An approach to express measurement results in ontologies of facts

Eric Benoit, Cedric Deffo Sikounmo, Stephane Perrin
Univ. Savoie Mont Blanc, LISTIC, F-74940 Annecy, France
E-mail: eric.benoit@univ-smb.fr

Abstract. Ontologies give a convenient framework to model complex systems. They are especially efficient to perform reasoning on these systems but the alignment of ontology based model with the empirical world can be questionable. Considering that alignments of a model with the empirical world is performed by the way of a measurement, the uncertainty related to such process must be considered. The goal of this paper is to present a solution that includes measurement results in an ontology even if this one modifies the structure of the ontology.

1. Introduction
The fields concerned by the modeling of complex systems are multiples. Not only physics, but also social sciences [1], software design, biology, navigation [2], get benefits from the ontological approaches. Independently from the field, the measurement process can be seen as a way to update a model i.e. to update an informational world with the observations of the empirical world [3]. The ontologies of fact, versus the ontologies of concepts, are specifically adapted to the modeling of the empirical world. Populated with abstract entities representing the real entities of the studied field, they link them with relations and associate properties to them. Some of these properties are measurands, i.e. quantities under measurement. According to the measurement science, a measurand can have several states, each one materialized by a manifestation. The observation of the measurand produces indications and a measurement result is an abstract representation of these states and is inferred from the indications. Following an erroneous popular practice, most of factual ontologies represent a measurement result by a single indication and do not take into account the measurement uncertainty. After reminding the modeling of empirical worlds with ontologies of facts, this paper exposes the specific issue related to the influence of a measurement result on the ontology structure. Then a solution is proposed to take this influence into account.

2. Ontology and measurement
2.1. Ontology of facts
An ontology of facts, also called factual ontology, is an ontology based on an empirical world. It is used to synthesize the knowledge about entities of the real world and about their mutual relationships. The RDF ontologies are especially adapted to the modeling of empirical worlds. A RDF ontology is defined by a set $N$ of entities, a set $L$ of literals i.e. terms or numbers, a set $R$ of relations and it is made of instances of these relations on $N \times (N \cup L)$. Each instance
named relationship or relation item is denoted $<r, a, b>$ where $r$ is the relation, $a$ is the entity and $b$ is the entity or the literal in relation with $a$.

2.2. Measurement uncertainty

Within a classical approach, the knowledge about a measurement result is represented by a probability distribution on a universe of discourse. When a measurement result concerns only one entity the statistical approach allows to represent it with a mean value and a standard deviation that are expressed as literals on an ontology. For example, the measurement result of a table length is expressed by the piece of ontology as presented on Fig. 1: $\{ <\text{hasLength}, \text{table}, \text{anon}_1 >, <\text{hasMean}, \text{anon}_1, 3.44 >, <\text{hasUnit}, \text{anon}_1, \text{m} >, <\text{hasStandardDeviation}, \text{anon}_1, 0.012 > \}$. The term $\text{anon}_1$ represents an unnamed entity also called anonymous entity. It is used to concentrate the properties related to the length. After another measurement, the literals of this ontology are simply updated to reflect the new knowledge.

Unfortunately, some measurement results can affect the structure of the ontology. This appears when the manifestations of a quantity are represented by entities into the ontology or when they are linked to entities into the ontology. This situation appears mainly with measurements based on nominal or ordinal scales. For example, if the result of a recognition process is interpreted as a measurement result, the different manifestations of the measurand are linked to the different entities to recognize. Let’s consider a practical situation including a camera, able to process face recognition, and people such that somebody can stand in front of the camera. The measurand is a picture and the people are the entities to be recognized and are related to the manifestations of this measurand. The existence of the relation item $<\text{inFrontOf}, \text{human}_1, \text{camera}_1 >$ then depends on the measurement result. But in fact, the situation is more complex. According to a probabilistic approach of measurement, the measurement result may be expressed as a probability distribution (PDF). Similarly, a possibilistic approach will express the measurement result as a possibility distribution. As the people can be involved in relations with other entities than the camera, the distribution cannot be simply shifted on the people by the use of literals. One solution can be to add in the ontology a new entity in similar ways than [4] or [5]. This solution has the disadvantage to increase dramatically the number of entities in the ontology.

Another solution is to shift the distribution on a set of relation items $\{ <\text{inFrontOf}, \text{human}_x, \text{camera}_1 > | x = 1, \ldots, n \}$. Most of ontologies do not offer this possibility, but the concept of infons can solve this issue.

2.3. The concept of infons

In [6], Devlin presents a discrete informational item called infon and defined by $\ll R, a_1, \ldots, a_n, \text{polarity} \gg$ where $R$ is an n-ary relation and $a_1, \ldots, a_n$ are objects appropriate for $R$. The polarity takes the values 0 or 1 and reflects the membership of $a_1, \ldots, a_n$ to the
relation $R$. The formalization of this membership actually expresses a truth value of the relation item $< R, a_1, \ldots, a_n >$, also called relationship [7]. Note that with a restriction to binary relations, infons are simply defined by a RDF item $< R, a, b >$ (Usually denoted $< a, R, b >$ but we keep in this paper the notation close to the original infons notation for the matter of consistency).

So the polarity is a binary value that has the meaning of a truth of a piece of knowledge expressed by a relation item. According to the various approaches to express knowledge, we propose first to extend the polarity value domain to $[0, 1]$. The polarity then reflects the fuzzy membership of $(a_1, \ldots, a_n)$ to the relation $R$. Then the polarity is able to hold a possibility degree: $\ll R, a, b, \pi \gg$.

Extending the polarity value domain to any positive number, allows us now to use the polarity to hold a probability density: $\ll R, a, b, P \gg$.

3. Proposal
3.1. Defining the universe of discourse
In the measurement process, the universe of discourse is the set on which the measurement results are expressed. The set of real numbers is the universe of discourse for interval scales and ratio scales. For nominal scales, the universe of discourse is a set of entities that can be anything including numbers. For ordinal scales the universe of discourse is a set of ordered entities. We previously proposed to express the distribution characterizing a measurement result on a set of relation items $\{ < \text{inFront, camera}_1, \text{human}_x > \mid x = 1, \ldots, n \}$. This solution can be performed by the use of infons like $\ll \text{inFront, camera}_1, \text{human}_x, 0.221 \gg$ but presents a critical issue: the advantage of using an ontology to model a given system is to include a large set of entities and multiple measurement results. The counterpart is the possibility for a given relation item to be related to several measurements. In this case, multiple universes of discourse can have non empty intersection.

Our main proposal is to uniquely identify each universe of discourse with an unnamed node and an original relation possibleSameAs. The idea is to split the relation item $< R, a, b >$ into 2 relation items $< R, a, \text{anon}_1 >$ and $< \text{possibleSameAs, anon}_1, b >$. With this solution, each measurement process is associated to an unique relation item $< R, a, \text{anon}_1 >$ and the universe of discourse is the set of all relation items $< \text{possibleSameAs, anon}_1, b >$. The name of the possibleSameAs relation has been chosen due to its meaning. It expresses the possibility for an entity to be replaced by another one and actually matches with the binary version of the possibility degree as defined in the possibility theory.

3.2. Management of uncertainty
Given an entity $a$ and an unknown entity linked to $a$ by a relation $R$. $< R, a, \text{anon}_1 >$ expresses the knowledge on the fact that some entity is linked to $a$ by the relation $R$, and $\{ < \text{possibleSameAs, anon}_1, b_x > \mid x = 1, \ldots, n \}$ is the universe of discourse (see Fig. 2). Then $\{ \ll \text{possibleSameAs, anon}_1, b_x, d(x) \gg \mid x = 1, \ldots, n \}$ is the distribution expressing the knowledge on the fact that a given entity $b_x$ is linked to $a$ by the relation $R$. The distribution can be either a probability distribution $\{ \ll \text{possibleSameAs, anon}_1, b_x, P(x) \gg \mid x = 1, \ldots, n \}$ or a possibility distribution $\{ \ll \text{possibleSameAs, anon}_1, b_x, \pi(x) \gg \mid x = 1, \ldots, n \}$.

Finally, the representation of the measurement uncertainty related to nominal scales and ordinal scales, is a particular case where the relation $< \text{measurementResult, measurand, anon}_1 >$ models a measurement and the distribution $\{ \ll \text{possibleSameAs, anon}_1, \text{indication}_x, d(x) \gg \mid x = 1, \ldots, n \}$ models the measurement result.

This way to represent the measurement uncertainty in ontologies allows to represent on the same ontology multiple measurement results of the same measurand. It also allows to aggregate these measurement results because these one are related to the same universe of discourse. By
Figure 2. General structure for the universe of discourse. The measurement result is defined as a distribution on indications. The possibleSameAs relation items are introduced to model the support of this distribution: the universe of discourse.

Figure 3. After the reasoning process, the possibleSameAs relation links an unknown entity (a human in the example) to the potential candidates.

extension, the impact of a measurement result on the structure of the ontology preserves the information on the uncertainty by the use of the possibleSameAs relation (see Fig. 3). In the previous example, the expression of the uncertainty on the identity of the human being in front of the camera is given by the relation \(< \text{InFront}, \text{camera}, \text{anon}_1 >\) and the infons \(\ll \text{possibleSameAs}, \text{anon}_1, \text{human}_x, P(x) \gg\) where \(P(x)\) is the probability for \(\text{human}_x\) to be the good candidate.

4. Conclusion

The necessity to update ontologies of facts used to model empirical systems imposes to include into this representational tool the main concepts of measurement science i.e. the measurement uncertainty. This paper proposed a solution to take into account the situations where this uncertainty appears at a macroscopic level into the ontology and has an influence on its structure. This approach gives a convenient solution to manage uncertainty in ontology based approaches with a minimal impact on the complexity of the model.

References

[1] Epstein B 2016 Philosophy of the Social Sciences 46 147–167
[2] Yang L and Worboys M 2011 Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness (ISA ’11 no November) (New York, NY, USA: ACM) pp 31–34 ISBN 978-1-4503-1035-2
[3] Mari L 2003 Measurement 34 17–30
[4] Bellenger A and Gatepaille S 2011 URL http://arxiv.org/abs/1106.3876
[5] Ding Z, Peng Y and Pan R 2006 Soft Computing in Ontologies and Semantic Web (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 3–29
[6] Devlin K 2006 Handbook of the History of Logic 7 601–664 ISSN 18745857
[7] Kokar M M, Matheus C J and Baclawski K 2009 Information Fusion 10 83–98 ISSN 1566-2535