

An Interactive Map of Semantic Web Ontology Usage

Sheila Kinsella, Uldis Bojārs, Andreas Harth, John G. Breslin, Stefan Decker
 Digital Enterprise Research Institute (DERI),
 National University of Ireland, Galway
 firstname.surname@deri.org

Abstract

Publishing information on the Semantic Web using common formats enables data to be linked together, integrated and reused. In order to fully leverage the potential for interlinking data by reusing existing schemas, an intuitive way of viewing current usage of RDF vocabularies is required. We present a system which allows a user to view the most frequently occurring namespaces and classes in a large Semantic Web dataset, and the main linkage patterns that exist between them. Users can select a namespace of interest in order to examine usage of a particular ontology, and see how it is being combined with other vocabularies.

1. Introduction

The way in which ontologies are used is important information to those who create and maintain them. The Linked Data¹ initiative encourages the interlinking of data in order to increase its value and usefulness. One way of increasing the potential for interlinking when publishing data is to reuse terms from existing vocabularies [2]. Therefore, creators of ontologies could benefit from a way to view schema usage in order to help select existing terms to use where possible, and only invent new ones if necessary. Also, maintainers of an ontology may want to monitor usage of classes and properties so that decisions on modifications to the schema can be based on how the ontology is actually being used.

Understanding how data is represented on the Semantic Web is difficult because of the scale and complexity of the data involved. The amount of semantic data available is large and growing. Google² estimates around 427k RDF documents on the Web, PingtheSemanticWeb³ lists 982k RDF documents, SWSE (Semantic Web Search En-

gine)⁴ provides query answering over 1.3m RDF sources, and Sindice⁵ indexes around 27m RDF documents. This data originates from many different sources and is highly heterogeneous. At the instance level, there is just too much data to sift through. In order to get an insight into ontology usage we need to abstract the problem to a higher level and look at classes and the relationships between them. However, even at the class level there is a huge amount of information. PingtheSemanticWeb is aware of 505 namespaces and over 5k classes. Our sample dataset which we describe later in the paper includes approximately 27k classes and 1.5k namespaces.

To assist people interested in exploring RDF schema usage, there is a need for a tool that can show ontology usage information in a way that is clear and comprehensible, preferably a visual system in order to show the linked nature of the data. The system should enable users to see a high level overview of the dataset and zoom in to see specific details. The ability to handle large datasets is also vital.

Previous studies of Semantic Web vocabulary usage have analysed the most used ontologies [4] and the distribution of class and property usage [3]. Existing work involving ontology visualisations include OntoVis [11], which uses structural and semantic abstraction to create graphical representations of large heterogeneous social networks, and Crop-Circles [9], which aims to provide insight into the complexity of class hierarchies in an ontology. Examples of systems that enable RDF visualisation at the instance level include IsaViz [10] and RDF Gravity [5]. Previous work on the reuse of ontologies includes Oyster [8], a peer-to-peer system for exchanging ontology metadata among communities to help people find and share ontologies. This work states the importance and difficulties of ontology reuse, and describes the vocabulary and system developed to enable ontology-sharing. Another related work investigates the construction of ontologies automatically from existing ones [1], and suggests that reuse of existing ontologies should encourage more participation in the Semantic Web.

¹<http://linkeddata.org/>

²<http://www.google.com/> visited 24/04/08

³<http://pingthesemanticweb.com/> visited 24/04/08

⁴<http://swse.deri.org/> visited 24/04/08

⁵<http://sindice.com/> visited 29/01/08

We present a system that allows a user to view the namespaces and classes present in a dataset and the linkages that exist between them. Users can view in detail a namespace of interest in order to examine usage of a particular ontology, and see how an ontology or a class is being used with other vocabularies. Sample visualisations are provided on a large dataset crawled from the Semantic Web.

2. Motivation

A system that allows viewing of Semantic Web namespaces, classes and their interconnections could be useful for multiple purposes including the following:

- **Designing a new ontology**

Designers of a new ontology can decide how they want to fit it in with existing ontologies, for example those which are already popular or well-connected to others. If there is already a class or property in existence which can be reused, then it is preferable to do so than to create a new one. In addition, if designers have already decided to use one schema to represent certain items of data, they may wish to also use other closely linked schemas, in order to have some consistency with other data publishers. These approaches increase the ease with which the data could be interlinked with other sources. However, in order for a data creator to take these factors into consideration when they choose the way in which they will represent their data, a method is required to be able to assess the extent to which properties and classes are already being used.

- **Viewing usage of an ontology**

By exploring the class level view of a particular ontology, creators of an ontology can compare actual usage of an ontology to the way in which the ontology was designed to be used. They can see which other vocabularies an ontology is being combined with, and the most popular terms that are being used. In some cases, knowledge about how users utilize a vocabulary could help determine future development.

- **“Geeky exploration”**

Finally, the system provides a way for those with a general interest in the Semantic Web and Linked Data to get an idea of the most common types of data being published on the Semantic Web and how information is being represented.

3 System Overview

In the following we describe the functionality and implementation of our ontology map system.

3.1. Functionality

The system provides two levels of views, a high level view at the namespace level and a lower level view at the class level. Users can move from the namespace level to the class level by selecting a particular ontology that they wish to see in more detail. Rather than displaying full URIs, shorthand prefixes are used in order to make the maps more easily readable.

- **Namespace level**

At the top level view, the most frequently occurring namespaces are displayed, as in Figure 1, with edges linking those that are commonly connected. Each namespace is shown as a node and labelled with its shorthand URI and a number that indicates the number of times that an instance is defined as belonging to a class of this namespace. The user can hover the mouse pointer over the arrowhead of an edge between two namespaces to view the number of links between instances belonging to classes of the relevant namespaces. The user can click on the shorthand URI of a namespace in order to view the corresponding class level map.

- **Class level**

The class level view shows the most frequently occurring classes belonging to a particular namespace, as well as classes from other namespaces that they are directly connected to. Those that are commonly connected are linked with an edge. An example is shown in Figure 2. To view usage of properties, the user can hover the mouse pointer over the arrowhead connecting two classes, and a box will appear showing a ranked list of the properties that most frequently link instances of these classes, as in Figure 3. The user can click on the shorthand URI of a class or property to look up that class or property in SWSE and retrieve related information.

3.2. Implementation

The creation of the browsable ontology map consists of the following steps, which are summarised in Algorithm 1.

3.2.1 Counting classes and links

The RDF dataset under investigation is indexed and stored using YARS2 [7] in quadruples of the form $\langle \text{subject, predicate, object, context} \rangle$ or *spoc* where context indicates the URI of the document containing the statement. From the `rdf:type` statements, a separate index allowing faster

lookups is created in BerkeleyDB⁶ to store a map of instance URIs to their corresponding classes.

The class counts are derived by iterating over the index and counting the objects of `rdf:type` statements. A *spo* statement that occurs multiple times in different contexts will increment counts as many times as it occurs, so one entity that is defined in two different documents will be counted twice. Therefore, the counts do not indicate how many instances there are of a certain type, but are only an indication of relative usage frequency. We also derive counts at the namespace level by considering the namespace of each class.

The link counts are calculated by iterating over the index and for each statement that indicates a link between two instance resources, the types of the subject instance and the object instance are retrieved. A count is kept for each distinct combination of subject class, predicate and object class. We also derive link counts at the namespace level by considering the namespace of the subject class and namespace of the object class.

3.2.2 Generating the graphs

Having generated class and link counts at both the namespace and class level, the next step is to represent these with graphs and enable visualisation of the dataset at either level. One namespace level graph is created, as well as a class level graph for each namespace in the namespace level graph. The class level maps contain classes within the corresponding namespace, and also directly connected classes from other namespaces. In order to keep the graphs at a manageable size, only the most frequently occurring (top-k) nodes and edges are selected for display. Other criteria for selecting nodes and edges can also be used. Rather than using full URIs, we replace the namespace with a prefix as in Table 1. Each map is defined using the GraphViz⁷ format and SVG⁸ images of the graph are generated using GraphViz's `neato` utility.

3.2.3 Adding interactivity

Finally, we connect the set of individual maps to form an interactive system. The namespace map is edited to link each namespace to its class level map, and the class level maps are edited to include links from each class and property to its result page in SWSE. We use Javascript to enable the display of popup boxes showing the link counts between namespaces in the namespace map, and showing the link counts between classes and the most frequently occurring properties between classes in the class level maps. Zoom and pan facilities are also added.

⁶<http://sleepycat.com/>

⁷<http://www.graphviz.org/>

⁸<http://www.w3.org/Graphics/SVG/>

Algorithm 1 Algorithm for generating ontology usage map

Require: Graph G in quadruples *spoc*

Require: List L of namespaces and corresponding prefixes

Require: Integer k for selecting top-k nodes and edges

for each quad in G where *p* is `rdf:type` **do**

 increment class count for *o*

 increment count for namespace of *o*

 add entry to database mapping *s* to class *o*

end for

for each quad in G where *s* and *o* are resources AND *p* is not `rdf:type` **do**

for each class of *s* **do**

for each class of *o* **do**

 increment count of links from class of *s* to class of *o* with predicate *p*

 increment count of links from namespace of class of *s* to namespace of class of *o*

end for

end for

end for

generate namespace graph from top-k namespace counts, using prefixes from L

for each namespace in namespace-level graph **do**

 generate top-k class-level graph from class counts, using prefixes from L

end for

4. Sample Map

The process described in Section 3.2 to generate an interactive map of ontology usage was applied to a crawl of RDF data that was carried out during January 2008 using Multi-crawler [6]. The dataset contains over 90m RDF statements, and includes entities from approximately 27k classes, from over 1.5k namespaces. In the following we present and discuss screenshots of the interactive map, which is also viewable online⁹. Table 1 shows a list of the namespaces and prefixes that appear in the following screenshots.

In these graphs, we display at most 20 nodes and 45 edges. Additionally, only nodes and edges with a count of at least 1k are included. In the class-level graphs, first the most popular classes within the namespace are added, in order to ensure that the included nodes and edges are those that are most relevant to the particular namespace. Next added are links to the classes inside or outside the namespace that are most strongly linked to classes within the namespace. Finally, we add any other edges between classes in the graph.

Figure 1 shows the top level view of the most common namespaces in the dataset. Only frequently occurring links are shown, so namespaces that are just occasionally connected do not have a link displayed between them. This

⁹<http://sw.der1.org/2007/06/ontologymap/>

Prefix:	Namespace URI (http://):	Ontology describes:
admin	webns.net/mvcb/	Terms for describing vocabularies
affx	www.affymetrix.com/community/publications/tmsplice#	A genetic information dataset
akt	www.aktors.org/ontology/portal#	A project demonstrator (AKTive Portal)
eco	purl.org/obo/owl/ECO#	Experimental and evidence statements
fips55	www.daml.org/2003/02/fips55/fips-55-ont#	US Federal Information Processing
fips55t	www.daml.org/2003/02/fips55/types#	FIPS class codes
foaf	xmlns.com/foaf/0.1/	People and relationships
geo	www.w3.org/2003/01/geo/wgs84_pos#	Geographical information
obo	www.geneontology.org/formats/oboInOwl#	Open Biomedical Ontologies in OWL
owl	www.w3.org/2002/07/owl#	The Web Ontology Language
rdfs	www.w3.org/2000/01/rdf-schema#	RDF vocabulary description language
rss	purl.org/rss/1.0/	Web feed formats schema
sc	purl.org/science/owl/sciencecommons/	Biomedical information ontology
sioc	rdfs.org/sioc/ns#	Information in online communities
sioc	rdfs.org/sioc/types#	Extends core SIOC ontology
skos	www.w3.org/2004/02/skos/core#	Taxonomies and other vocabularies
tapxmlns	tap.xmlns.com/data/	Project to make web of machine-readable data
wn	www.cogsci.princeton.edu/~wn/schema/	WordNet(English language lexical database)
wordnet	xmlns.com/wordnet/1.6/	WordNet(another version)
words	truesense.net/words.en.01.owl#	WordNet(another version)

Table 1. Prefixes, URIs and descriptions of the most common namespaces in the dataset

screenshot shows that the FOAF (Friend-of-a-Friend)¹⁰ vocabulary in particular is strongly connected to many other popular schemas. As a result of hovering the mouse pointer over the arrowhead connecting the RDF namespace to the FOAF namespace, the number of links connecting them is also displayed.

By using the mouse to click on the name of any namespace in Figure 1, a class level view of that namespace is obtained. Figure 2 shows a view of the FOAF namespace, displaying the most frequently occurring FOAF classes and related classes, and the most common links that exist between them. An interesting point to note from this image is that one of the most commonly used classes is `foaf:chatEvent`, a class which is not defined in the official FOAF specification¹¹, but which is often used to store Internet Relay Chat (IRC) logs in RDF.

Figure 3 shows the result of hovering the mouse pointer over the over the arrowhead connecting `foaf:Person` to `foaf:Document`. The properties that most commonly connect these two classes, and the count of each one, are displayed. One of the properties in this list, `foaf:knows`, actually has a domain and range of `foaf:Person`. This may indicate either that some entities are being declared as both a `foaf:Person` and a `foaf:Document`, or that `foaf:knows` is being used to link to instances of type

`foaf:Document` (perhaps personal homepages) instead of to URIs representing the persons who are actually known.

5. Conclusions

This paper shows how visualising Semantic Web schema usage and interconnectivity at the namespace level and the class level can give enhanced insight into how vocabularies are being used. Future work includes displaying the top literal properties for each class. The current system displays only information about the links between instances of classes, but the properties linking instances of classes to literals would also be of interest. Additionally, we plan to carry out task-based evaluation in order to assess the usefulness of the system. We hope that this system provides interesting and useful information for people exploring ontology usage on the Semantic Web.

Acknowledgments

This work has been partially funded by Science Foundation Ireland under grant number SFI/02/CE1/I131 and a Supplemental Equipment Grant.

¹⁰<http://www.foaf-project.org/>

¹¹<http://xmlns.com/foaf/spec/>

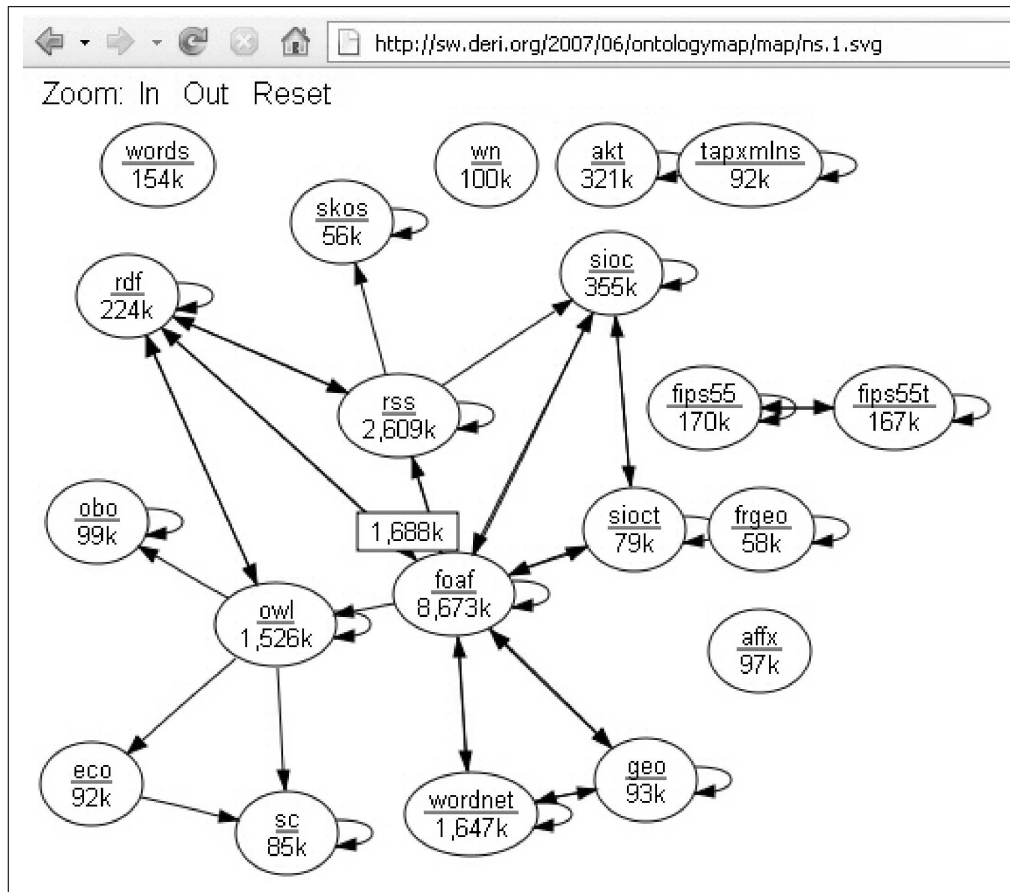


Figure 1. Namespace level view, showing the most frequently occurring namespaces in the dataset. The number of links from the RDF namespace to the FOAF namespace is shown as a result of hovering the mouse pointer over the arrowhead connecting the two namespaces.

References

- [1] H. Alani. Position Paper: Ontology Construction from On-line Ontologies. *Proceedings of the 15th international conference on World Wide Web*, pages 491–495, 2006.
- [2] C. Bizer, R. Cyganiak, and T. Heath. How to Publish Linked Data on the Web, 2007.
- [3] L. Ding and T. Finin. Characterizing the Semantic Web on the Web. *Proceedings of the 5th International Semantic Web Conference*, 2006.
- [4] L. Ding, L. Zhou, T. Finin, and A. Joshi. How the Semantic Web is Being Used: An Analysis of FOAF Documents. *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*, 2005.
- [5] S. Goyal and R. Westenthaler. RDF Gravity (RDF Graph Visualization Tool). *Salzburg Research, Austria*, 2004.
- [6] A. Harth, J. Umbrich, and S. Decker. Multicrawler: A pipelined architecture for crawling and indexing semantic web data. In *5th International Semantic Web Conference, Athens, GA, USA.*, 2006.
- [7] A. Harth, J. Umbrich, A. Hogan, and S. Decker. YARS2: A Federated Repository for Searching and Querying Graph-Structured Data. In *Proceedings of the 6th International Semantic Web Conference*, November 2007.
- [8] R. Palma, P. Haase, and A. Gómez-Pérez. Oyster: sharing and re-using ontologies in a peer-to-peer community. *Proceedings of the 15th international conference on World Wide Web*, pages 1009–1010, 2006.
- [9] B. Parsia, T. Wang, and J. Golbeck. Visualizing Web Ontologies with CropCircles. *Proceedings of the 4th International Semantic Web Conference*, pages 6–10, 2005.
- [10] E. Pietriga. Isaviz: a visual environment for browsing and authoring rdf models. *The Eleventh International World Wide Web Conference (Developers day)*, 2002.
- [11] Z. Shen, K. Ma, and T. Eliassi-Rad. Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction. *Visualization and Computer Graphics, IEEE Transactions on*, 12(6):1427–1439, 2006.

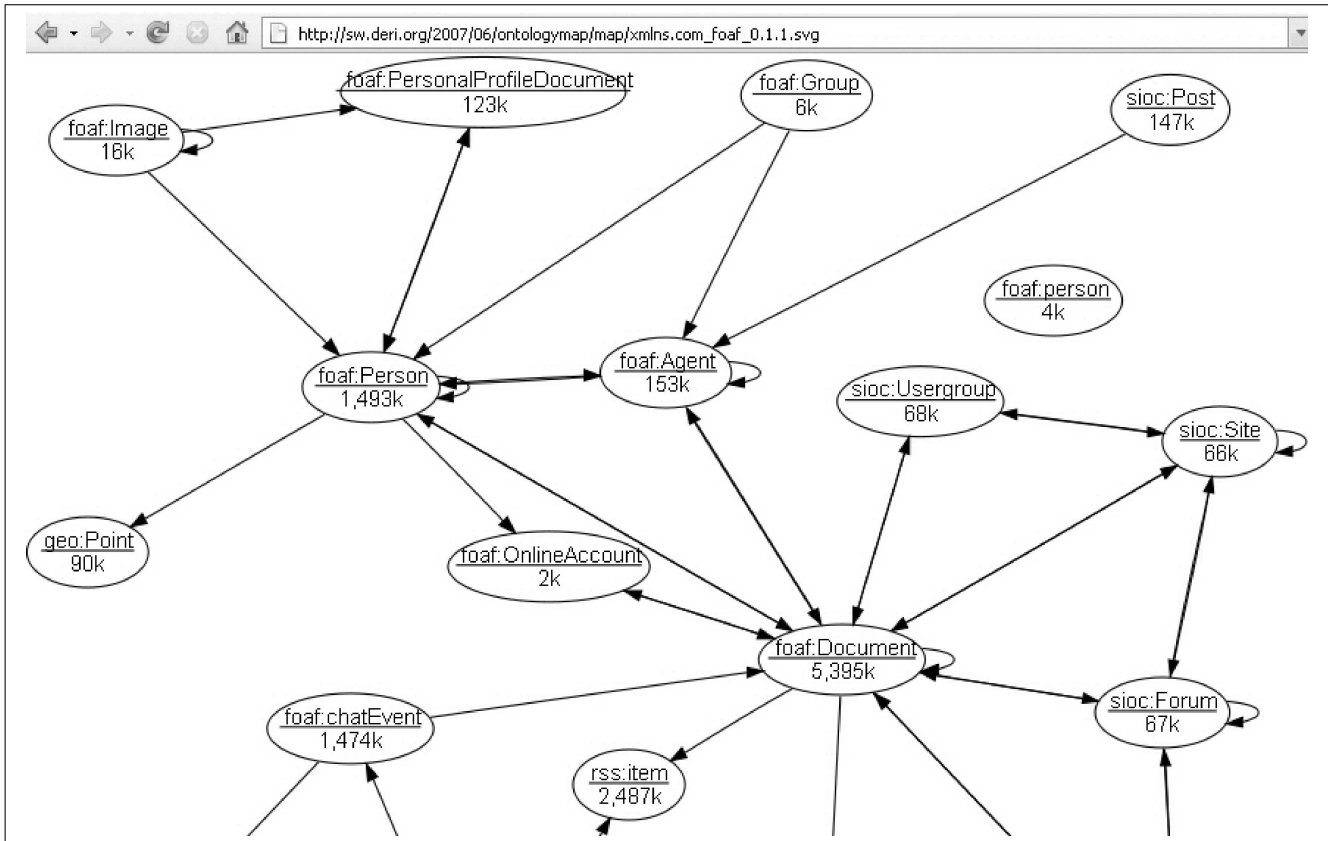


Figure 2. Class level view of the FOAF namespace, showing the most common FOAF classes and closely related classes, and the links between them. This view is reached by clicking the text reading “foaf” in Figure 1.

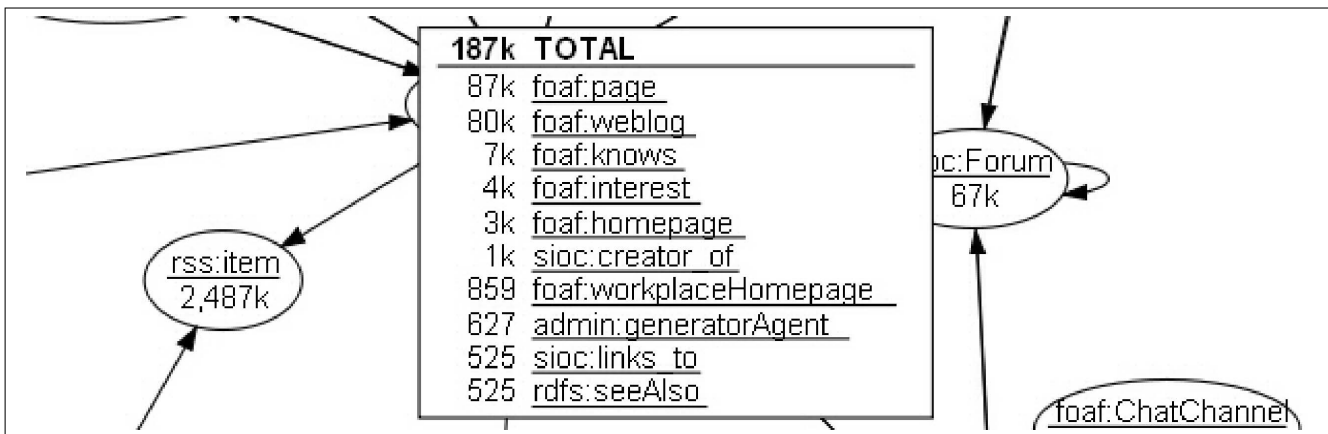


Figure 3. Popup box showing the properties connecting foaf:Person to foaf:Document and their counts. This box appears as a result of hovering the mouse pointer over the arrowhead connecting foaf:Person to foaf:Document in Figure 2.