A Protégé Plug-in Development to Support the NEPOMUK Representational Language

Milena C. Caires[1], Simon Scerri[1], Siegfried Handschuh[1], Michael Sintek[2], Ludger van Elst[2]
[1] DERI, National University of Ireland, Galway, Ireland
{firstname.surname}@deri.org
[2] German Research Center for Artificial Intelligence (DFKI) GmbH, Kaiserslautern, Germany
{surname}@dfki.uni-kl.de

Abstract—**In this document, we present the challenges faced to develop a plug-in for Protégé that supports requirements for the NEPOMUK[1] Representational Language (NRL). NRL is a language built on top of RDF/S and although it was designed to fulfill requirements for the NEPOMUK Social Semantic Desktop project, it is available for global use due to its novel improvements over the RDF model and given its independence of the domain for which it was initially designed.**

***Index Terms*—NRL, Protégé and plug-in**

I. Introduction

NRL [1] combines a set of useful and accepted RDF/S extensions within a comprehensive and sound conceptual framework and it was inspired by the need for a robust representational language for the Social Semantic Desktop [3]. It targets the shortcomings of RDF/S, the most important one being the lack of support for handling multiple models. NRL introduces Named Graphs to the enhanced RDF model. This means that each RDF dataset can have a unique identifier as its name, and metadata information associated with it. Thus NRL tackles the issues of handling data provenance information and trust [2]. Besides provenance information other metadata can be associated with named graphs, like *Graph Roles*, e.g. Ontology or Instance Base. Additionally, NRL has support for property constraints inspired by OWL and Protégé (e.g., inverse property and cardinality constraints among others).

To illustrate the issues tackled by NRL that cannot be handled with plain RDF/S, we will take a use case based on desktop users. Desktop users may be interested in different aspects of data at different times. Looking at a document in a folder, the user might want to learn about all other documents that are semantically related. However, the user might not want to see all available information but just one particular aspect of it. The Decker problem[2] is of particular relevance here - given that a resource is defined as the subclass of another resource, then it can be stated that the parent resource has at least one child. In some scenarios, the user might need to know nothing about the children. In others, the actual number of children is of interest and yet for others, extensive information pertaining to the children is also of relevance for the user. Returning all this aggregated information at once would confuse rather than help the user.

NRL's solution to this problem is its ability to handle multiple interpretations for a dataset at different times. Bearing in mind that the information is stored in named graphs, this required interpretation is accomplished by extracting a simplified form of the Graph. NRL refers to these interpretations as *Graph Views*, and they are one of the core concepts in the representation language. In other situations, more complex views over graphs might be required. These views are obtained by realizing the *Declarative Semantics* carried by a named graph. For example, a user browsing through a hierarchy of concepts might want to see not just the direct parent (defined via rdfs:subClassOf) of an interesting concept, but all parent concepts. This would require a Graph view that realizes the semantics of rdfs:subClassOf by performing its transitive closure. By allowing these arbitrary tailored interpretations for established named graphs, graph views respect the idea that basic named graphs should not innately carry realized semantics or assumptions on semantics.

On the contrary, Protégé's (excluding Protégé-OWL) meta-model operates under a closed-world assumption, i.e. it imposes assumptions on semantics for all processed data. This is unaligned with the RDF/S semantics [4] which state that RDF/S is open-world. It would also be incompatible with the multiple semantics approach of NRL. NRL allows data to carry arbitrary declarative semantics, which includes the possibility of having named graphs with closed-world (NRL-CW) or open-world (NRL-OW) assumptions. On the other hand, RDF/S and OWL by definition always have an open-world assumption. Therefore we want to ensure that a NRL graph that is loaded in Protégé should be handled according to its declarative semantics. This implies that whereas a loaded graph that is marked as being a NRL-OW requires no extra action, a NRL-CW graph should be validated accordingly.

II. Handling multiple models: named graphs

In the Social Semantic Desktop domain we take a Named Graphs [2] approach to semantic data. Named Graphs (NGs) are an extension on the top of RDF, where a unique identifier

[1] Networked Environment for Personalized, Ontology-based Management of Unified Knowledge

[2] http://lists.w3.org/Archives/Public/public-sws-ig/2004Feb/0037.html

is assigned to each RDF graph. Most important, NGs allow the metadata information to be associated to each uniquely identified graph, making data handling more manageable. Named graphs should not carry any realized semantics or assumptions on the semantics. Instead these can be realized through designated views on graphs.

An RDF triple can exist in a named graph or outside any named graph. However, for consistency reasons, all triples must be assigned to some named graph. For this reason, NRL provides a special named graph (nrl:DefaultGraph). Triples existing outside any named graph automatically form part of this default graph, to ensure backward compatibility with triples that are not based on named graphs. A Graph set is composed of a default graph and an unlimited number of distinct named graphs.

NRL distinguishes between Graphs and Graph Roles, in order to have orthogonal modeling primitives for defining graphs and for specifying their role. A graph role refers to the characteristics and content of a named graph (e.g. an instance base, ontology, a knowledge base, etc) and how the data is intended to be handled. The NEPOMUK Graph Metadata vocabulary (NGM) provides the vocabulary for annotating graph roles, i.e., providing metadata about graphs. This is attached to graph roles rather than to the graph themselves, because it is more intuitive to annotate an ontology, for example, than the underlying graph. Roles are more stable than the graphs they represent, and while the graph for a particular role might change constantly, evolution of the role itself is less frequent. An instantiation of a role represents a specific type of graph and the corresponding triple set data.

III. Imposing semantics on graphs: Graph Views

A named graph consists only of the enumerated triples in the triple set associated with the name, and does not inherently carry any form of semantics (apart from the basic RDF semantics). However, in many situations it is required to work with an extended or restricted interpretation of simple syntax-only named graphs. These can be realized by applying some algorithm which enhances named graphs with entailment triples, returns a restricted form of the triple set, or an entirely new triple set. To preserve the integrity of a named graph, interpretations of one named graph should never replace the original. To model this functionality and retain the separation between original named graph and any number of their interpretations, NRL introduced the concept of Graph Views. Formally, a view is an executable specification of an input graph into a corresponding output graph. Informally, they can be seen as arbitrary wrappings for a named graph.

IV. Challenges

The NEPOMUK project requires an ontology modeling tool to support NRL. In other words, the modeling tool must handle the following issues:

- **Named Graphs:** the idea behind named graphs consists in having multiple RDF graphs in a single document/repository (graph set) and naming them with URIs. Currently, there is no support for handling named graphs in Protégé;
- **Metadata Graph:** each Named Graph must have an associated special graph, called the Metadata graph, that contains its metadata information, i.e. Graph Role, author, date of creation, etc;
- **Multiple Views:** each Named Graph may have several views and each Graph View is a graph corresponding to an extended or simplified version of the original named graph. View specification provide rules to extract views from the original graph;
- **Multiple Semantics:** In Protégé a closed-world assumption is imposed on all managed data. We would like to extend Protégé so that no semantics are ever imposed on loaded NRL data. Instead, we want the data concerned to be treated solely with regards to its declarative semantics. Given a graph with declared NRL-CW Semantics, this would require validation on the graph to be performed. Such a graph does not necessarily have to be self-contained, i.e., a NRL-CW graph can still be valid if it uses properties from an *external graph*, given that the domain, range and other given constraints are satisfied. Although an external graph must be accessible in order to perform consistency checking, there should be no need to wholly import it within the current project, as is currently required by Protégé.

V. Protégé Plug-in development

The modeling tool chosen by the NEPOMUK project was Protégé because it is: open-source, a well-known tool by ontology developers, easy to extend given the framework based on plug-in; there is good documentation available; well maintained and up-to-date with recent W3C recommendations; and the mailing lists are very helpful. However, Protégé does not have many features required to deal with the NRL challenges described.

We decided to develop an export plug-in for Protégé as a starting point in order to develop later a plug-in that is able to provide Protégé with full support for NRL data. The NRL export option in the plug-in enables the provision of information necessary to create the metadata graph for the named graph that corresponds to the model extracted from the current project. The current export plug-in only applies to RDF/S projects and it is not yet possible to convert OWL projects to NRL. OWL semantics are relatively harder to handle and thus a conversion from OWL to NRL is far from straightforward. We plan to address this problem as future work. The exported data is serialized as TriG[3], a language that supports named graphs. This data typically consists of the named graph itself, plus its associated metadata graph. However the user can choose to output the data serialized as XML/RDFS. In this case, the graph and its metadata graph are stored within two separate files. Fig. 1. shows an example of such an output, where a graph, *mine:gn* and its associated metadata graph, *mine:gn_metadata* are defined. In particular the latter states that it is a metadata graph, that *mine:gn* is a named graph with a role of KnowledgeBase and carries both RDF/S and NRL-OW declarative semantics.

```
@prefix nrl:
<http://www.semanticdesktop.org/ontologies/2006/1
1/24/nrl#> .

@prefix nao:
<http://www.semanticdesktop.org/ontologies/2007/0
3/nao#> .

@prefix mine:
<http://www.mynamespace.org/my#> .

@prefix rdfs:
<http://www.w3.org/2000/01/rdf-schema#> .

@prefix rdf:
<http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

mine:gn {mine:Van
          a       rdfs:Class ;
          rdfs:label "Van" ;
          rdfs:subClassOf mine:MotorVehicle .

mine:Person
          a       rdfs:Class ;
          rdfs:label "Person" ;
          rdfs:subClassOf rdfs:Resource .

mine:rearSeatLegRoom
          a       rdf:Property ;
          rdfs:domain mine:MotorVehicle ;
          rdfs:label "rearSeatLegRoom" ;
          rdfs:range rdfs:Literal .


mine:MotorVehicle
          a       rdfs:Class ;
          rdfs:label "MotorVehicle" ;
          rdfs:su bClassOf rdfs:Resource .


mine:PassengerVehicle
          a       rdfs:Class ;
          rdfs:label "PassengerVehicle" ;
          rdfs:subClassOf mine:MotorVehicle .

mine:MiniVan
          a       rdfs:Class ;
          rdfs:label "MiniVan" ;
           rdfs:subClassOf mine:Van ,
           mine:PassengerVehicle .
}

mine:gn_metadata {mine:gn_metadata
          a       nrl:GraphMetadata .
           mine:gn
           a       nrl:KnowledgeBase ,
           nrl:DocumentGraph ;
          nrl:hasSemantics mine:NRLOW ,
                   mine:RDFS ;
          nao:contributor
          <http://sw.deri.ie/people/Milena> ;

           nao:creator
           <http://sw.deri.ie/people/UserName> ;
           nao:engineeringTool "Protege" ;
           nao:hasNamespace
           "http://www.mynamespace.org/my#" ;
          nao:hasNamespaceAbbreviation
                  "mine" ;
          nao:lastModified
               "2007-04-27T19:14:56.941Z" ;
          nao:serializationLanguage
                  "TriG" ;
          nao:status "Stable" ;
          nao:updatable "1" ;
          nao:version "1" .

    nrl:NRLOW
          a       nrl:Semantics ;
           nrl:semanticsDefinedBy
           "http://www.semanticdesktop.org/ontolo
           gies/nrl#" ;
          nrl:semanticsLabel "NRL/OW" .

    nrl:RDFS
          a       nrl:Semantics ;
          nrl:semanticsDefinedBy
          "http://www.w3.org/TR/rdf-schema#" ;
          nrl:semanticsLabel "RDF/S" .
}
```

Fig. 1. NRL Example using TriG Serialization

VI. Conclusion

The first version of the export plug-in (which is publicly available for download[4]) was the first step towards the creation of a plug-in that is able to provide Protégé with full support for NRL This can become possible in the near future, given that Protégé supports graph sets that handle multiple views on each graph. Handling graph views is the biggest challenge so far.

Acknowledgements

This work was supported by the European Union IST fund (Grant FP6-027705, Project NEPOMUK). The authors would especially like to thank all contributors to NEPOMUK's ontology taskforce.

References

[1] M. Sintek, L. van Elst, S. Scerri, S. Handschuh. Distributed Knowledge Representation on the Social Semantic Desktop: Named Graphs, Views and Roles in NRL. In Proceedings of the 4th European Semantic Web Conference, ESWC07, Innsbruck, Austria.

[2] J.J. Carrol, C. Bizer, P. Hayes, and P. Stickler. Named graphs, provenance and trust. In WWW'05: Proceedings of the 14th International Conference on World Wide Web, pages 613-622, New York, NY, USA, 2005. ACM Press.

[3] S. Decker and M. Frank. The social semantic desktop. In Proc. of the WWW2004 Workshop Application Design, Development and Implementation Issues in the Semantic Web, 2004.

[4] P. Hayes. RDF semantics. W3C recommendation, W3C, February 2004.http://www.w3.org/TR/rdf-mt/.

[3] http://sites.wiwiss.fu-berlin.de/suhl/bizer/TriG/

[4] http://smile.deri.ie/tools/NRLExporter