

# Introducing Context into Semantic Web Knowledge Bases

Heiko Stoermer

University of Trento,  
Dept. of Information and Communication Tech.,  
Via Sommarive 10, 38050 Trento, Italy  
stoermer@dit.unitn.it

**Abstract.** Knowledge Representation in the Semantic Web is mainly characterized by the capabilities of the languages OWL and RDF. When attacking non-trivial problems such as the creation and maintenance of large-scale knowledge bases in RDF, it becomes evident that RDF lacks a central feature, namely the capability of restricting the truth value of a statement to a context, with far-reaching consequences which we will hint at in this paper. We propose a solution approach to this problem: the extension of RDF knowledge bases with context features. We will present theoretical background, envisioned experiments as well as a comparative study based on two technically different approaches, and illustrate the applications and benefits of contextual Knowledge Representation in the Semantic Web.

**Keywords:** Context, Knowledge Representation, Semantic Web, RDF

## 1 Introduction

The general setting of this thesis proposal is the area of the Semantic Web. Since the first concrete published ideas in this direction in 2001 [3], many interesting and progressive technologies have emerged to pursue the vision of a Web that will not only simply contain data but semantic information that is machine-processable in a *meaningful* way.

The original approach was to keep things simple, as for the authors of [3] it was obvious that complex approaches in Artificial Intelligence (AI) and especially in the field of Human Language Technology (HLT) had not advanced to a level that could be called usable for the desired machine interpretation of Web documents authored by humans.

So at the base of the layered architecture sketched in [3] there was supposed to be the meaningful annotation of Web documents to transform them into Semantic Web Documents. Technologies such as RDF and RDF Schema [10] were among the first to come into existence – RDF to annotate documents and RDFS the define the regarding annotation vocabulary. Together with OWL – a language for conceptual modeling – these two languages and the related

tools play a central role in the Semantic Web and have become standards for representing what a machine can *know* about a document (and the world).

From the original idea of rather simple (and possibly manual) annotation, things have developed farther, not only through the influence of the AI community, on who's work for example in Description Logics<sup>1</sup> and Automated Reasoning the different variants of the OWL language are based.

More recent projects in the Semantic Web use a more self-confident vocabulary and talk about *knowledge bases* instead of metadata, and use inference techniques and Automated Reasoning to provide semantic-enabled services. This development has brought the efforts of the Semantic Web closer to research that has been performed for many years in the area of AI and specifically Knowledge Representation (KR). This growing overlap opens opportunities for applying theories from AI to the Semantic Web to ensure that the efforts taken so far receive the best theoretical base they can have and also to contribute new findings back to the AI field.

This thesis proposal will deal with one aspect of KR, the notion of *context*, and how to improve current KR approaches in the Semantic Web by incorporating context theories to RDF Knowledge Bases. I will illustrate motivations of why sensible KR for the Semantic Web should be based on contexts, and underline this opinion by outlining requirements of the large ongoing European project VIKEF<sup>2</sup> that I am involved in. I will furthermore give details about work that has already been performed in this area, by myself and other people, as well as describe the vision for the proposed research in more depth. Finally, I will deliver an outline of which benefits and applications a good approach in this topic will be able to provide.

## 2 Problem Description and Motivation

The VIKEF project deals with creating large-scale information systems that base on Semantic Web technology. At the center of the envisioned systems there is an RDF knowledge base (KB) that contains a large amount of information about documents and their contents. This information is stored in an RDF triple store that has been developed in the course of the project by other partners in the consortium. The triple store has originally been envisioned to be a single bag of RDF triples, i.e. all statements are stored in the same information space together. However, from a Knowledge Representation point of view RDF statements in general are context-free, and thus represent statements of *universal truth*, while documents contain *context sensitive information*, i.e. information whose interpretation depends on the context in which the document was written. In effect, the original way to proceed can easily generate contradictory statements to be stored in the KB, such that for instance "Silvio Berlusconi is the Prime minister

---

<sup>1</sup> As one of many relevant sources for information about Description Logics please refer to [1]

<sup>2</sup> For more information about the VIKEF architecture, features and application scenarios please refer to the project's website: <http://www.vikef.net>

of Italy” and “Romano Prodi is the Prime minister of Italy” as the result of articles written at different points in time. Such phenomena are however unwanted in a logical system because they would seriously interfere with the higher level reasoning that is to be performed to provide the envisioned services. Additionally, we would like to be able to model other aspects such as relevance, credibility and validity of a statement, all of which require further qualification of a single statement or a defined set of statements.

An innumerable amount of similar examples can be constructed for a single information system, although in such a system it would at least be possible to implement some heuristics to try to keep track of these issues, at least at the time of insertion of a triple, to minimize the problem. However, if we think about the Semantic Web as a whole, with a large number of uncoordinated information systems, the problem becomes even more evident. If every peer builds up a KB of unqualified RDF statements, the set of universally true facts in the Semantic Web becomes enormously large and impossible to handle from a semantic point of view. In our opinion, such contradictions, contradictory beliefs and facts which become semantically incorrect in the absence of additional pragmatic or contextual information are likely to impose serious problems on the coordination and interoperation of information systems in the Semantic Web.

### 3 Research Proposal: Context in RDF Knowledge Bases

We think that the mentioned issues can be attacked by introducing the notion of *context* into RDF, to limit the scope of an RDF statement to the context in which it is relevant. Especially, we want to model that a statement is true *only under a certain set of conditions*, which will help us store information in the KB that would cause contradictions or inconsistencies in a non-contextual RDF KB.

#### 3.1 Context in KR - Multi Context Systems

The ideas presented in this paper base on the logical theory of *Multi Context Systems* (MCS) and the principles of *Locality* and *Compatibility* presented e.g. in [11], with influences from [8, 9]. Basically, this theory states that contexts can be seen in a peer-to-peer view, resembling more general aspects such as human beliefs, agent knowledge and other distributed systems. The important aspect is that reasoning within a context follows standards mechanisms, as the non-elementary view on the axioms does not require to keep track of the context they are relevant for. Relations between contexts however, i.e. to reason across contexts, are to be expressed in so-called *compatibility relations* (CRs), that formalize exactly how under certain circumstances knowledge from other contexts becomes relevant. Regarding RDF in this case we claim that a context can be thought of as a locally coherent set of axioms, and we envision CRs to be modeled as a semantic attachment [16].

### 3.2 Basic Work

The basic idea is to have all statements that belong to a context in a separate named RDF graph, and extend the RDF semantics in a way to enable contexts to appear as standard objects in RDF statements of other contexts. One way to achieve this has been described in [6].

Then, we want to model the mentioned *compatibility relations* (CRs) between contexts, to allow for reasoning across contexts. This aspect is probably the most important one, because from an application perspective it is crucial that sensible queries can be issued and all relevant information is taken into account - which requires reasoning across contexts and reasoning on the relations between contexts.

Several approaches can be thought of to model compatibility relations in our architecture. First of all, one could think of allowing the implementer of an information system to provide their own vocabularies (ontologies) to describe relations between contexts. A similar option would be for us to provide such an ontology as part of the architecture. However, in our opinion the basic problem with these approaches is the fact that many interesting relations between architectures semantically cannot be fully formalized with the help of a Semantic Web ontology, which is based on Description Logics<sup>3</sup>.

One of the questions that might arise is how these compatibility relations are supposed to be modeled. At the moment, we see three approaches to do this:

1. Formally define a fixed set of compatibility relations as part of our architecture and require all systems that implement the architecture to take care of providing also an implementation of the relations.
2. Provide an ontology for context relations, so that there exists a vocabulary to describe these relations with the help of RDF. This approach is slightly more flexible, because the ontology would be extendable.
3. Define a CR to be implemented as a *semantic attachment* [16], which can be thought of as a sort of plugin to the system, one attachment per CR. This has the positive effects that i) there is no restriction on how many and which kind of CRs are part of such a system and ii) implementation of the CRs is generally not restricted to any specific language or system.

Very basic and preliminary results including some of the above ideas have been presented in [6].

### 3.3 Open Issues and Further Work

As we are only at the very beginning of our work, there are numerous open issues and questions to be explored. We will only list a few, to illustrate the direction

<sup>3</sup> As an example for this claim take a relation such as  $\langle c' \text{ EXTENDS } c \rangle$ . The semantics of this relation have to be expressed algorithmically:  $c$  and  $c'$  are taken to be compatible in the sense that one does not contain facts that contradict with facts of the other and that the relevant context parameters are the same; then, if no answer to a query to  $c$  can be given, the query will be propagated to  $c'$ .

of the proposed work. To begin with, there are aspects about the underlying theory that in our opinion have to be cleared. Among these are:

- What are the semantics of *overlapping* contexts and how are they to be modeled?
- What are the semantics of adding a statement to the context, or adding a statement to the description of a context?
- In which way are statements *in* a context affected by statements *about* a context?
- How do we discover that a statement is not coherent with a context and thus may not be added to it?
- Should contexts be transparent or opaque to a query language?

Secondly, as one of the major aspects of the research to be performed, we intend to invest a reasonable amount of effort into the exploration of compatibility relations. As already mentioned before, there are several implementation options that come to mind, but we want to make sure all possibilities are thought of. Additionally, one of our objectives is also to see if there are limitations to the Multi Context Systems theory, what their nature is and if they can be overcome. Regarding the relations between contexts themselves, we would like to investigate if there are general relations that can always be assumed – such as for example the ones existing in relational database systems – and how we want to deal with application-specific relations. It is our goal to explore the possibilities that the theory offers, and to define the formal properties of relations we find to be relevant.

### 3.4 Comparative and Experimental Objectives

As a further step in addition to the more theoretical aspects mentioned before, we are currently establishing close collaboration with two other research groups to lay the base for comparative experimental work that is envisioned to be part of the outcomes of the proposed dissertation. With the group responsible for developing the mentioned RDF triple store we are currently working on an implementation based on *RDF named graphs*, which we have described in [15]. Additionally, with the group behind the RDF-based P2P system DBin<sup>4</sup> we will try to develop a second implementation based on RDF's reification feature, to explore if there are ways to overcome the limitations mentioned later in Sect. 4.

A substantial part of the proposed work will be the comparison and evaluation of the two prototypes regarding soundness, limitations, performance and development effort.

The setting of the proposed work within a large-scale research project is very fortunate and beneficial for this kind of comparative studies. We expect very large datasets to be generated and available for intense evaluation, which will hopefully make the results interesting and highly relevant for the Semantic Web community.

---

<sup>4</sup> <http://www.dbin.org/>

## 4 State of the Art and Related Work

We are certainly not the first to raise the issue of underqualified statements in RDF. The straight forward approach to tackle this problem would be to use the reification capability of RDF: for every statement inserted into an RDF graph we also insert a number of meta-statements about this statement, containing all relevant context parameters, e.g.:

- <1996 IsTheTimeOf "<RomanoProdi PrimeMinisterOf Italy>">
- <2005 IsTheTimeOf "<SilvioBerlusconi PrimeMinisterOf Italy>">

There exist different opinions about the questions whether parameters describing a context can be limited or not [2]. In any case, this approach would be implementable using standard RDF. However, with a potentially unlimited number of context parameters, we foresee a statement explosion with the this kind of approach, because for every statement we would have to add a significant number of meta-statements describing the relevant context dimensions, so the overhead is immense.

A completely different approach of representing contexts in RDF is to extend RDF with the ability to represent a reference to a context directly in the data model. There have been proposals in the past, by Guha [12] or Klyne [14, 13], which do not use reification but implement context as a real extension of the RDF model theory, by moving from triples to quadruples for identifying the context to which a statement belongs. To the best of our knowledge, these ideas have not been pursued any further. Moreover all the currently available RDF tools would have to be extended in order to deal with such an RDF model.

One related approach that has lead to actual results is that of the W3C Named Graph Interest Group<sup>5</sup>. There has been published a substantial article in 2005 [7], and implementational results are now part of the Named Graphs API for Jena (NG4J)<sup>6</sup>, with Jena being the de-facto standard for the development of software that deals with RDF. We think that this approach it could be well used as an underlying implementation for the smaller part of what we want to do, and to be extended for the very important aspect for Compatibility Relations discussed above.

From a different point of view, we are also trying to discover the relations between our ideas and works from the field of Description Logics, e.g. Context-OWL [5] and Distributed Description Logics [4]. These works are especially important with regards to Compatibility Relations and the modelling of contexts in general.

## 5 Benefits and Applications

The benefits of a good solution for the issues mentioned in Sect. 2 would in our opinion be the base for advanced Knowledge Representation in the Semantic Web. We hope to provide three major outcomes with our work:

<sup>5</sup> <http://www.w3.org/2004/03/trix/>

<sup>6</sup> <http://www.wiwiss.fu-berlin.de/suhl/bizer/ng4j/>

1. A detailed solution to the problem of modeling contexts in the Semantic Web in a coherent and general way.
2. An evaluation of the MCS theory. We would be able to put this theory to the test, and explore its limitations.
3. Provision of comparative experimentation results, to illustrate which possibilities exist, how they behave and whether they prove appropriate for real-world applications.

Possible applications for this kind of knowledge representation are manifold. Aspects such as beliefs, trust, incomplete knowledge or knowledge base evolution in our opinion can all be tackled with a sensible context system as a base. We believe that in the long run, the vast amount of knowledge represented in the Semantic Web can be handled much easier if represented *in context*.

However, we envision the outcomes of this work to go beyond local aspects and also become relevant from a distributed point of view. As the nature of the Semantic Web is inherently distributed, we think we can contribute to the semantic coordination of Semantic Web agents, firstly by offering the capabilities to make explicit that two knowledge bases belong to their respective agents and to enable the agents to establish semantic links to the KBs of other peers with the help of compatibility relations.

## 6 Conclusion

In this proposal I have presented an envisioned framework of representing context in the Semantic Web, as a way to make the currently available methods of knowledge representation in this area become more precise, meaningful and sensible. I have illustrated the motivations, theoretical and practical, as well as the beneficial setting of the proposed work within a large European research project. We have already contributed to the community with a preliminary publication that received positive response. Certainly, there are many open issues that will have to be explored. But we are establishing the relevant cooperations to provide for a high level of expertise in this area and to provide not only a theoretical contribution but also practical results to illustrate how contextual knowledge representation in the Semantic Web can help to overcome a number of existing issues and limitations.

## 7 Acknowledgements

The work described in this paper has been partly funded by the European Commission through grant to the project VIKEF under the number IST-507173. I would like to thank Luciano Serafini and my advisor Prof. Paolo Bouquet for their continuous input and support regarding this topic, and for keeping me on track.

## References

- [1] Franz Baader, Peter Patel-Schneider, Diego Calvanese, Deborah L. McGuinness, and Daniele Nardi, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.
- [2] Massimo Benerecetti, Paolo Bouquet, and Chiara Ghidini. Contextual reasoning distilled. *J. Exp. Theor. Artif. Intell.*, 12(3):279–305, 2000.
- [3] Tim Berners-Lee, James A. Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, May, 2001. <http://www.sciam.com/2001/0501issue/0501berners-lee.html>.
- [4] Alexander Borgida and Luciano Serafini. Distributed description logics: Assimilating information from peer sources. In Stefano Spaccapietra, Salvatore T. March, and Karl Aberer, editors, *J. Data Semantics I*, volume 1 of *Lecture Notes in Computer Science*, pages 153–184. Springer, 2003.
- [5] Paolo Bouquet, Fausto Giunchiglia, Frank van Harmelen, Luciano Serafini, and Heiner Stuckenschmidt. C-owl: Contextualizing ontologies. In Dieter Fensel, Kattia P. Sycara, and John Mylopoulos, editors, *International Semantic Web Conference*, volume 2870 of *Lecture Notes in Computer Science*, pages 164–179. Springer, 2003.
- [6] Paolo Bouquet, Luciano Serafini, and Heiko Stoermer. Introducing Context into RDF Knowledge Bases. In *Proceedings of SWAP 2005, the 2nd Italian Semantic Web Workshop, Trento, Italy, December 14-16, 2005. CEUR Workshop Proceedings, ISSN 1613-0073, online <http://ceur-ws.org/Vol-166/70.pdf>*, 2005.
- [7] Jeremy Carroll, Christian Bizer, Patrick Hayes, and Patrick Stickler. Named Graphs, Provenance and Trust. In *Proceedings of the Fourteenth International World Wide Web Conference (WWW2005), Chiba, Japan*, volume 14, pages 613–622, May 2005.
- [8] G. Criscuolo, F. Giunchiglia, and L. Serafini. A Foundation for Metareasoning, Part I: The proof theory. *JLC*, 12(1):167–208, 2002.
- [9] G. Criscuolo, F. Giunchiglia, and L. Serafini. A Foundation for Metareasoning, Part II: The model theory. *JLC*, 12(3):345–370, 2002.
- [10] Eric Miller Frank Manola. *RDF Primer*, February 2004. <http://www.w3.org/TR/rdf-primer/>.
- [11] Chiara Ghidini and Fausto Giunchiglia. Local models semantics, or contextual reasoning=locality+compatibility. *Artif. Intell.*, 127(2):221–259, 2001.
- [12] Ramanathan V. Guha, Rob McCool, and Richard Fikes. Contexts for the semantic web. In Sheila A. McIlraith, Dimitris Plexousakis, and Frank van Harmelen, editors, *International Semantic Web Conference*, volume 3298 of *Lecture Notes in Computer Science*, pages 32–46. Springer, 2004.
- [13] Graham Klyne. *Contexts for RDF Information Modelling*. Content Technologies Ltd, October 2000. <http://www.ninebynine.org/RDFNotes/RDFContexts.html>.
- [14] Graham Klyne. *Circumstance, provenance and partial knowledge - Limiting the scope of RDF assertions*, 2002. <http://www.ninebynine.org/RDFNotes/UsingContextsWithRDF.html>.
- [15] Heiko Stoermer, Ignazio Palmisano, Domenico Redavid, Luigi Iannone, Paolo Bouquet, and Giovanni Semeraro. RDF and Contexts: Use of SPARQL and Named Graphs to Achieve Contextualization. In *Proceedings of the First Jena User's Conference, Bristol, UK, 2006* (to appear).
- [16] R.W. Weyhrauch. Prolegomena to a Theory of Mechanized Formal Reasoning. *Artificial Intelligence*, 13(1):133–176, 1980.