Facets

Giovanni Bartolomeo[1],

[1] University of Rome Tor Vergata, Via del Politecnico 1,
00100 Rome, Italy
Giovanni.Bartolomeo@Uniroma2.it

Abstract. We present a novel approach to the resource identity problem in RDF documents based on the notion of context-dependant facets.

Keywords: RDF, Named Graph, eXtensible Resource Identifier (XRI).

1 Introduction

RDF was introduced to provide machine-understandable descriptions of "resources". The term "resource" was originally intended as a synonym of a generic "object of the Network" (RFC 1630). In recent times however the notion of "resource" has been widened to include "anything that has an identity", even abstract concepts (RFC 3986). The reuse of RDF on such a much broader domain has caused some critical challenges to emerge.

One of these challenges is related to the quality of information in Linked Data. It is well known that Linked Data allows anyone to provide statements on any resource, even on those resources which are not under one's control. Despite it is still possible to distinguish between authoritative and non authoritative descriptions, a single trusted predicate (owl:sameAs) could be enough to let infer statements that - expressing personal opinions, outdated facts or simply misplaced attributes - might lead to inconsistencies or might have other negative side effects (such as unwanted disclosures of personal information and privacy violations).

In this paper we first provide a brief overview of works on contexts, ontology mapping and resource identity in RDF, OWL and Linked Data. Subsequently, we introduce our intuition which reconsiders Hayes and Halpin's thoughts on "reference by description" [6] and leverages on the difference between referring to a resources in a context and referring to one of its "facets" from outside that context. Finally we propose a possible grounding of our idea in RDF using Named Graphs [1].

2 Related Works

Giunchiglia et al. [2] have attempted to add the notion of context to OWL, observing that ontology stands to context as global stands to local: ontologies are useful to share information whereas contexts are local models that encode a party's view of a domain not (necessarily) to be shared with others. An ontology is contextualized when its contents are kept local and put in relation with the contents of another ontology via explicit mappings called ”bridge rules”.

In [3] the authors state that this approach aims at addressing the context-insensitive issue of OWL, however it does not provide a solution to knowledge reuse in RDF. In fact, RDF intentionally does not specify how much knowledge from a referenced document should be reused in the referring document. To selectively transfer knowledge between documents the authors introduce a set of new RDF predicates.

Hayes and Halpin address the problem of resource identity in Linked Data, presenting four “alternative readings” of the predicate `owl:sameAs` [4]: misplaced references, referential “opacity”, identity in different contexts, and similarity. They admit that choosing alternative predicates to `owl:sameAs` might be difficult, as their interpretation could be quite subjective.

Many approaches to contextualization have focused on maintaining tracks of provenance information, rather than providing mappings between resources in different contexts. An extensive investigation on this topic has been performed by the Provenance Incubator Working Group at W3C, which has derived explicit requirements on the usage and management of provenance information [5].

3 Contexts and Facets

Again Hayes and Halpin [6], trying to understand the conceptual problem of defining identity for non accessible entities, propose an unusual viewpoint. They observe that reference “by description” is the only possible alternative to reference “by acquaintance” when an entity (e.g., a non information resource) is not accessible. Terms in a sentence determine the meaning of the sentence; however it is the sentence which contributes to determine (at least part of) the meaning of its terms. The meaning of a term is determined by the normative usage of all the sentences about that term.

Let us consider a common Linked Data scenario: two authors might use different URIrefs in different namespaces to talk about, e.g. a given car from a technical perspective and a commercial perspective [2]; independently, they would assert different statements for each of them, without mixing their properties. However the URIrefs they use would still refer to the same car seen from different viewpoints. Thus, a third author could legally relate each other by applying an `owl:sameAs` predicate, turning the two distinct URIrefs into aliases and destroying their contextualization.

We think that the mistake in this approach is to use the same identifier to refer both to a resource and to its “facets” [6]. If one wanted to refer to Napoleon as depicted in different contexts, e.g. in history, essays, fictions, or poetry, one would probably say “*that* Napoleon during the Waterloo battle”, or “*that* Napoleon in Alessandro Manzoni’s poem *Il Cinque Maggio*”. *Inside* each of these sentences the subject is Napoleon; but *outside* them the subject is a *facet* of Napoleon.

According to our intuition, by default an identifier can be used to refer to a resource, *in* a given context; but, *outside* that context, one should emphasize that this identifier is the "context-dependent" term to refer to that resource, i.e., one should refer to a *local facet* of the resource, and not to the resource itself.

4 Facets in Named Graphs

RDF represents statements about resources as a graph of nodes and arcs ("links"). We think that when a graph is bound to a context, its nodes and links, seen from *outside* that context, should be better interpreted as local facets of resources and *facts about these facets* - rather than as *claims* about the original resources. Therefore one would need the capability to refer to nodes and links *as parts of* a graph.

Unfortunately, RDF does not provide such facility. This is not possible even using the reification dictionary, as there is no way in the "conventional usage of reification" to associate the subject of the reification triples composing the reification quad to an individual triple in the original document.

Named Graphs have emerged among several different proposals as a solution characterized by its simplicity, backward compatibility and straightforward applicability to existing software. A Named Graph is not a RDF graph, rather it is a pair *(n,g)* where *n* is the "name" (identifier) of the Named Graph and *g* is a RDF graph. A Named Graph has to be intended as a given "copy" of a RDF graph, to which an identifier has been rigidly bound.

Even Named Graphs currently do not provide the ability to assign a "name" to (better, to *refer to*) their internal nodes, but we think that this facility could be seamlessly introduced. Assuming that prefixes `ex:` and `ex2:` refer to, respectively, namespaces `http://example.org/` and `http://example2.org/`, let us consider the following Named Graph:

`ex:myGraph{ex2:John.Smith foaf:knows ex:Alice.Doe .}` **(1)**

Each URIref appearing in (1) can be replaced by a "cross reference", a notation originally developed for the eXtensible Resource Identifier (XRI) [7]. Syntactically, a cross reference simply consists in the original URIref enclosed into round brackets. For example, `ex2:John.Smith` is cross referred by `(http://example2.org/John.Smith)`. Obviously a cross reference is not a URIref; however it can be resolved into a URIref [8]. To this end, we use the name of the graph as the "resolution context" [7] to which the cross reference is bound. Therefore, node `ex2:John.Smith` in Named Graph (1) is identified by the URIref

```
http://example.org/res/myGraph/(http%253A%252F%252Fe
        xample2.org%252Fres%2523John.Smith)
```
(2)

where a trailing slash has been added to the graph's identifier to mark the transition between the URIref identifying the graph and the cross references; the percent encoding rules recommended in [7] have been applied. This URIref refers to a *facet* of `ex2:John.Smith` in the context of Named Graph `ex:myGraph`. In short, we are suggesting that if URIref *u* identifying resource *r* is in Named Graph *g*, then a graph-outsider should use *g/(u)* to refer to that facet of *r* described in *g*.

Generally, an entity authoritative for a Named Graph is not authoritative for the URIrefs contained inside that Named Graph (in our example, prefixes `ex2:` and `foaf:` denote different authorities than `ex:`); however, one could note that an entity authoritative for a Named Graph is also authoritative for all the proposed URIrefs identifying facets of the resources referenced in the Named Graph. According to the technique described in [9], a HTTP GET request on these URIrefs could return a 303 HTTP response redirecting to the corresponding informative resource containing authoritative statements about *these* facets, i.e. the Named Graph in which they appear. In our example, an HTTP GET request on (2) would return (1).

5 Conclusions

In this paper we presented a possible technique to refer to different context-dependent "facets" of a resource. Facets are represented by nodes and links of Named Graph that are assigned global identifiers distinguished from URIrefs used to refer to resources. We have not (yet) provided a formal semantics for facets. At the time of writing, we are evaluating several possibilities, some of them based on the semantics defined for conventional reification.

References

1. Bizer, C., Carroll, J.J.: Named Graphs, Provenance and Trust. In: WWW 2005, May 10-14,2005, Chiba, Japan. ACM 1595930469/05/0005 (2005)
2. Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L, Stuckenschmidt, H.: C-OWL: Contextualizing ontologies. In D. Fensel, K. P. Sycara, and J. Mylopoulos, editors, International Semantic Web Conference, volume 2870 of Lecture Notes in Computer Science, pages 164–179. Springer, 2003
3. Ding, L., Shinavier, J., Finin T., McGuinness, D.: owl:sameAs and Linked Data: An Empirical Study. Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, 2010.
4. Halpin, H., Hayes, P.P., McCusker, J., Mcguinness, D., Thompson, H.: When owl:sameas isn't the same: An analysis of identity in linked data. In Proceedings of the 9th International SemanticWeb Conference, 2010. DOI: 10.1007/978-3-642-17746-0_20 23
5. Zhao, J., Bizer, C., Gil, Y., Missier, P., Sahoo, S.:Provenance Requirements for the Next Version of RDF, W3C Workshop - RDF Next Steps, June 26-27, 2010, Stanford, Palo Alto, CA, USA February (2010)
6. Hayes, P., Halpin, H., In defense of ambiguity. International Journal of Semantic Web and Information Systems, 4(3), 2008.
7. Reed, D. (Ed): Extensible Resource Identifier (XRI) Version 3.0 Working Draft 03 (2010) http://docs.oasis-open.org/xri/xri/v3.0/WD01/xri-syntax-3.0-wd03.pdf
8. Thompson, H.S., Orchard D. (Eds): URNs, Namespaces and Registries. (2006) http://www.w3.org/2001/tag/doc/URNsAndRegistries-50.html
9. Sauermann, L., Cyganiak, R., Volkel, M.: Cool URIs for the Semantic Web, Technical Report, TM-07-01, DFKI, 2007