Named Graphs: an Approach to Trust and Provenance

Chris Bizer[1], Jeremy J. Carroll[2,⋆], Pat Hayes[3], and Patrick Stickler[4]

[1] Freie Universität Berlin, Germany
[2] Hewlett-Packard Labs,Bristol, UK
[3] IHMC, Florida, USA
[4] Nokia Research Centre, Tampere, Finland

Abstract. Placeholder. **Todo:** *Decide on order of authors, the current order is alphabetically. I suggest that the final order should approximate to the amount of work each has put in, to be decided when the paper is nearly finished, perhaps decided by the first person who volunteers to go last*

1 Introduction

Todo:

2 Abstract Syntax

The abstract syntax of named graphs is based on that of RDF[3]. A set of named graphs is a partial function relating nodes (URIrefs and blank) to RDF graphs.

In more detail a set of named graphs $\mathbf{N}$ is a 5-tuple $\langle N, V, U, B, L\rangle$ where: U is a set of URIrefs; L is a set of literals (both plain and typed); B is a set of 'blank' nodes; $V = U \cup B \cup L$ is the set of *nodes* of $\mathbf{N}$; N is a partial function from $U \cup B$ to $V \times U \times V$. U, B and L are pairwise disjoint. $N(n)$ is hence an RDF graph[5] (a set of triples) which is *named* n. When $n \neq n'$ the blank nodes used in triples from $N(n)$ are all distinct from those used in triples from $N(n')$, i.e. blank nodes cannot be shared between different graphs named in N. For technical reasons, we require all nodes $n \in U \cup B \cup L$ to either be a name in the domain of N or to appear in a triple in some graph in the range of N.

Two sets of named graphs $\mathbf{N} = \langle N, V, U, B, L\rangle$ and $\mathbf{N}' = \langle N', V', U, B', L\rangle$ are isomorphic if there is a bijection $\phi : V \to V'$ such that ϕ is the identity on $U \cup L$ and:

$$\langle s, p, o\rangle \in N(n) \text{if and only if } \langle \phi(s), \phi(p), \phi(o)\rangle \in N'(\phi(n)) \qquad (1)$$

In this case we see that the graphs named by $\mathbf{N}$ and the graphs named by $\mathbf{N}'$ are pairwise equivalent (in the sense of [3]).

⋆ Jeremy Carroll is a visiting researcher at ISTI, CNR in Pisa, and thanks his host Oreste Signore.
[5] We have removed the legacy constraint that a literal cannot be the subject of a triple.

3 Concrete Syntax

We offer three concrete syntaxes for named graphs: RDF/XML[1] on the Web; TriX[2]; and a new informal syntax used in this paper.

The URL from which an RDF/XML file is retrieved can act as a name for the graph given by the RDF/XML file using the normal rules. This has some disadvantages:

– It is not clear where the boundary of a set of named graphs lies, the URL provides the name for a single graph, whereas the advantageouos of named graphs is the ability to consider a collection of graphs.
– It is not possible to use a blank node as the name of a graph, or a URIref which is not a URL.
– The known constraints and limitations of RDF/XML apply. For instance, it is not possible to serialize graphs which have predicates that do not end with a sequence matching the NCName production from XML Namespaces. Nor is it possible to use literals as subjects.
– It confuses the URL as a means of identifying the document, and the URL as a means of identifying the graph described by the document.

In balance, there is the major advantage of a deployed base, and current technology.

The TriX serialization of Carroll and Stickler is given by the following DTD

```
ToDo
```

In this paper we use an informal notation, TriG, derived from the informal notation used in the RDF and OWL recommendations. It is roughly N-triple[**?**] with qnames. We extend that notation by using '(' and ')' to group triples into multiple graphs, and to (optionally) precede each by the name of that graph. When the name is omitted it is a b-node that does not occur elsewhere.

4 Semantics

Todo: *Todo. Needs to cover how the graph naming is captured in the semantics either extensionally say with a direct restriction on I, or intensionally with say an additional graph extension interpretation function. Also should cover semantics for `rdfx:equivalentGraph` and `rdfx:subGraphOf`, and maybe `rdfx:Triple` and `rdfx:graphSize`. Beginnings of suggestion ...*

The meaning of a set of named graphs depends on a separate decision about which of the graphs to accept. We represent this decision as a set A of nodes naming the accepted graphs. The meaning of a set of accepted named graphs $\langle A, \mathbf{N} \rangle$ is given by taking the graph merge $\bigcup_{a \in A} N(a)$, and then interpreting that graph using the semantics of RDF[**?**]. Any extension semantics of RDF can be used; in this paper we uniformly use those of OWL Full[4].

5 A Simple Query Language

Todo.

6 Trust

Todo.

6.1 Using a Public Key Infrastructure

Todo:

6.2 Publishing and Asserting RDF

Todo:

7 Provenance

Todo:

8 Apparent Paradoxes

Todo: *We point out that rdfs:comment can be used to introduce Russell's paradox in RDFS, and that the problem is an abuse of rdfs:comment as a class or property defn. We then concern log:implies and log:True and log:False and show that absolute semantic definitions are broken, whereas ones that talk operationally about CWM are OK (if unilluminating).*

9 Summary of New Vocabulary

We have introduced new vocabulary for named graphs using the `rdfx:` namespace. The classes are listed in table **??**, the properties in table **??**. **Todo:** *Write the tables*

10 Conclusions

Todo.

Todo: *Update bib file*

References

1. D. Beckett. RDF/XML Syntax Specification (Revised). `http://www.w3.org/TR/rdf-syntax-grammar/`, 2003.
2. J. Carroll and P.Stickler. RDF Triples in XML. Submitted to WWW2004, 2003.
3. G. Klyne and J. Carroll. Resource Description Framework (RDF): Concepts and Abstract Syntax. `http://www.w3.org/TR/rdf-concepts/`, 2003.
4. P. F. Patel-Schneider, P. Hayes, and I. Horrocks. OWL Web Ontology Language Semantics and Abstract Syntax. `http://www.w3.org/TR/owl-semantics/`, 2003.