

rSim: Simplifying an RDF Graph at the Visualization Tier for Non-Expert Users

Rathachai Chawuthai*, Hideaki Takeda

SOKENDAI (The Graduate University for Advanced Studies), Japan
National Institute of Informatics, Tokyo, Japan
{rathachai,takeda}@nii.ac.jp

Abstract. Using a concept-map or a node-link diagram for representing knowledge can enhance the learning ability of users. However, the queried RDF graph can be complex and difficult for non-expert users to read, because many redundant triples entailed by RDFS and OWL rules are also included. Existing tools can rearrange, filter, highlight and summarize a graph for visualization but they require human effort. For this reason, this research attempts to use Semantic Web rules and schemas to simplify a graph automatically at the visualization tier. The prototype shows that the simplified graph is easier to read than the original one. In future, this method will be positioned as a plugin for other visualization tools in order to serve attractive RDF-based knowledge to users.

Keywords: Graph Simplification, RDF Visualization, Semantic Web Rule

1 Introduction

This research introduces a method to simplify a highly connected RDF (Resource Description Framework) graph at the visualization tier, and demonstrates an easily readable graph for non-expert users. This contribution is suitable for consumers who (1) are willing to use RDF graph visualization as learning and teaching tools, and (2) believe in the role of Linked Open Data (LOD) for future knowledge management.

It is known that displaying RDF data to general users is very challenging as RDF expressions are not human-friendly. For this reason, graph visualization becomes an alternative way to enhance the learning ability of users from beginners to advanced levels. In this paper, a term “beginner” is used as a user who does not know Semantic Web well (lay-user) [1], and one is new to domain knowledge (non-domain-expert) [2]. In terms of using LOD for education, concept maps can be successfully adopted for e-learning for beginners [3], and the graph should be sparse and easy to read [4].

However, in practice, a displayed graph is sometimes extremely difficult to read as shown in Figure 2(a) due to a large number of inferred triples. They are resulted from some well-prepared RDF repositories having reasoning on schemas, rules, and same-as instances (*owl:sameAs*). In general, a schema defines members of classes, hierarchies of classes and properties, and characteristics of properties. For example, *rdf:type*, *rdfs:subClassOf*, *rdfs:subPropertyOf*, *owl:equivalentProperty*,

owl:TransitiveProperty, *owl:inverseOf*, etc. are always found in any schema. After having RDFS and/or OWL entailments, a network dataset becomes more complex because giant components and node pairs with multiple edges are arisen. This condition makes a visualization tier is not suitable for reading.

To display an easy and readable RDF graph, there are several pieces of research working on this issue. *Motif Simplification* compresses some topologies of subgraphs into basic shapes such as diamonds and fans [5]. *RDF Gravity* offers a zoomable graph with text overlays [6]. *Fenfire* gives a clear display for immediate nodes, but fades away the others, which are far from the focus [7]. *IsaViz* uses graph style sheets to draw a graph with meaningful icons and it groups metadata into tables to display in a canvas [8]. *Motif Simplification* aims to give the overall structure of a network rather than detail. *RDF Gravity* and *Fenfire* do not reduce the issue of redundant data. *IsaViz* requires human effort to filter some nodes or properties to sparsify a dense graph. For this reason, this study is an attempt to simplify an RDF graph for beginners by reducing some redundant nodes and inferred links with less human effort.

According to the mentioned issue, this paper proposes a rule-based methodology named *rSim* where its scope is to give a simplified graph visualization for beginners to read and learn knowledge represented by RDF data. A prototype is implemented in order to demonstrate the possibility and feasibility of the proposed approach.

2 *rSim* Approach

To visualize an RDF query result for beginners, *rSim* (RDF Simplification) approach is introduced to eliminate some redundant triples which are resulted from reasoning on schemas. This research initiates a set of rules (*rSim's rules*) that works with the available schemas to clean up a graph of query result. Thus, it requires queried triples, schemas, and *rSim's rules* as inputs to produce a clean graph for the visualization tier as shown in Figure 1. The involved components are displayed in the dash rectangular. The query result is in RDF format, the schemas such as *owl* files are provided by any repositories, the *rSim's rules* are described hereafter, and the process “Simplify” does the reasoning on the three input components for generating simplified RDF data.

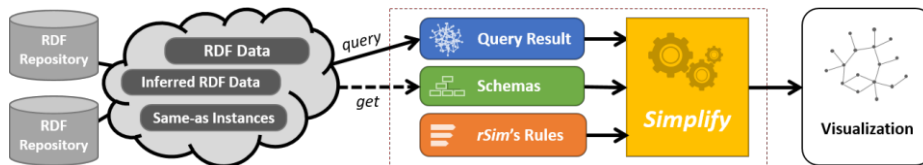


Fig. 1. Overview of *rSim* approach

The *rSim's* ruleset is formally defined in Table 1 and the objective of each rule is described in the following list.

- Rules #1-4 are to merge either nodes or triples that contain some same-as resources by selecting high frequently occurring nodes to display.
- Rule #5 is to remove some transitively inferred individual-class relationships.

- Rule #6 is to eliminate some links that are resulted from transitive entailment.
- Rules #7-8 are to select only one link among multiple links of a node pair based on the characteristics of properties that are transitive and inverse links.
- Rule #9 is to choose the most frequent property from some equivalent properties.
- If a canvas cannot well-display multiple links between a node pair, the Rule #10 is used to select only one link that is the most frequent occurring property.

The occurrence number should be counted across repositories, because a popular URI has a high chance to discover more knowledge in next queries. Further, the graph should display human-readable labels instead of URIs.

Table 1. List of *rSim*'s rules for transforming a dense graph into a better visualization (In this table, $\langle s p o \rangle$ is a triple including a subject s , a predicate p , and an object o . The term $|x|$ is a number of the URI " x " occurred in datasets.)

#	Queried Triples	Schema/Matched Data	Count	Display only
1	$\langle s_1 p_1 o_1 \rangle, \langle s_2 p_1 o_1 \rangle$	$\langle s_1 owl:sameAs s_2 \rangle$	$ s_1 > s_2 $	$\langle s_1 p_1 o_1 \rangle$
2	$\langle s_1 p_1 o_1 \rangle, \langle s_1 p_1 o_2 \rangle$	$\langle o_1 owl:sameAs o_2 \rangle$	$ o_1 > o_2 $	$\langle s_1 p_1 o_1 \rangle$
3	$\langle s_1 p_1 o_1 \rangle, \langle s_2 p_2 o_2 \rangle$	$\langle s_1 owl:sameAs s_2 \rangle$	$ s_1 > s_2 $	$\langle s_1 p_1 o_1 \rangle, \langle s_1 p_2 o_2 \rangle$
4	$\langle s_1 p_1 o_1 \rangle, \langle s_2 p_2 o_2 \rangle$	$\langle o_1 owl:sameAs o_2 \rangle$	$ o_1 > o_2 $	$\langle s_1 p_1 o_1 \rangle, \langle s_2 p_2 o_1 \rangle$
5	$\langle s_1 rdf:type c_1 \rangle, \langle s_1 rdf:type c_2 \rangle$	$\langle c_1 rdfs:subClassOf c_2 \rangle$		$\langle s_1 rdf:type c_1 \rangle$
6	$\langle x_1 p_1 x_2 \rangle, \langle x_2 p_1 x_3 \rangle, \langle x_1 p_1 x_3 \rangle$	$\langle p_1 rdf:type owl:TransitiveProperty \rangle$		$\langle x_1 p_1 x_2 \rangle, \langle x_2 p_1 x_3 \rangle$
7	$\langle s_1 p_1 o_1 \rangle, \langle s_1 p_2 o_1 \rangle$	$\langle p_1 rdfs:subPropertyOf p_2 \rangle$		$\langle s_1 p_1 o_1 \rangle$
8	$\langle s_1 p_1 o_1 \rangle, \langle o_1 p_2 s_1 \rangle$	$\langle p_1 owl:inverseOf p_2 \rangle$		$\langle s_1 p_1 o_1 \rangle$
9	$\langle s_1 p_1 o_1 \rangle, \langle s_1 p_2 o_1 \rangle$	$\langle p_1 owl:equivalentProperty p_2 \rangle$	$ p_1 > p_2 $	$\langle s_1 p_1 o_1 \rangle$
10	$\langle s_1 p_1 o_1 \rangle, \langle s_1 p_2 o_1 \rangle$		$ p_1 > p_2 $	$\langle s_1 p_1 o_1 \rangle$

3 Discussion

To demonstrate the possibility and feasibility of the proposed approach, a prototype is developed and accessible at "<http://rc.lodac.nii.ac.jp/rsim/>". The dataset and schema are from the LODAC repository¹ that gathers a large number of inferred triples of biodiversity data. In our experiment, we queried the whole neighborhood of a given node within two hops, and compared the original query result with our simplified graph. Figure 2(a) shows about 200 raw triples queried using the resource *lodac:Bubo*, whereas *rSim* simplifies them into 20 triples as shown in Figure 2(b).

For the user perspective, any beginners, who have less background knowledge about biodiversity and RDF, are appreciated the result of *rSim*. This work helps to inspire users to use a graph visualization as a learning tool. It also induces them to understand and realize the power of Linked Open Data for knowledge management.

In future, we intend to integrate *rSim* with some existing visualization features from the related works such as zoom in/out, blur unfocused nodes, replace blurred area by motif shapes, group node-attribute triples into a table, show meaningful icons, and allow user-defined conditions for filtering triples.

¹ LODAC: Linked Open Data for Academia (<http://lod.ac>)

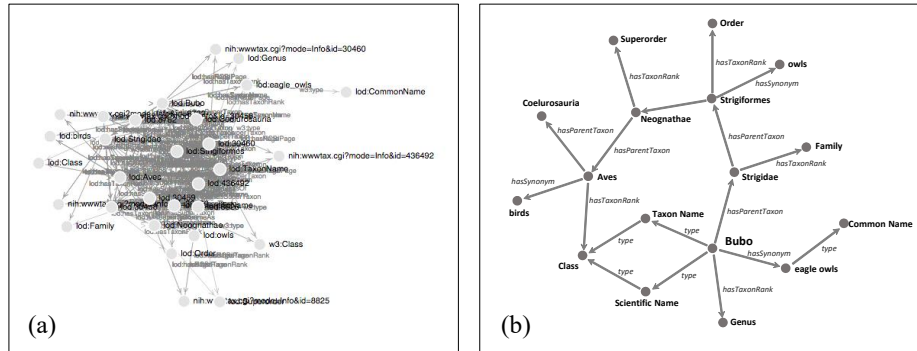


Fig. 2. Graph visualization of information queried from the LODAC repository. In this figure, (a) is an original query result, and (b) is a result of *rSim* work.

4 Conclusion and Future Work

This research introduced the method name *rSim* to simplify an RDF graph for visualizing data to general users by using Semantic Web rules to eliminate some redundant triples. The proposed rules can be viewed as inversions of RDFS and OWL in order to handle some cases of same-as resources, equivalent properties, class hierarchies, property hierarchies, transitive properties, inverse properties, etc. Most redundant triples are eliminated while the remaining triples can serve the overall information that is easy to be read by for non-domain-expert users.

Since this work focuses on RDF visualization for beginners, more details should be added according the learning ability of a user when they becomes more expert. This research is planned to be enhanced for more flexibility to every level of users based on domain knowledge, user preference, and ability to understand RDF expression.

References

1. Dadzie, A., Rowe, M.: Approaches to visualising Linked Data: A survey. In: Semantic Web Journal, pp. 89-124 (2011)
2. Edelson, D., Gordin, D.: Visualization for learners: a framework for adapting scientists' tools. In: Computers & Geosciences, pp. 607-616 (1998)
3. Schwendimann, B.: Concept maps as versatile tools to integrate complex ideas: From kindergarten to higher and professional education. (2015)
4. Novak, J., Cañas, A.: The theory underlying concept maps and how to construct and use them. In: Florida Institute for Human and Machine Cognition (2006)
5. Dunne, C., Shneiderman, B.: Motif simplification: improving network visualization readability with fan, connector, and clique glyphs. In: SIGCHI2013, pp. 3247-3256 (2013)
6. Goyal, S., Westenthaler, R.: RDF gravity (RDF graph visualization tool). (2004)
7. Tuukka, H., et al.: Browsing linked data with Fenfire. In: LDOW 2008. (2008)
8. Pretorius, J., et al.: What does the user want to see? What do the data want to be?. In: Information Visualization, vol. 8(3), pp. 153-166. (2009)