

Reuse Globally Unique URIs in Protégé Software

Paolo Bouquet, Heiko Stoermer, Xin Liu (University of Trento, Italy)
{bouquet, stoermer, liu@disi.unitn.it}

An ontology is a principal form of data in the Semantic Web (SW). Especially individuals play an important role for information generation and reasoning. We currently face the problem of globally unique identifiers of individuals in ontologies^[1]. In the current state of Semantic Web, it is fairly common that individuals in ontologies have different identifiers even if they are meant to denote the same resource in the real world. Different users or applications assign and use different identifiers for the same real-world entity. This makes data integration from various data sources at the level of individuals quite difficult.

If we could assume that individuals in different ontologies which represent the same actual resources have the same URI, the heterogeneous data problem would be simplified to a large extent. In this way, recognizing the same resource distributed in heterogeneous data sources would be only based on simply comparing their identifiers rather than depending on their data structures or content similarities for approximating identity.

The Okkam project¹ is dedicated to solving the issue of identifiers crisis in the SW. It proposes a web-scale infrastructure called the Entity Name System (ENS)^[2] to support the reuse of existing URIs for any type of individuals (person, location, organization, event, etc). The central principle of the ENS is to match the request provided by users or agents and the entity properties stored in the ENS repository in order to return a list of candidate individuals, in which users can reuse the correct URIs. This can enable knowledge creation applications to reuse globally unique identifiers for individuals.

Protégé software is a widely used ontology creation and editing tool. But the naming approach of items is either the stochastic mechanism of application itself or the manual input of human beings. So we developed a plug-in called Okkam4P^[3] for protégé-OWL editor to extend its function of reusing pre-existing URIs for individuals. Now, the plug-in is compatible with protégé 3.x and protégé 4.

Okkam4P makes use of an extension point which allows a plug-in to appear in the right-click popup menu of a selected individual and can be effortlessly hooked into the application architectures. In order to reuse an existing URI for an individual, Okkam4P starts to collect the datatype properties of the selected individual and send a request to the ENS. The request consists of the individual type and its characteristic description, which is a collection of datatype property name-value pairs with the ability of distinguishing different individuals. After dealing with the request, the ENS returns a ranked list of candidate individuals which are similar ones with the description. Users can pick up one correct individual and reuse its URI. As a result, the local identifier of the corresponding individual in XML code will be substituted to the globally unique URI. At the same time, the prefix will be changed from “rdf:ID” to “rdf:about”, which means the value is an absolute URI and refer to a global resource. In this way, the identifiers of individuals in independently created ontologies are global URIs, thus causing the integration of different ontology resources at individual level to be much easier.

Okkam4P can also work at the client side for Collaborative Protégé², which supports collaborative ontology editing, thus allowing multiple clients to add or edit different individuals and to reuse the globally unique URIs simultaneously. This is of benefit to building a huge database which comes from different users' consciousness. The diffusion of the use of Okkam4P can help to address the difficulties during the process of data integration.

References

- [1]. Paolo Bouquet, Heiko Stoermer, and Daniel Giacomuzzi. OKKAM: Enabling a Web of Entities. In *i3: Identity, Identifiers, Identification*. Banff, Canada, May 8, 2007.
- [2]. Paolo Bouquet, Heiko Stoermer, Claudia Niederee, and Antonio Mana. Entity Name System: The Backbone of an Open and Scalable Web of Data. In *Proceeding of ICSC 2008*, number CSS-ICSC 2008-4-28-25.
- [3]. Paolo Bouquet, Heiko Stoermer, and Liu Xin. Okkam4P - A Protégé Plug-in for Supporting the Re-use of Globally Unique Identifiers for Individuals in OWL/RDF Knowledge Bases. In *Proceedings of SWAP2007*, Italy.

¹ <http://www.okkam.org>

² http://protegewiki.stanford.edu/index.php/Collaborative_Protege