Context on the Semantic Web: Why and How

Loris Bozzato[1], Martin Homola[2], and Luciano Serafini[1]

[1] Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy
[2] FMFI, Comenius University, Mlynská dolina, 84248 Bratislava, Slovakia

{bozzato,serafini}@fbk.eu, homola@fmph.uniba.sk

Abstract. It is becoming increasingly apparent that knowledge published via the Semantic Web (SW) and Linked Open Data (LOD) resources is typically valid w.r.t. some assumed context. The contextual information, however, is often left implicit and not explicitly indicated. What is more, the means offered by SW technologies to represent this type of knowledge and link it to the resource itself are rather limited. In this position paper we argue that more advanced means of treatment of context in the SW and LOD resources are needed. Contextual meta knowledge has to be explicitly represented and logically treated. We propose a set of properties that we think such a representation should have and finally we review the known existing approaches to contextual representation on the SW.

1 Introduction

An increasing number of ontologies and data sets are being published using the SW languages such as RDF and OWL. Especially under the more recent LOD initiative, large knowledge sources such as DBpedia and Freebase but also many others were conceived and populated. However, rarely these large portions of knowledge are absolutely valid. They are typically qualified with respect to some *context*, i.e., they are assumed to hold under certain circumstances – relative to certain time period, a geo-political or geo-cultural region, certain specific topics, etc. By context we will therefore mean the situation that limits the validity of information and by contextual information (or meta information) we will mean any kind of description of this situation.

On the other hand, there is a lack of a widely accepted standard mechanism to qualify knowledge with the context in which it is supposed to hold. Sometimes the contextual meta data is mixed directly with other data; more often this meta information is crafted in annotation properties like rdfs:comment or owl:AnnotationProperty which do not affect reasoning at all, or it is even left implicit in many cases.

Recognizing this problem, a number of extensions to SW languages have been proposed with the aim to handle context. Among other proposals [4,12,5,1,15,8], there are approaches such as aRDF [18], Metaview [17], two-dimensional description logics of context [10,9] and CKR [14,6]. All these approaches offer some solution for this problem, however, a recognized and widely accepted consensus has not yet been reached by far.

This paper appeared at the 4th International Workshop on Acquisition, Representation, and Reasoning with Contextualized Knowledge (ARCOE-12), collocated with ECAI-12, August 28 2012.

In this position paper, we argue that contextual representation of knowledge is one of the most needed ingredient that is still missing in the standardized SW technology family. We underline the relevance of this topic and propose to the community to try to come up with a set of standard and universally acknowledged features of such a representation for the SW. In order to foster the discussion about such features, we then present a number of properties that we believe a contextualized representation suitable for the SW should satisfy.

It is to be stressed that these properties mirror our ideas and intuitions. We leave it open for the community to discuss and evaluate the proposed properties, and to come up with modifications or with other proposals. Towards the end of the paper we give a brief overview of the existing approaches and conclude with discussion on future directions.

2 Properties of Contextual Representation for the Semantic Web

In the search for a suitable knowledge representation framework it is a relevant question what properties it should have in order to serve best its purpose, in this case the purpose of enabling context to be explicitly represented and reasoned about within the SW.

According to our opinion the following properties should be considered. It is not, however, our stance that this is the canonical set of properties to be applied in context-enabled SW knowledge representation from this moment. This is something the community needs to discuss and anticipate. The following list should therefore not be considered as final, but rather as a starting point in this discussion.

Property 1 (Encapsulation). Knowledge that shares the same context should be encapsulated, easily identified and accessed.

A good example for encapsulation can be drawn from the *context as a box* metaphor known from the previous works on context [3]. Under this metaphor a context is perceived as a "box" whose boundaries are given by a set of contextual attributes, i.e., all knowledge that fits into common boundaries should be grouped within the same box. In accordance with this paradigm, these "boxes", that is, collections of knowledge that shares the same contextual qualification, are often themselves referred to as *contexts*, as we also do in this paper. Encapsulation does not automatically imply that knowledge bases need to be physically split into subsets for each context. On the other hand, the part of knowledge bound to the same context has to be easily retrieved and manipulated as needed. In this sense, also, the context should serve as navigation axis when working with the knowledge base.

Property 2 (Explicit meta knowledge). Knowledge about contexts should be explicitly represented in a logical language.

The knowledge inside the context, that is, the knowledge represented in the first place, will be called *object* knowledge. We assume that this knowledge has a logical representation, as it normally is in the SW. The property implies that in addition the contextual information that qualifies the object knowledge should also be explicitly represented in a logical language. This information will be called *meta* knowledge.

There are multiple ways how to represent meta knowledge. In the context as a box approach [3] and also other works [11] it is formalized using a set of contextual dimensions together with their possible values.

For example, given some knowledge about the current US presidential election, the meta data that we want to possibly represent may be that this information is relative to the location USA, that it is relative to a certain time period, e.g., year 2012, and to a specific topic, e.g., "presidential election". This information should be explicitly stated in the representation language, for instance, in OWL[3] using properties location, time and topic and individuals such as USA, 2012 and presidential_election. Assuming that the individual c42 acts as the context identifier for our context, we may assert axioms such as $\mathsf{location}(\mathsf{c42}, \mathsf{USA})$, $\mathsf{time}(\mathsf{c42}, 2012)$, and $\mathsf{topic}(\mathsf{c42}, \mathsf{presidential\_election})$. There might be more knowledge that we want to formalize in relation with the meta data. For instance we may want to assert that USA is a country, and that every country has a capital. To do this we may add more axioms, e.g., $\mathsf{Country}(\mathsf{USA})$ and $\mathsf{Country} \sqsubseteq \exists \mathsf{hasCapital}.\mathsf{City}$. The important point in this example is not this particular formalization, but instead the fact that the meta knowledge is represented in some logical language on top of which reasoning can be done.

Property 3 (Separation of meta knowledge and object knowledge). Meta knowledge is clearly distinguished from object knowledge.

By this property we mean that one can immediately tell which statements belong to object knowledge and which belong to meta knowledge, i.e., it should be apparent that the properties location, time and topic as used above belong to meta knowledge, and that the statements formed using these properties represent meta information. Again, the meta knowledge does not have to be physically separated from the object knowledge, but at least the two vocabularies used for each one respectively should be disjoint. One good reason to stick to this rule is simply to avoid the user to confuse object and meta statements, since each type of statements has different purpose and influences the knowledge base in different ways.

Property 4 (Relations between contexts). If one context is related to another, there should be a way how to represent this within the framework.

Studying contexts in separation makes little sense, as if there was just one context, representation of meta data and contextual reasoning would not be needed in fact. Hence contextual representation must be able to deal with multiple contexts and with the implications that the knowledge present in a context has on the other contexts. Therefore the literature of contextual reasoning has dedicated significant attention to possible relations between contexts [3,2,11]. A distinctive attention was given to the context coverage relation, that enables to organize the contexts w.r.t. the specificity–generality axis [11,2]. Other possible relations are for instance that of neighboring context (in space) or consecutive contexts (in time). Clearly such relations have significant influence of how much and which part of knowledge should be carried over from one context to another during reasoning. Therefore they need to be considered also in the case of contextual knowledge representation for the SW.

[3] For convenience we use description logic-like syntax for OWL expressions.

For example, given the context of US presidential election 2012, and given another context, say, one of US politics of early 21st century, the second context is clearly broader then the first one – it contains the first context in some sense. This also means that the first context is narrower then the second one. Similarly, considering a third context, this time of US presidential administration 2013–2016, we can see not only that this context is also narrower then the second context, but also that it immediately succeeds the first context in time. We discuss later on how such relations should be taken into account in reasoning (see Property 7), which, however, may only be possible if the relations between contexts are explicitly represented.

Property 5 (Contextual reasoning). Reasoning should take into the account as well the contextual meta knowledge and relations between contexts.

This property states that both the contextual meta knowledge and the context structure should be taken into the account in reasoning. Continuing our example, we should be able to assert a constraint requiring that if the same individual is an instance of President in two consecutive US administration contexts, then it cannot be an instance of Candidate in the following presidential election context. Clearly, the reasoner needs to be able to access and process the meta knowledge and it must further consider also the relations between contexts in order to reason with such a constraint.

In our previous work [14] we have argued that while meta knowledge may (and in fact should) influence the object knowledge, it should not be the case vice versa. This was relevant in order to maintain low complexity of reasoning. This desideratum is a bit too strict, however, if satisfiable complexity can be maintained (see Property 9) of course also more complex meta-reasoning patterns can be exploited.

One example of such a complex modeling pattern is the case in which we would organize the context structure differently depending on the assertions in the object knowledge. Theoretically such use cases are appealing, however, their practical use and especially the implications on the computational costs have to be thoroughly investigated.

Property 6 (Locality of knowledge). In each context we should be able to state axioms with local effect, that do not affect other contexts, and are not affected by other contexts.

For instance, in any context related to US presidential administration there will be a concept President but in each of these contexts this concept may have different instances. Furthermore, in all of these context we may want to add an axiom which would guarantee that the concept has only one instance (i.e., there is always only one president). On the other hand, in the broader context related to early 21st century US politics, there is also a concept President, in this case however we do not want to have such an axiom – there may be multiple presidents in this broader period.

Property 7 (Knowledge lifting). If needed, specific knowledge can be lifted from one context and reused by another.

Locality of knowledge, on the other hand, should not imply opacity. If needed, knowledge from one context should be accessible to another context, especially if the contexts are related in a favourable fashion. Such propagation of knowledge between contexts is commonly referred to as *knowledge lifting*, and it has been intensively studied in the literature [13,11,2]. It is usually implemented by specific lifting axioms, or possibly by some automated lifting mechanism.

In our running example we might want to assure that the president in the administration of 2013–2016 is featured as an instance of the concept President in the context of early 21st century US politics. This is an example of knowledge lifting and it could be asserted using a specific lifting axiom that would lift President from the administration 2013–2016 into the latter context as President2013-2016 and then asserting a local axiom $\mathsf{President2013\text{-}2016} \sqsubseteq \mathsf{President}$.

Consider also the case of two consecutive contexts, e.g., the one of US presidential election 2012 and the one of the US presidential administration of 2013–2016. With specific lifting axioms, we should be able to assert that the individual which represents the elected president in the first context is the same as the one which represents the actual president in the latter context. The amount of knowledge reuse may possibly be constrained by relations between contexts – for example, if the contexts are unrelated the lifting may not be possible, or may at least be very limited.

Note that Properties 6 and 7 are not mutually conflicting. Both cases should be possible at the same time: one approach is to have knowledge with local meaning, if no lifting axioms were specified. Knowledge lifting can then be implemented by giving the user the option to specify lifting axioms in specific cases as needed.

Property 8 (Overlapping and varying domains). Objects can be present in multiple contexts, but not necessarily in all contexts.

The RDF semantics relies on the fact that the same URI should always have the same meaning in every RDF document. Accordingly, we believe constants should have the same meaning, at least in related contexts, and therefore the interpretation domains should partly overlap. We refrain from this requirement when contexts are completely unrelated. Also, we need not require to the existence of the interpretation for all constants of the language within a given context, as it is usual in DL-like OWL semantics. Therefore constants should naturally appear (and be interpreted) in those contexts for which they are relevant.

In this respect, in our example the constant obama which is present in the context related to US presidential administration of 2009–2012 need not necessarily appear in the context of the next administration. On the other hand obama should certainly appear in the early 21st century US politics context, as given the relation of the latter context to the one of the 2009–2012 administration, it is clearly a relevant constant there.

Property 9 (Complexity invariance). The contextual layer should not increase the complexity of reasoning.

The last property is concerned with the computational cost that we have to pay to add the contextual dimension to the SW. Given the fact that some of the SW languages already exhibit quite high complexity (i.e., OWL 2, based on the $\mathcal{SROIQ}$ DL, is 2NExpTime-complete [7]), we believe that the contextual layer needs to be added without any increase in the complexity – if only possible. Indeed, a minor increase in complexity, especially with more tractable OWL fragments, may not be as harmful.

Therefore we could more generally require that the complexity of the resulting contextualized formalism should be acceptable.

3 Existing Approaches and Proposals

Perceiving the need for some mean of representing context in the SW, both aRDF [18] and Context Description Framework [8] extend RDF triples with an n-tuple of qualification attributes with partially ordered domains. In such an approach, each formula is annotated separately, seemingly violating the encapsulation property. However, this can also be seen as an optimization issue easy to resolve during the implementation.

Straccia et al. [16] enable RDFS graphs to be annotated with values from a lattice. The semantics of the framework is based on an interpretation structure that is common in multi-valued logics. This effectively restricts the dimensional structure to a complete lattice: for every two contexts there exists a meet ($\wedge$) and a join ($\vee$) and also the global bottom ($\bot$) and top ($\top$). This permits to study relations between contexts, e.g., looking for least common super context and greatest common sub context in some sense.

Another extension of RDFS to cope with context was proposed by Guha et al. [5] and further developed in Bao et al. [1]. A new predicate $\mathsf{isin}(c, \phi)$ is used to assert that the triple ϕ occurs in the context c. A set of operators to combine contexts ($c_1 \wedge c_1$, $c_1 \vee c_2$, $\neg c$) and to relate contexts ($c \Rightarrow c_2$, $c \rightarrow c_2$) is defined, making the approach particularly suited for manipulating contexts. Unfortunately, to our best knowledge no decision procedure is known so far.

The Metaview approach [17] enriches OWL ontologies with logically treated annotations and it can be used to model contextual meta data, albeit on per-axiom basis. In the Metaview framework, however, the contextual level has no direct implications on ontology reasoning, but it makes possible to reason about the ontology or even data. Also a contextually sensitive query language MQL is provided. Presented examples concentrate especially on modeling provenance of data and associated confidence, and the framework seems well suited for this purpose.

Two-dimensional description logics of context [10] allow for a multi-modal extension of one description logic by another. In this fashion, a combination of some object language $\mathcal{L}_O$ and another context language $\mathcal{L}_C$ results into the language $\mathfrak{C}_{\mathcal{L}_O}^{\mathcal{L}_C}$. This permits a structured knowledge base composed of a number of contexts in $\mathcal{L}_O$ with contextual meta data expressed in $\mathcal{L}_C$. For instance $\mathfrak{C}_{\mathcal{ALC}}^{\mathcal{ALC}}$ (previously called $\mathcal{ALC}_{\mathcal{ALC}}$) was studied [9]. An interesting feature of this logic are the contextual modal operators such as $[C]_r A$ representing "all objects of type A in all contexts of type C reachable from the current context via relation r". The language also provides rich representation possibilities for meta knowledge using a separate knowledge base in $\mathcal{L}_C$.

Finally, Contextualized Knowledge Repository (CKR) [14,6] allows for contextualized knowledge bases structured into contexts. Local language of the contexts can be as expressive as $\mathcal{SROIQ}$ (i.e., OWL 2) or any of its sublanguages. Contextual meta data is assigned to contexts in form of dimensional attributes, which are formalized in a separate meta knowledge base. A set of attributes assigned to some context is called a dimensional vector. To access information across contexts so called qualified symbols are used: the concept $A_{\mathbf{c}}$ represents the meaning of concept A in the context identified by the dimensional vector **c**. The choice to implement knowledge lifting with qualified symbols is one of the main differences between CKR and the previous approach.

4 Discussion

Coming back to the existing approaches that we discussed above, we first have to say that CKR was indeed modeled based on similar desiderata [14] as proposed in this paper, being proposed by the same authors. Remarkably, also the two-dimensional description logics of context seem to be in accord with most of the properties. Both these frameworks provide good encapsulation of contexts, rich representation of meta knowledge, and contextual reasoning. We stop the comparisons here; a more complex evaluation of all of the formalisms is beyond the scope of this position paper.

Considering the increasing importance of contextual knowledge representation for the SW and LOD resources, we have proposed in this paper a set of properties that we believe such a representation should satisfy. Our intention here is not to propose the canonical set of properties that should be immediately adopted: instead, we intend this just as the first step in a longer process, in which the community should review these properties, try to reach consensus in discussion, and work towards a standardization. Only then a well suited solution can be found, which can be adopted by a number of SW and LOD data sources to the benefit of the users.

Acknowledgements. This research was supported from the LiveMemories project. Martin Homola is also supported from the Slovak national project VEGA no. 1/1333/12.

References

1. Bao, J., Tao, J., McGuinness, D.: Context representation for the semantic web. In: Procs. of WebSci10 (2010)
2. Benerecetti, M., Bouquet, P., Ghidini, C.: Contextual Reasoning Distilled. Experimental and Theoretical AI 12(3), 279–305 (2000)
3. Benerecetti, M., Bouquet, P., Ghidini, C.: On the Dimensions of Context Dependence. In: Perspectives on Contexts. CSLI (2007)
4. Ding, L., Finin, T., Peng, Y., da Silva, P.P., McGuinness, D.: Tracking RDF graph provenance using RDF molecules. In: ISWC (2005)
5. Guha, R., McCool, R., Fikes, R.: Contexts for the semantic web. In: ISWC (2004)
6. Homola, M., Serafini, L.: Contextual representation and reasoning with description logics. In: DL (2011)
7. Kazakov, Y.: $\mathcal{RIQ}$ and $\mathcal{SROIQ}$ are harder than $\mathcal{SHOIQ}$. In: KR (2008)
8. Khriyenko, O., Terziyan, V.: A framework for context sensitive metadata description. IJSMO 1(2), 154–164 (2006)
9. Klarman, S., Gutiérrez-Basulto, V.: $\mathcal{ALC}_{\mathcal{ALC}}$: a context description logic. In: JELIA (2010)
10. Klarman, S., Gutiérrez-Basulto, V.: Two-dimensional description logics of context. In: DL (2011)
11. Lenat, D.: The Dimensions of Context Space. Tech. rep., CYCorp (1998)
12. Liao, H.C., Tu, C.C.: A RDF and OWL-based temporal context reasoning model for smart home. Inform. Tech. J. 6, 1130–1138 (2007)

13. McCarthy, J.: Notes on formalizing context. In: IJCAI (1993)
14. Serafini, L., Homola, M.: Contextualized knowledge repositories for the semantic web. Journal of Web Semantics, Special Issue: Reasoning with context in the Semantic Web 12 (2012)
15. Stoermer, H.: Introducing context into semantic web knowledge bases. In: CAiSE DC (2006)
16. Straccia, U., Lopes, N., Lukácsy, G., Polleres, A.: A general framework for representing and reasoning with annotated semantic web data. In: AAAI (2010)
17. Tran, T., Haase, P., Motik, B., Cuenca Grau, B., Horrocks, I.: Metalevel information in ontology-based applications. In: AAAI (2008)
18. Udrea, O., Recupero, D., Subrahmanian, V.S.: Annotated RDF. ACM Trans. Comput. Log. 11(2), 1–41 (2010)