



HAL
open science

Linking and Negotiating Uncertainty Theories Over Linked Data

Ahmed El Amine Djebri, Andrea Tettamanzi, Fabien Gandon

► **To cite this version:**

Ahmed El Amine Djebri, Andrea Tettamanzi, Fabien Gandon. Linking and Negotiating Uncertainty Theories Over Linked Data. WWW 2019 - LDOW/LDDL Workshop of the World Wide Web Conference, May 2019, San Francisco, United States. 10.1145/3308560.3317074 . hal-02064075

HAL Id: hal-02064075

<https://hal.science/hal-02064075>

Submitted on 11 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Linking and Negotiating Uncertainty Theories Over Linked Data

Ahmed El Amine Djebri
Université Côte d'Azur, Inria, CNRS,
I3S, France
djebri.emp@gmail.com

Andrea G.B. Tettamanzi
Université Côte d'Azur, Inria, CNRS,
I3S, France
tettamanzi@univ-cotedazur.fr

Fabien Gandon
Inria, Université Côte d'Azur, CNRS,
I3S, France
fabien.Gandon@inria.fr

ABSTRACT

There is no credibility insurance measure for the information provided by the Web. In most cases, information cannot be checked for accuracy. Semantic Web technologies aimed to give structure and sense to information published on the Web and to provide us with a machine-readable data format for interlinked data. However, Semantic Web standards do not offer the possibility to represent and attach uncertainty to such data in a way allowing the reasoning over the latter. Moreover, uncertainty is context-dependent and may be represented by multiple theories which apply different calculi. In this paper, we present a new vocabulary and a framework for handling generic uncertainty representation and reasoning. The meta-Uncertainty vocabulary offers a way to represent uncertainty theories and annotate Linked Data with uncertainty information. We provide the tools to represent uncertainty calculi linked to the previous theories using the *LDScript* function scripting language. Moreover, we describe the semantics of contexts in uncertainty reasoning with meta-uncertainty. We describe the mapping between RDF triples and their uncertainty information, and we demonstrate the effect on the query writing process in *Corese*. We discuss the translatability of uncertainty theories and, finally, the negotiation of an answer annotated with uncertainty information.

CCS CONCEPTS

• Information systems → Uncertainty; Web data description languages; Ontologies; • Theory of computation → Incomplete, inconsistent, and uncertain databases.

KEYWORDS

Uncertainty, Ontology, Negotiation, Linked Data, Semantic Web

ACM Reference Format:

Ahmed El Amine Djebri, Andrea G.B. Tettamanzi, and Fabien Gandon. 2019. Linking and Negotiating Uncertainty Theories Over Linked Data. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW'19 Companion)*, May 13–17, 2019, San Francisco, CA, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3308560.3317074>

1 INTRODUCTION

The Web is a massive source of information, and the emergence of Semantic Web [3] technologies allowed the transition from a document-driven view of the Web to a data-driven one. As no credibility insurance measure is taken in most cases, information on the

Web cannot be checked for accuracy. In fact, many websites can post whatever they want including biased, subjective, incomplete or uncertain information [20]. In addition, several issues involving the modeling of uncertain Web Data and services are raised by Benslimane *et al.* [2] arguing that uncertainty should be modeled and measured in order to be associated with a piece of information or a service. Such a model prepares the ground for querying, searching, composing and harnessing uncertain data. Semantic Web technologies aimed to offer both structure and sense to the existing Web. However, Semantic Web standards do not offer the possibility to represent and attach uncertainty to such pieces of data while keeping standard semantics without further extension.

Various theories exist and can be used to model different aspects or types of uncertainty. Moreover, the choice of one theory among the others depends on the context and the application. Thus we argue that the bridge between uncertainty and the Semantic Web is crucial because of the dependence of many applications on uncertain data, the existence of many uncertainty theories, the need of interoperability and reusability of uncertain data.

First of all, the definition of uncertainty itself is challenging. It can be *epistemic*, i.e., stemming from our ignorance (incomplete knowledge, lack of a model) of an entity or process, or *ontic*, i.e., representing the inherent randomness of a phenomenon or system (a roulette, for instance). In addition, the border between these two types of uncertainty is somehow blurred and arbitrary, in that it depends on our point of view and on the level of abstraction of knowledge representation.

Multiple uncertainty theories are currently applied in Artificial Intelligence in order to reconcile the data, correct them, extrapolate or predict new values or simply assess the degree of uncertainty of information or error comparing to a dataset [14]. Probability theory is but one example, if not the best known and most time-honored, of such theories, which focuses on the representation and manipulation of the ontic uncertainty. Other theories are interested in the quantization (or “granulation”) of vague data. As the words “tall” or “cold” do not have a specific measure, they still can be modeled using fuzzy-sets and manipulated using fuzzy logic. A non-exhaustive list of uncertainty theories and their formalization can be found in [17].

To integrate uncertainty in the Semantic Web stack, two aspects must be taken into account:

- **Syntax:** uncertainty needs to be represented in order to be queried or published. The format must be machine-readable and interchangeable. The vocabulary proposed by the W3C URW3-XG group [18] enables to annotate data with the type, the model and the derivation of uncertainty. The group offers a limited list of models (Fuzzy sets, rough sets, etc.) with which neither information regarding the quantification

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

© 2019 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-6675-5/19/05.

<https://doi.org/10.1145/3308560.3317074>

of uncertainty nor the specificities of each approach and theory are provided.

- **Semantics:** reasoners are used on top of ontologies to infer new triples (based on OWL profiles) or validate the knowledge base with the schema. Direct semantics in OWL [8] is intended to give sense to ontology structures compatible with the *SROIQ* description logic. Faced with uncertainty as described below and assuming that triples can be annotated with uncertainty values, OWL does not provide the necessary tools to manipulate uncertainty, as, with the generation of new information, uncertainty must be provided too. Several extensions to the standards are done in the Schema layer itself such as FuzzyOWL [22] or PossOWL [19], while other addressed the 'Unifying logic' layer like FuzzyDL [11]. Such standards can be extended too to enable inferring new information about uncertainty or keep only triples that follow a defined set of rules in the knowledge base.

Dividino *et al.* [12] presented a framework extending both RDF and SPARQL to enable meta-knowledge querying. Their framework enables the system administrator to define meta-knowledge properties and for each, should define the intended semantics and the knowledge dimension inside the application in a non-standard format. Thus, neither semantics for the dimensions are publishable, nor semantics for uncertainty itself are provided.

Another problem that arises from the different theories is the relationship between theories and data. Some theories are suitable for specific contexts and applications. Which requires either finding the suitable theory to annotate the data, or a suitable transformation of the existing annotations to fit the requirement. Both reasons raise the question of alternatives and negotiation of content with the data sources, alongside with the possibility for users to represent their preferences in queries.

In this paper, we focus on a generic representation for uncertainty information, one that can be publishable and reusable. We summarize the *mUnc* vocabulary enabling both the representation of uncertainty theories with their calculi and the annotation of data with uncertainty information. We present an extension to RDF Semantics by giving a contextual meaning to Named Graphs. We offer the possibility to use *mUnc* alongside with our framework built on top of the *Corese* Semantic Web engine [10] and using the *LDScript* function scripting language [9] based on the SPARQL filter language, to map between triples and their uncertainty metadata, using multiple mapping modes. We discuss the translatability of uncertainty theories and propose to extend *mUnc* to fill that gap, and then discuss the negotiation of theories over HTTP as a special case of *conneg* (Content Negotiation) where clients can state their preferences in terms of uncertainty theories and servers can select or translate among theories to serve the best possible answer.

The rest of the article is organized as follows. Section 2 first recalls the uncertainty in information theory and discusses a model for the existing uncertainty theories in the literature. Then, Section 3 discusses the integration of uncertainty in the semantic web, offering a presentation of the *mUnc* vocabulary, discussing the publishing of uncertainty approaches and their calculi on the Semantic Web, and giving an overview of metadata mapping modes and how to query over uncertain Linked Data. Section 4 is dedicated to the

translatability of uncertainty theories and the negotiation details. Related works are discussed in Section 5 while the last section summarizes our work and offers a glimpse over the perspectives.

2 UNCERTAINTY IN INFORMATION

Information uncertainty can present itself in different forms depending on the domain of the definition. Data itself can be wrong from the source: entry mistakes, deceit, ambiguity, sensor errors, etc. In the context of decision-making, we link uncertainty to the outcome by defining uncertain data as "entries leading to a wrong output". Whether the data are incomplete (or imprecise), vague (or fuzzy, ambiguous), incorrect (or invalid). We find the previous specifications under other definition in the literature. For example, Dubois *et al.* [14] state that incompleteness, uncertainty, graduality, and granularity are different. They consider uncertainty as a measure for ignorance of the truth of a primitive item of information (proposition, statement, a subset of possible values, etc.), and quantified by a numerical or symbolic token located in the metalevel. In their context, the authors discuss assigning to each primitive item (fine grain) of information A , a number $g(A) \in \{0, 1\}$ which evaluates the confidence of an agent in the truth of a proposition asserting $v \in A$. The presented property is also called *confidence function*, sometimes *capacity* [7] or *fuzzy measure* [23]. This measure preserves:

- monotonicity : If $A \subseteq B$ then $g(A) \leq g(B)$;
- $g(\perp) = 0, g(\top) = 1$;

The confidence function can have multiple types according to the data in measure. It can be a *possibility*, *necessity* or even a *probability* measure. The former two measures are dual and the latter satisfies the additivity property. Several theories were established in order to formalize uncertainty, such as probability theory, possibility theory, Dempster-Shafer evidence theory, belief functions, etc. Each of the previous has a set of measures and a logic to read or to infer new values. For instance, possibility theory states that given an event A , which is a set of interpretations or outcomes, the possibility of the event to occur is given by the possibility distribution:

$$\Pi(A) = \sup_{T \in A} \pi(T)$$

While laying under the Open World Assumption, inconsistency is another challenging aspect of uncertain data where the existence of multiple interpretations of the same knowledge base leading to undecidability. As an alternative of two-valued logic, other types of logic are exploited: defeasible logic based on rules or paraconsistent logic are among those. Such alternatives offer to draw conclusions from partial and conflicting information.

Uncertainty is also affected by the context of the reading [12]. A distributive reading of uncertainty means that statements in a set provide each a portion of the uncertainty value of the whole set. Collective reading, on the other hand, assigns to the whole set an uncertainty measure. The two readings are not exhaustive, an alternative would be inheriting the uncertainty measure of the set.

Klir *et al.* [17] observe that dealing with uncertainty consists of four distinct levels: representation, calculus, measurement, and development of methodological aspects of the theory. Our focus in this paper is on the first two levels.

3 INTEGRATING UNCERTAINTY IN THE SEMANTIC WEB

Semantic Web [3] is meant for machines to understand and infer knowledge as humans do while exploring the Web. Granularizing and linking pieces of data allow machines to provide more relative content and help in the query resolution process. Schema and ontology languages offered a backbone to the existing information by providing the vocabulary of a defined topic, the relationships ruling over the different terms and the semantics allowing to logically establish links in between. Several standards (RDF, RDFS, OWL, SPARQL, etc.) of the Semantic Web technology stack enable the representation and querying of RDF data.

To integrate uncertainty in the Semantic Web, this dimension must comply with the standards and, like any other data on the web, must be reusable and publishable. In this section, we offer an ontology covering uncertainty representation, and we show the possibility of replicating the semantics of uncertainty using calculi represented in the *LDScript* function scripting language.

3.1 mUnc: a vocabulary for uncertainty theories

To enable uncertainty representation on the Semantic Web, we need to opt for an interchangeable format to write uncertainty theories. We propose *mUnc*¹, an RDFS ontology for uncertainty theories. *mUnc* (for *meta-Uncertainty*) enables publishing uncertainty information based on uncertainty theories. Figure 1 gives an overview of the core concepts and properties of *mUnc*.

We have adopted the definition of *sentence* and *world* proposed in the URW3-XG ontology. A *sentence* is an expression evaluating a truth value, while the *world* represents the context in which a *sentence* is stated. Still, unlike the previous definition, both sentences and worlds can be annotated with *meta* information. For instance, the sentence `ex:S1` representing the triple $\langle \text{ex:StefanoTacconi}, \text{dbo:height}, 188 \rangle$ referring to the height of the football player is stated in the context of the French language chapter of DBpedia [4], assuming that the latter is consistent [5]. *Uncertainty* is considered a specialization of the general concept of *meta*. This simplifies the task if any future extensions for other types of metadata such as provenance or trust are proposed. We do not include the concept of *Agent*, as it can be included using other vocabularies like the W3C PROV² Ontology.

An uncertainty theory (*Uncertainty Approach*) is linked to a set of *features* and *operators*. The *features* are the metrics on which uncertainty theory is based to indicate the degree of truth, credibility, or likelihood of a sentence. Each *feature* links a *value* to the *uncertainty* information. The *operators* represent the logic to apply to the previous *values*, while the *operations* are the implemented calculus for such logic. Other concepts in the URW3-XG ontology like the type or the derivation of uncertainty can be represented as features of an uncertainty approach.

To illustrate the previous definitions, we can annotate the previous sentence using probability theory. It can be represented using only one feature: the probability value. We choose three logical operators to include with the definition: and, or, not. The example in

```

ex:S1 a rdf:Statement;
    rdf:subject ex:StefanoTacconi;
    rdf:predicate dbo:height;
    rdf:object "188"^^xsd:decimal.

ex:Probability a munc:UncertaintyApproach;
    munc:hasUncertaintyFeature ex:probabilityValue;
    munc:hasUncertaintyOperator ex:and, ex:or, ex:not.

ex:probabilityValue a munc:uncertaintyFeature;
    rdfs:range xsd:decimal.

ex:S1 munc:hasMeta [a munc:Uncertainty;
    munc:hasUncertaintyApproach ex:Probability;
    ex:probabilityValue 0.7].

```

Listing 1: Representing uncertainty with *mUnc*

the listing 1 shows how to assert that a sentence `ex:S1` is true with a probability of 0.7. For the sake of illustration, we use reification to attach an IRI to the previous sentence, although no preference about metadata representation methods is stated [16].

3.2 Uncertainty calculi

Semantic Web ontology languages focus on classification ontological knowledge and do not support the provision of procedural attachments or functions inside ontologies. Our model allows linking the features of uncertainty approaches to their proper calculi (arithmetic, logical, comparison, etc). To represent the calculi, we rely on the *LDScript* function definition language [9], a programming language whose objects are RDF entities. It is built on top of SPARQL and relies on the SPARQL filter expression language. *LDScript* as a language permits variable declaration, assignment, function call, return, etc. Using *LDScript*, we can define functions named with an IRI and one or several arguments that are variables in the SPARQL syntax. This enables defining uncertainty operations and linking them to uncertainty features.

To continue with the previous example, considering the fact that the sentence `ex:S1` is true with a probability of 0.7, and is stated in a context `ex:C1` where all contained facts are considered to be true with a probability of 0.9. The probability of a conjunction of two supposed independent events *A* and *B* is given by Equation (1):

$$P(A \wedge B) = P(A) \times P(B). \quad (1)$$

Such value can be calculated for the user using the function referenced by `ex:multProbability` and defined in *LDScript* as shown in the following example:

```
function ex:multProbability(?pA, ?pB){?pA * ?pB}
```

Therefore, binding the function `ex:multProbability(0.7, 0.9)` during a SPARQL query execution will return 0.63. The former definition of the probabilistic approach using *mUnc* can be enriched by linking the IRI of the function to the declared feature, simply by adding the triple:

```
ex:ProbabilityValue ex:and ex:multProbability.
```

As stated before, each function is considered as a resource, due to the IRI defining its name. We can store such functions in SPARQL files all over the Web, and access their code using their reference.

¹<http://ns.inria.fr/munc/>

²<http://www.w3.org/TR/prov-o/>

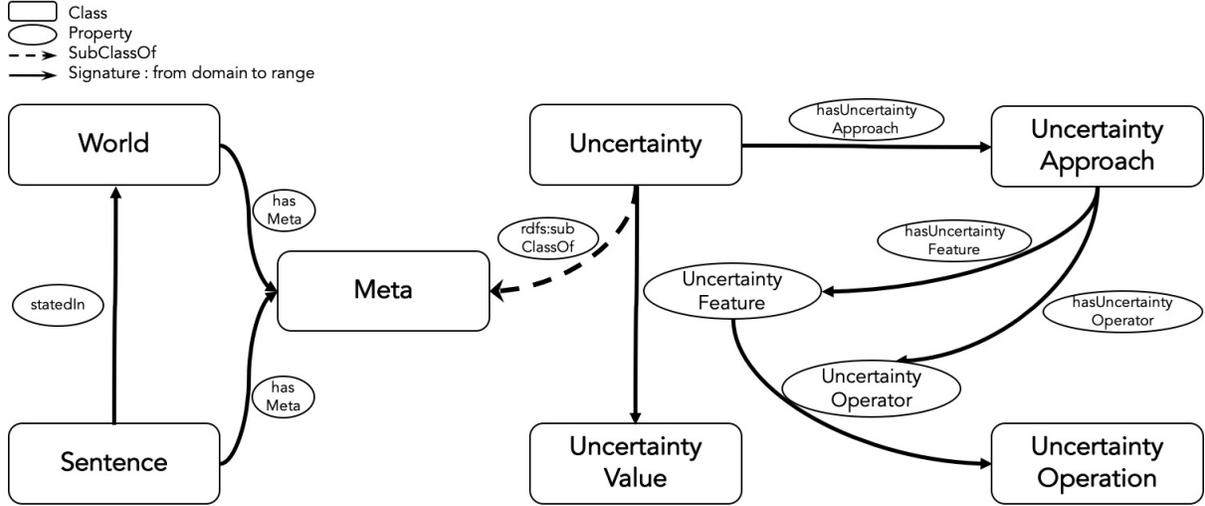


Figure 1: Overview of the *mUnc* ontology and its core concepts

3.3 Contextualizing, Mapping and Querying Uncertain Linked Data

As mentioned before, uncertainty depends on a context. A single world may accept different interpretations using different uncertainty measurements. Dubois [13] explains that beliefs are the different views upon a single world. We rely on the fact that data is issued by querying n uncertain data sources s_1, s_2, \dots, s_n , each possibly containing several contexts C_{ij} , each representing consistent information. This means that each context contains a set of triples that do not lead to contradictory reasoning.

mUnc does not provide an extension of RDF Semantics. Instead, we rely on the SPARQL query language to provide a mapping between sentences and the uncertainty information presented to the user. Moreover, we consider *mUnc* as an approach to providing definitions of known and custom uncertainty theories, for which we do not provide any specific semantics. The possibility of defining a calculus alongside with the ontology is an alternative to generalize and to reuse of the shared rules between uncertainty theories such as maximizing or minimizing a feature.

We note $\mathcal{U}_{S_{C_{ij}}}$ the uncertainty information about the sentence S cited in the context C_{ij} and $\mathcal{U}_{C_{ij}}$ the uncertainty information about the context C_{ij} . Each sentence S stated in a context C_{ij} of a source s_i , will be mapped to a combined set of pairs (*Uncertainty Feature*, *Uncertainty Value*) issued from both sentence and context metadata (noted $\mathcal{U}_{S_{C_{ij}}}$). This requires defining a metadata-mapping mode (see table 1).

The modes depend on the purpose of the application, the data itself, and the semantics of uncertainty theories. In the first mode, only uncertainty information linked to context C_{ij} is considered. The second mode considers only pairs from the lowest level of granularity, while the third mode enables inheriting context metadata but overrides the values for existing features in uncertainty information linked to the sentence.

In this paper, we use a specific meta mapping mode which relies on uncertainty calculus to evaluate a new set of pairs based on both

```

@public
function munc:metaList(?xT, ?xG){
  let(SELECT ?xT ?xG (group_concat(?FV;separator="-") as ?metaD) WHERE
  {{
    SELECT ?xT ?xG (CONCAT(?xF,'=',?xV) AS ?FV) WHERE
    {
      ?xG ?xF ?xV1
      OPTIONAL {?xT ?xF ?xV2}
      ?xF rdfs:subPropertyOf munc:uncertaintyFeature
      ?xF ex:and ?xFFFunction
      BIND(IF(BOUND(?xV2), funcall(?xFFFunction,?xV1,?xV2),?xV1) AS ?xV)
    }
  } GROUP BY ?xT ?xG)
  UNION
  {{
    SELECT ?xT ?xG (CONCAT(?xF,'=',?xV) AS ?FV) WHERE
    {
      ?xT ?xF ?xV
      ?xF rdfs:subPropertyOf munc:uncertaintyFeature
      FILTER NOT EXIST {?xG ?xF ?xV2}
    }
  } GROUP BY ?xT ?xG)
  }
  {?metaD}
}

```

Listing 2: *metaList* algorithm in *LDScript*

information from sentences and contexts. In the *metaList* algorithm we implement and use the last mode in the previous table (see Listing 2),

The `munc:metaList` function is declared in the example as “@public”. This keyword is implemented in *Corese* as many others (`@define`, `@visitor`, `@trace`, ...) defining specific routines in the Semantic Web engine. The former keyword allows the previous code to be accessed globally in the engine through its reference, and without the need to rewrite the function with each query. The listing 2 translates the `metaList` algorithm into *LDScript*. The result of binding this function in a SPARQL query is a string that groups all uncertainty features and their corresponding values from the Universal Uncertainty Information set of the corresponding sentence.

Corese also implements *Linked Functions* enabling storing *LDScript* [9] functions in external SPARQL query files on the Web. Such functions, referenced by IRIs, may be called at the moment of query

Table 1: Metadata mapping modes

Considered level of granularity	Mode	Explanation
Context	Imposed	$\hat{\mathcal{U}}_{S_{C_{ij}}} = \mathcal{U}_{C_{ij}}$
Lowest granularity only	Overriding	$\hat{\mathcal{U}}_{S_{C_{ij}}} = \begin{cases} \mathcal{U}_{S_{C_{ij}}} & \text{if } \mathcal{U}_{S_{C_{ij}}} \neq \emptyset \\ \mathcal{U}_{C_{ij}} & \text{otherwise} \end{cases}$
Lowest granularity first	Override if Exists	$\hat{\mathcal{U}}_{S_{C_{ij}}} = \mathcal{U}_{S_{C_{ij}}} \cup \{(F, V) \mid (F, V) \in \mathcal{U}_{C_{ij}}, \nexists V' (F, V') \in \mathcal{U}_{S_{C_{ij}}}\}$ $\hat{\mathcal{U}}_{S_{C_{ij}}} = \{(F, V) \mid (F, V) \in \mathcal{U}_{C_{ij}}, \nexists V' (F, V') \in \mathcal{U}_{S_{C_{ij}}}\}$ $\cup \{(F, V) \mid (F, V) \in \mathcal{U}_{S_{C_{ij}}}, \nexists V'' (F, V'') \in \mathcal{U}_{C_{ij}}\}$
Multi-level	Uncertainty Calculus	$\hat{\mathcal{U}}_{S_{C_{ij}}} = \mathcal{U}_{C_{ij}} \cup \{(F, V) \mid \exists V_{C_{ij}}, V_{S_{C_{ij}}} (F, V_{C_{ij}}) \in \mathcal{U}_{C_{ij}}, (F, V_{S_{C_{ij}}}) \in \mathcal{U}_{S_{C_{ij}}}, V = eval(Calculus(F), V_{C_{ij}}, V_{S_{C_{ij}}})\}$

execution. The former feature permits publishing and executing the calculi of uncertainty approaches. Additionally, this feature may be extended to allow the capitalization of existing software libraries from other programming languages like C++ or Java.

The Semantic Web engine also allows defining specific routines preceding the query execution. One can integrate query transformation or precalculations of some variables. We implemented the previous meta-mapping mode in extension to the visitor "@metadata" and enabled rewriting SPARQL queries to simplify querying for uncertainty information. Using "@metadata" and with `munc:metaList` publicly defined, querying for the height of the football player Stefano Tacconi in a data source is as follows.

```
@metadata
prefix ex: <http://example.org/> .
prefix munc: <http://ns.inria.fr/munc/> .

SELECT * WHERE {
  ex:StefanoTacconi dbo:height ?o
}
```

Listing 3: Query rewriting using visitors implementing the access to uncertainty information

4 NEGOTIATING UNCERTAINTY ON THE SEMANTIC WEB

In addition to the previous two-step process leading to the generation of Universal Uncertainty Information Sets alongside with sentences, users may actually have a preference for one theory or another. In this section, we will discuss the translatability between uncertainty theories and how, using HTTP content negotiation (conneg), users may negotiate the theory they want for their results.

4.1 Translating uncertainty between theories

Many examples reject the claim that uncertainty can be represented only using probability theory. However, the belief about uncertainty being the lack of information or the deficiencies due to a shortage of knowledge urge researchers to believe that it may be unified, or at least, that the different views may be linked. Dubois *et al.* [15] stated that transformation is useful in any problem where heterogeneous uncertain and imprecise data must be dealt with (e.g. subjective, linguistic-like evaluations and statistical data). Zadeh [25] cites

the example of Dempster–Shafer theory which is a theory of random sets. The latter are a probability distribution of possibility distributions. An interesting analysis of the possibility-probability transformation and its links to graphical models can be found in [1].

With the use of our framework, each and every context will issue an answer to the user. If the answers are annotated with the same theory and the same set of features, this enables ranking the results or offers more options to control the results. In the example of search engines, this could support a uniform criteria to order the results shown to the user. However, on an open Web where several open sources are queried, the results might use different features from different theories.

A translation must offer to transform a Universal Uncertainty Information Set $\hat{\mathcal{U}}_{S_{C_{ij}}}$ of a sentence S annotated following an uncertainty approach T_1 , to another set annotated with a different uncertainty approach T_2 . The translatability of theories should take into account several issues such as the symmetry, the reversibility, and the possible loss of information.

To fit in with the previous requirements, we define a translatability relationship between two uncertainty theories as follows:

Definition 4.1. A theory T_1 has a *translatability* relationship with a theory T_2 , if there exists a mapping $M : F_{T_1} \rightarrow F_{T_2}$ from the set of features F_{T_1} represented in theory T_1 to the set of features F_{T_2} represented in theory T_2 such that every possible feature of F_{T_1} is mapped to a set of feature of F_{T_2} semantically coherent with the uncertainty initially expressed in T_1 . We note $T_1 >| T_2$.

The former definition is valid for all theories that have a relationship allowing the conversion of features from one theory to another, regardless the loss of information. In case the conversion does not generate a loss of information allowing the reversibility of the operation, we define the relationship as follows:

Definition 4.2. A theory T_1 has an *ideal translatability* relationship with a theory T_2 , if T_1 is translatable to T_2 ($T_1 >| T_2$) and there is no loss of information in the translation. We note $T_1 \gg T_2$

We should mention that an ideal translatability might not be reversible, regardless the semantics of the translatability. The loss of information disables the backward operation. If the other case is considered, where we have no loss of information, then we can define a full translation as follows:

Definition 4.3. A theory T_1 has a *full translatability* relationship with a theory T_2 , iff T_1 ideally translatable to T_2 ($T_1 \gg T_2$) and, inversely, T_2 ideally translatable to T_1 ($T_2 \gg T_1$). We note $T_1 \otimes T_2$.

Using our *mUnc* vocabulary and the framework previously proposed, we are able to formalize the translation (if it exists) between the different theories. For this, we extended *mUnc* with the set of the following properties:

- `munc:hasTranslation` (definition 4.1)
- `munc:hasIdealTranslation` (definition 4.2)
- `munc:hasFullTranslation` (definition 4.3)

These QNames respectively identify the previous definitions. Figure 2 shows the extension, where each property of the previous set has for domain an uncertainty approach, and for range a blank node pointing to both the destined theory and the IRI of the translation function, written in *LDScript*.

By definition, the translativity properties have a number of algebraic properties:

LEMMA 4.4. (*transitivity of translatability*) Let $T_i, i = 1, 2, 3$ be three uncertainty theories. If $T_1 > T_2$ and $T_2 > T_3$ then $T_1 > T_3$.

LEMMA 4.5. (*reflexivity of translatability*) Each uncertainty theory has a *full translatability* with itself. We note $T \otimes T$.

LEMMA 4.6. (*Symmetry of full translatability*) Let T_1, T_2 be two uncertainty theories. If $T_1 \otimes T_2$ then $T_2 \otimes T_1$.

LEMMA 4.7. (*full translatability is an equivalence relation*) because it is transitive, symmetric and reflexive.

We can note that full translatability being an equivalence relation it allows to form equivalence classes by transitive closure i which we have translatability with no loss of information from a theory T_i to any other theory T_j of its class. We note this set $TC_{\mathcal{U}}(T_i)$.

To illustrate the previous extension, we propose to represent the example proposed in [21] about the Optimal Transformation (OT) from probability to possibility. We declare a translatability relationship between probability theory `ex:Probability` and possibility theory `ex:Possibility` representing the two different uncertainty theories. We enrich the data source with the triples below, where `ex:translateProbaToPoss` is an *LDScript* function.

```
ex:Probability munc:hasTranslation ex:Possibility.

ex:translateProbaToPoss munc:translateFrom ex:Probability.
ex:translateProbaToPoss munc:translateTo ex:Possibility.
```

4.2 Negotiation of Uncertainty Headers

Based on the previous model we can now support the possibility of negotiating answers annotated with different uncertainty theories. Content negotiation can be based on HTTP headers or non-HTTP methods such as query arguments in IRIs. Following the W3C working draft proposed by Svensson *et al.* [24] we propose that clients may negotiate a representation annotated with a specific uncertainty theory, using q -values to express their preference regarding the uncertainty theories they are to receive. Since uncertainty theories are already defined using *mUnc* and named with IRIs, both server and client can exchange and verify the conformity of their options. We propose to handle three use cases:

- (1) Uncertainty information exists in the queried source in one or many requested uncertainty theories. We answer with the first theory selected by the user. In the example, uncertainty information is issued from a context annotated with evidence.

```
GET /some/resource HTTP/1.1
Accept: text/turtle;q=0.9;uncertain="http://example.org/probability",
text/turtle;q=0.7;uncertain="http://example.org/evidence";

HTTP/1.1 200 OK
Content-Type: text/turtle;uncertain=<http://example.org/evidence>
```

- (2) The data source does not offer information about all requested theories, but a translation from existing uncertainty information to one or more requested theories is available. In this example, uncertainty is available in probability theory. The returned information are evaluated using the function `ex:translateProbaToPoss` and presented to the client with an indication about the type of translation the data underwent.

```
GET /some/resource HTTP/1.1
Accept: text/turtle;q=0.7;uncertain="http://example.org/possibility";

HTTP/1.1 200 OK
Content-Type: text/turtle;uncertain=<http://example.org/possibility>;
translation=full
```

We note that the selection of a suitable translation starts from the transitive closure of full translations $TC_{\mathcal{U}}(\text{possibility})$ offering more information and graduates to the normal translatability relationship.

- (3) The data source has no information about the theory and no available translations, we answer the user with the existing information. The default uncertainty information proposed by the server is returned in such case.

```
GET /some/resource HTTP/1.1
Accept: text/turtle;q=0.9;uncertain="http://example.org/probability",
text/turtle;q=0.7;uncertain="http://example.org/evidence";

HTTP/1.1 200 OK
Content-Type: text/turtle;uncertain=<http://example.org/possibility>;
default=true
```

5 RELATED WORKS

mUnc can extend the work done by Cabrio *et al.* [6] by enriching the proposed fuzzy labeling algorithms with definitions of other uncertainty theories that can be more suitable to the data. The Linked Data sources can adopt this approach to enrich federated queries with uncertainty information and, progressively, build a consensus-based Linked Data source. A set of other applications such as fake news detection (definition of a theory and logic for fake news), argumentation-based systems and even community-based data sources such as DBpedia can use *mUnc* to enrich their future content with uncertainty information. Furthermore, *mUnc* shares the same objectives as the W3C Credible Web community group [20] as for exchanging data which bears directly on credibility assessment while keeping standardization in data interchange.

The focus on uncertainty translatability was mainly in AI-based applications. We point that Semantic Web requirements in term of interoperability and information usage are very important and the suitability of uncertainty theories to different types of data and applications need to be further discussed.

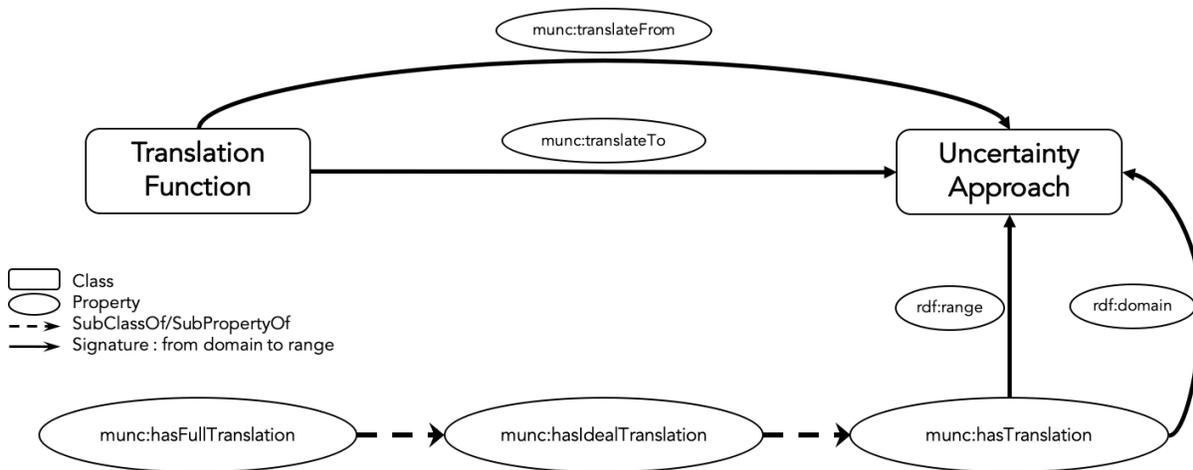


Figure 2: Extending *mUnc* ontology with translation properties

6 CONCLUSION AND PERSPECTIVES

In this paper, we discussed the representation and publication of uncertainty on the Semantic Web. We presented a vocabulary allowing the representation of uncertainty theories and the annotation of sentences using the Semantic Web standards. We explained the publishing of reusable uncertainty calculus using *LDScript*. We also offered the possibility to translate between uncertainty theories and to negotiate uncertainty information following a specific theory.

Uncertainty representation is the first step of a long process, including the preliminary calculus of uncertainty values and the propagation of uncertainty among interconnected Linked Data sources. The translation process is also a first step enabling, to some extent, merging uncertain data annotated using different uncertainty approaches. In our future work, we would like to implement context overlapping, allowing the selectivity inside the source between contexts and the optimization of the storage. We would also study the relationship between data, applications, and uncertainty theories that are used.

ACKNOWLEDGMENTS

We thank Olivier Corby for his insights about *Corese* and *LDScript*. This work is supported by the Algerian government.

REFERENCES

- [1] Salem Benferhat, Amélie Levray, and Karim Tabia. 2015. On the analysis of probability-possibility transformations: Changing operations and graphical models. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer, 279–289.
- [2] Djamal Benslimane, Quan Z Sheng, Mahmoud Barhamgi, and Henri Prade. 2016. The uncertain Web: Concepts, challenges, and current solutions. *ACM Transactions on Internet Technology (TOIT)* 16, 1 (2016), 1.
- [3] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific american* 284, 5 (2001), 34–43.
- [4] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. DBpedia-A crystallization point for the Web of Data. *Web Semantics: science, services and agents on the world wide web* 7, 3 (2009), 154–165.
- [5] Elena Cabrio, Serena Villata, and Fabien Gandon. 2014. Classifying Inconsistencies in DBpedia Language Specific Chapters.. In *LREC*. 1443–1450.
- [6] Elena Cabrio, Serena Villata, and Alessio Palmero Arosio. 2017. A RADAR for information reconciliation in Question Answering systems over Linked Data 1. *Semantic Web* 8, 4 (2017), 601–617.
- [7] Gustave Choquet. 1953. Theory of capacities. *Annales de l’institut Fourier* 5 (1953), 131–295. <https://doi.org/10.5802/aif.53>
- [8] World Wide Web Consortium et al. 2012. OWL 2 web ontology language document overview. (2012).
- [9] Olivier Corby, Catherine Faron-Zucker, and Fabien Gandon. 2017. *LDScript: a Linked Data Script Language*. In *International Semantic Web Conference*. Springer, 208–224.
- [10] Olivier Corby and Catherine Faron Zucker. 2002. Corese: A corporate semantic web engine. In *International Workshop on Real World RDF and Semantic Web Applications, International World Wide Web Conference*.
- [11] Valerie Cross and Shangye Chen. 2018. Fuzzy Ontologies: State of the Art Revisited. In *Fuzzy Information Processing*, Guilherme A. Barreto and Ricardo Coelho (Eds.). Springer International Publishing, Cham, 230–242.
- [12] Renata Dividino, Sergej Sizov, Steffen Staab, and Bernhard Schuele. 2009. Querying for provenance, trust, uncertainty and other meta knowledge in RDF. *Web Semantics: Science, Services and Agents on the World Wide Web* 7, 3 (2009), 204–219.
- [13] Didier Dubois. 2011. Information Fusion and Revision in Qualitative and Quantitative Settings. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer, 1–18.
- [14] Didier Dubois and Henri Prade. 2010. Formal Representations of Uncertainty. In *Decision-making Process: Concepts and Methods*. 85–156. <https://doi.org/10.1002/9780470611876.ch3>
- [15] Didier Dubois, Henri Prade, and Sandra Sandri. 1993. On possibility/probability transformations. In *Fuzzy logic*. Springer, 103–112.
- [16] Johannes Frey, Kay Müller, Sebastian Hellmann, Erhard Rahm, and Maria-Esther Vidal. 2017. Evaluation of Metadata Representations in RDF stores. *Semantic Web Journal* (2017).
- [17] George J Klir and Richard M Smith. 2001. On measuring uncertainty and uncertainty-based information: recent developments. *Annals of Mathematics and Artificial Intelligence* 32, 1-4 (2001), 5–33.
- [18] Kenneth J Laskey and Kathryn B Laskey. 2008. Uncertainty Reasoning for the World Wide Web: Report on the URW3-XG Incubator Group. In *URSW*. Citeseer.
- [19] Bal-Bourai Safia and Mokhtari Aicha. 2014. Poss-OWL 2: Possibilistic Extension of OWL 2 for an uncertain geographic ontology. *Procedia Computer Science* 35 (2014), 407–416.
- [20] Sandro Hawke. 2018. Technological Approaches to Improving Credibility Assessment on the Web. <https://credweb.org/report/20181011>. Accessed: 2019-01-20.
- [21] Yosra Ben Slimen, Raouia Ayachi, and Nahla Ben Amor. 2013. Probability-possibility transformation. In *International Workshop on Fuzzy Logic and Applications*. Springer, 122–130.
- [22] Giorgos Stoilos, Giorgos B Stamou, Vassilis Tzouvaras, Jeff Z Pan, and Ian Horrocks. 2005. Fuzzy OWL: Uncertainty and the Semantic Web.. In *OWLED*.
- [23] Michio Sugeno. 1977. Fuzzy measures and fuzzy integrals: A survey. 78 (01 1977).
- [24] Svensson, Lars G. and Atkinson, Rob and Car, Nicholas J. 2018. *HTML 5.3*. W3C Working Draft. W3C. <https://www.w3.org/TR/dx-prof-conneg/>.
- [25] Lotfi A Zadeh. 2011. Generalized Theory of Uncertainty: Principal Concepts and Ideas. In *Fundamental Uncertainty*. Springer, 104–150.