic interaction between these two components.

4 Extending PODM with a Probabilistic Assertion

OWL2 supports a rich set of axioms for stating assertions (axioms about individuals that are often also called facts). In this paper, we are concentrated only to these assertions: *ClassAssertion*, *ObjectPropertyAssertion* and *DataPropertyAssertion*. We have extended PODM with a list of new probabilistic element named probabilistic Assertion (see Fig. 2) and its sub-classes (*ProbabilisticClassAssertion*, *ProbabilisticObjectPropertyAssertion* and *ProbabilisticDataPropertyAssertion*) for attaching the uncertainty to Assertion. The *ClassAssertion* axiom allows one to state that an individual is an instance of a particular class. The new axiom *ProbabilisticClassAssertion*

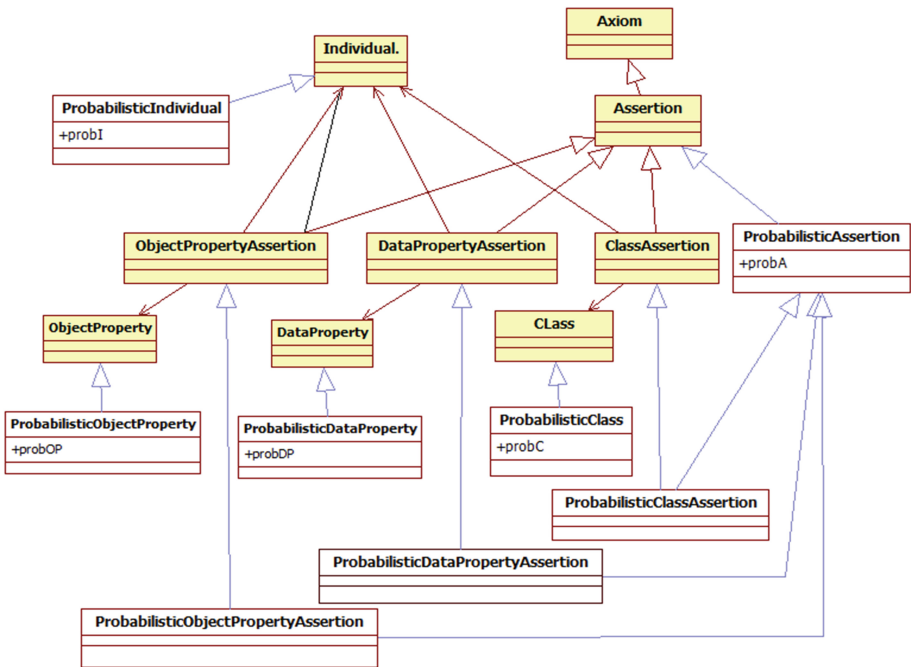


Fig. 2. Probabilistic Assertion. Note that the classes with color white represent the new probabilistic components and the classes with color yellow represent the classical components of ODM of OWL2.

allows to state that a probabilistic individual is an instance of a class with a particular probability. The *ObjectPropertyAssertion* axiom allows one to state that an individual is connected by an object property expression to an individual. The *ProbabilisticObjectPropertyAssertion* allows to state that an individual is connected by a probabilistic object property expression to an individual with a probabilistic value for example “john” is interested to “Films” with a probability of 0.9. The *DataPropertyAssertion* axiom allows one to state that an individual is connected by a data property expression to literal. The *ProbabilisticDataPropertyAssertion* axiom allows to state that an individual is connected by a probabilistic data property expression to literal for example; “Smith” prefers the hobby “travel” with a probability of 0.5 and prefers “music” with a probability of 0.4.

5 Constructing a Probabilistic Ontology by Using PODM

During the past years, the ontologies are widely used for representing knowledge of most real world domains. They provide a definition of concepts, relationships, and other features related to modeling knowledge of particular domain (Gruber 1995). Thanks to these elements, they are used to model the reality (real world applications). However, this world includes inaccuracies and imperfections which cannot be represented by classical or traditional ontologies (COs). For allowing agents to deal with uncertainty, an extension of ontologies which has the capability of supporting uncertain and probabilistic knowledge is mandatory. POs have come to remedy this defect (Costa and Laskey 2006). We can define the PO simply as a CO enriched with uncertain and probabilistic knowledge. Indeed, POs augment COs with the ability to represent the uncertainty (Hlel et al. 2015; Hlel et al. 2014). A reader interested by CO can find various works describing in detail the process of construction of ontology (CO), its components, various automatic, semi-automatic or manual construction methods of CO, etc. (Gómez-Pérez et al. 2006; Stephen and Adam 2015). However, this is not available for PO such as the majority of researchers have focused only on the proposal of extensions of probabilistic description logics and languages of semantic web to model the uncertain knowledge of a particular field (Yang 2007; Costa and Laskey 2006).

In this section, we propose a new method to guide the users for constructing POs based on PODM which provides support for defining POs. This method includes these phases: Specification of requirements, Identification and description of certain (deterministic) knowledge and uncertain (probabilistic) knowledge of domain of interest and Construction of probabilistic ontology by using a formal language. In the following, we will present these phases.

Specification of requirements. This step determines the domain and the purpose of ontology: It is important to be clear identified the purpose (goal) of the ontology. In addition, the ontologist must verify the necessity of the creation of PO through research of uncertainties and inaccuracies in the field of study. For more explaining our method of construction of a PO by using PODM, we have tried to build a probabilistic ontology, named O, which describes a list of peoples as well as their preferences (animal, music, etc.). We assume that these preferences are determined by an automatic

system from social network based on techniques of text mining and natural language processing. Generally, these techniques us provide probabilistic and uncertain knowledge. In our case, the discovered preferences are considered as probabilistic knowledge. So, it is necessary to construct a probabilistic ontology for describing a list of people as well as their preferences.

Identification of probabilistic and deterministic knowledge. In this step, we have determined and described the knowledge of domain of interest by specifying the main probabilistic components of PO (probabilistic concepts, probabilistic individual, probabilistic proprieties) and their characteristics. Moreover, it is necessary identify and

Table 1. Description of components (probabilistic and deterministic) of the ontology O.

Components of ontology		Description of components
Classical Classes	Interest	This classical concept expresses the interests of different people
Classical DataProperties	Name	The class Person is characterized with two Data Properties: name and address
	Address	
	description-Interest	The class Interest is characterized with this property description
Classical ObjectProperties	have-friend	This ObjectProperty expresses a friendly relationship between persons
Probabilistic Classes	Person	This probabilistic concept expresses a list of people. It is probabilistic because it has some probabilistic instances like “tom” and “loulou”
	Animal	This probabilistic concept expresses a list of animals. It is probabilistic because it has some probabilistic instances like “tom” and “loulou”. Note that “tom” is a probabilistic instance of Person with a probabilistic value P1 and it is a probabilistic instance of Animal with a probabilistic value P2 (same to “loulou”)
Probabilistic ObjectProperties	be-interested	This property represents a probabilistic relation between Person and Interest. It expresses that a person can be interested to one or more interests. For example, Jhon is interested to “music” with a probability of 0.5 and to “travel” with a probability of 0.2, etc
	be-connected	This property represents a probabilistic relation between the instances of class Interest. It expresses that an interest can be connected to one or more interests. For example, “watch-TV” and “Films” are connected with a probability of 0.9, etc
Probabilistic DataProperties	prefer-Animal	This DataProperty expresses that an person can prefer one or more animals. For example, John prefers cats with a probability of 0.5 and dog with a probability of 0.3, etc

describe the deterministic components necessary to satisfy the purpose of ontology (classes, properties, etc.). This step requires serious effort to analyze the domain of interest for identifying this knowledge. Table 1 resumes the probabilistic and classical components of O.

Construction of probabilistic ontology. After determining the domain and the purpose of ontology and identifying the different components of ontology (probabilistic and deterministic), the process of ontology development can be started by using a formal language such as OWL. For constructing a probabilistic ontology of a particular domain by using our proposed meta-model PODM (Hlel et al. 2016), firstly we create the probabilistic and deterministic concepts. The probabilistic classes of our example are represented as follows:

```
Declaration(ProbabilisticClass(a:Person))
Declaration(ProbabilisticClass(a:Animal))
```

Secondly, we create the probabilistic and deterministic properties of PO. The probabilistic properties of our example are represented as follows:

```
Declaration(ProbabilisticObjectProperty(a:be-connected))
Declaration(ProbabilisticObjectProperty(a:be-interested))
Declaration(ProbabilisticDataProperty(a:prefer-Animal))
```

The next step allows to create the deterministic and probabilistic instances of the ontology as well as the relations between them (populate the ontology further with instances). The probabilistic instances of our example as well as their properties are as follows:

```
//create the probabilistic individuals and its properties
Declaration(ProbabilisticNamedIndividual(a:loulou))
Declaration(ProbabilisticNamedIndividual(a:tom))
ProbabilisticClassAssertion(a:Person a:loulou 0,7^^xsd:float)
ProbabilisticClassAssertion(a:Person a:tom 0,3^^xsd:float)
ProbabilisticClassAssertion(a:Animal a:loulou 0,8^^xsd:float)
ProbabilisticClassAssertion(a:Animal a:tom "0,6^^xsd:float")
ProbabilisticPropertyAssertion(a:prefer-animal a:Smith "chat"
0,3^^xsd:float)) /* smith prefer chat with a probability with 0,3 */
//add the relations between the instances of ontology
ProbabilisticPropertyAssertion(a:be-interested a:tom a:imagery
0,5^^xsd:float))
ProbabilisticPropertyAssertion(a:be-connected a:visual-communication
a:imagery0,8^^xsd:float))
```


6 Conclusion and Perspectives

In this paper, we have presented how we can construct a probabilistic ontology for a particular domain by using our proposed meta-model of OWL2 (PODM). Our proposed method contains three phases which are Specification of requirements, Identification and description of certain (deterministic) knowledge and uncertain (probabilistic) knowledge and construction of PO. So, PODM can be used by users for creating probabilistic ontologies of complex domains.

In the future work, we will focus on the determination of probabilities which are associated to elements of ontologies for making them probabilistic. In addition, we will extend PODM with other probabilistic components: probabilistic axioms, probabilistic class expressions, etc.

References

- Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F.: *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, New York (2003)
- Ben Mrad, A., Delcroix, V., Piechowiak, S., Leicester, P., Abid, M.: An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence-uncertain evidence in Bayesian networks. *Appl. Intell.* **43**(4), 802–824 (2015)
- Costa, P.C.G., Laskey, K.B.: PR-OWL: a framework for probabilistic ontologies. In: FOIS, Amsterdam, pp. 237–249 (2006)
- Ding, Z., Peng, Y.: A probabilistic extension to ontology language OWL. In: *Proceedings HICSS*, pp. 5–8 (2004)
- Ding, Z.: *BayesOWL: a probabilistic framework for uncertainty in semantic web*. Thesis, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, USA (2005)
- Fabio, G., Rodrigo, B.P., Takiyama, F.I., Kate R.C.: Computing inferences for credal ALC terminologies. In: *URSW*, pp. 94–97 (2011)
- Finn, V.J.: *An Introduction to Bayesian Networks*. UCL Press, London (1996)
- Giugno, R., Lukasiewicz, T.: P-SHOQ(D): a probabilistic extension of SHOQ(D) for probabilistic ontologies in the semantic web. In: *INFSYS*, Austria, Research report (2002)
- Gómez-Pérez, A., Fernández, L.M., Corcho, O.: *Ontological Engineering with Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer, London (2006)
- Hlel, E., Jamoussi, S., Turki, M., Ben Hamadou, A.: Probabilistic ontology definition meta-model. In: Czarnowski, I. et al. (eds.) *Intelligent Decision Technologies. Smart Innovation, Systems and Technologies*, vol. 56, pp. 243–254 (2016)
- Hlel, E., Jamoussi, S., Hamadou, A.B.: A probabilistic ontology for the prediction of author's interests. In: Núñez, M., Nguyen, N.T., Camacho, D., Trawiński, B. (eds.) *ICCCI 2015. LNCS (LNAI)*, vol. 9330, pp. 492–501. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-24306-1_48](https://doi.org/10.1007/978-3-319-24306-1_48)
- Hlel, E., Jamoussi, S., Ben Hamadou, A.: Intégration d'un réseau bayésien dans une ontologie. In: *IC 2014*, pp. 295–297 (2014)

- Horrocks, I., Sattler, U.: Ontology reasoning in the SHOQ(D) description logic. In: 17th Conference on Artificial Intelligence (IJCAI) (2001)
- Laskey, K.B.: MEBN: a language for first-order bayesian knowledge bases. *Artif. Intell.* **172**(2–3), 140–178 (2008)
- Motik, B., Patel-Schneider, P.F., Parsia, B.: OWL2 Web Ontology Language: Structural Specification and Functional-Style Syntax (2012)
- Stephen, K.R., Adam, P.: A framework for constructing cognition ontologies using WordNet, FrameNet, and SUMO. *J. Cogn. Syst. Res.* **33**, 122–144 (2015)
- Salvatore, F.P.: An individual-centric probabilistic extension for OWL: modeling the uncertainty. In: International Conference on Computational Science, pp. 1742–1751 (2015)
- Schmidt-Schauss, M., Smolka, G.: Attributive concept descriptions with complements. *Artif. Intell.* **48**(1), 1–26 (1991)
- Predoiu, L., Stuckenschmidt, H.: Probabilistic models for the semantic web: survey. In: Web Technologies: Concepts, Methodologies, Tools, and Applications, pp. 1896–1928 (2010)
- Yang, Y.: A framework for decision support systems adapted to uncertain knowledge. Thesis, der Universität Fridericiana zu Karlsruhe (TH) (2007)
- Yang, Y., Calmet, J.: Ontobayes: an ontology-driven uncertainty model. In: IAWTIC 2005, pp. 457–464 (2005)