

ZoomRDF: Semantic Fisheye Zooming on RDF Data

Kang Zhang[†], Haofen Wang[†], Duc Thanh Tran[‡], Yong Yu[†]

[†]Dept. of Computer Science & Engineering, Shanghai Jiao Tong University, China

{jobo, whfcarter, yyu}@apex.sjtu.edu.cn

[‡]Institute AIFB, University of Karlsruhe, Germany

dtr@aifb.uni-karlsruhe.de

ABSTRACT

With the development of Semantic Web in recent years, an increasing amount of semantic data has been created in form of Resource Description Framework (RDF). Current visualization techniques help users quickly understand the underlying RDF data by displaying its structure in an overview. However, detailed information can only be accessed by further navigation. An alternative approach is to display the global context as well as the local details simultaneously in a unified view. This view supports the visualization and navigation on RDF data in an integrated way. In this demonstration, we present ZoomRDF, a framework that: i) adapts a space-optimized visualization algorithm for RDF, which allows more resources to be displayed, thus maximizes the utilization of display space, ii) combines the visualization with a fisheye zooming concept, which assigns more space to some individual nodes while still preserving the overview structure of the data, iii) considers both the importance of resources and the user interaction on them, which offers more display space to those elements the user may be interested in. We implement the framework based on the Gene Ontology and demonstrate that it facilitates tasks like RDF data exploration and editing.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces

General Terms

Human Factors, Design

Keywords

RDF Data, Information Visualization, Semantic Web

1. INTRODUCTION

With the development of Semantic Web in recent years, more and more semantic data has been made available in form of RDF triples. Currently, there are billions of triples in publicly available datasets on the Web describing real-world entities of different domains in terms of their classes, attributes, attribute values and relations between them. For

instance, the Gene Ontology¹, which is widely used in life science, contains 472,041 triples. It describes more than 16,000 gene and gene product attributes of a large number of organisms.

High-quality RDF data has proven to be a useful tool in the areas of digital libraries, bioinformatics, and personalized information management. Recently, many Semantic Web applications have shown additional values and are gaining momentum by integrating visualization in the process of searching and browsing RDF data. As a prominent example, ThinkBase², a visualization interface for Freebase, displays a visual representation of the Freebase contents and its semantic relationships in the form of a network graph. Users can click and drag the visual graph to explore the data interactively.

While the amount of RDF data and the number of related applications are increasing rapidly, effective and efficient exploration on RDF data poses several requirements:

- An *intuitive visual presentation* which shows explicitly the structure of RDF data and fully utilizes the available space to maximize visibility of resources.
- An *efficient navigation* to target elements with as few user interactions as possible.
- A *context-aware exploration* considering both the properties of RDF data itself and information derived from user interactions.

In this demonstration, we present ZoomRDF, a framework for large scale RDF data visualization and navigation. To satisfy the above requirements, the framework has the following main features:

- It adapts a *space-optimized visualization* on RDF data, which not only displays the global structure of the data explicitly but also fully utilizes the display space to show more resources.
- It leverages a *fisheye zooming* so that more space is assigned to the selected elements while the overall context is still preserved, which ensures both global and local views to be presented simultaneously.
- It considers the importance of resources as well as the user interactions performed on them. By assigning each resource a *Semantic Degree of Interest* (SDOI),

¹<http://www.geneontology.org/>

²<http://thinkbase.cs.auckland.ac.nz/>

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.

ACM 978-1-60558-799-8/10/04.

ZoomRDF is able to identify more important elements users might be interested in and to reorganize the display layout to make them more obvious to the users.

2. RELATED WORK

Information visualization techniques have been used recently to provide more intuitive presentations of RDF data. RDF Gravity [2], OntoViz [6] and Welkin [3] render RDF data as node-link diagrams in which nodes represent resources, and links stand for relations between them. However, it is difficult for users to see the details of individual resources if a large RDF graph is visualized on a limited-size screen. Jambalaya [7] tries to solve this problem by allowing scrolling on a large node-link diagram to support navigation on it. Since there is no support for "orientation", users might have to try out and scroll to different directions. Navigating like this in a large graph might take much time to find a resource.

A more sophisticated approach is to navigate through RDF data using its class hierarchy, which can be regarded as guided scrolling. It can save the time as well as the number of interactions required for locating resources. Previous work based on this approach can be classified into two main categories according to different visualization algorithms they use: (1) connection based visualization draws the connections from the parent nodes to their children on a tree. It allows users to see the relations between nodes using a set of graphical edges [1] but may waste some display space, and (2) space-division based visualization divides the space into many enclosures to which each node on the tree is mapped [8, 10]. It has the advantage of one hundred percent space utilization [4] but cannot show the connections between nodes explicitly. In addition, most prior methods do not consider the selected elements and previous user interactions as the context for visualization.

3. FRAMEWORK

In this section, we introduce the data model, the workflow and the main components of ZoomRDF.

3.1 Data Model

Given a set of RDF data, we present its data model as a hierarchy tree. It is a rooted tree whose nodes are classes and instances from the data. There exists an edge between classes in the hierarchy tree if and only if there is a *subclass* relation between them in the RDF data. Besides, an edge exists between a class and an instance in the hierarchy tree if and only if there is a *type* relation between them in the data. Figure 1 shows a partial hierarchy tree extracted from the Gene Ontology. Notice that each node v in the hierarchy tree is attached with a score called *semantic degree of interest*, $SDOI(v)$. The definition and calculation of SDOI will be introduced in the next subsection.

3.2 SDOI Calculation

The user interest on the data can be captured by the importance of the data and user interactions on it. In particular, a SDOI value can be decomposed into two parts: (1) the *static semantic degree of interest* (SSDOI) which indicates the importance of a resource in the RDF data, and (2) the *dynamic semantic degree of interest* (DSDOI) determined by user interactions. More precisely, for a node v , we define

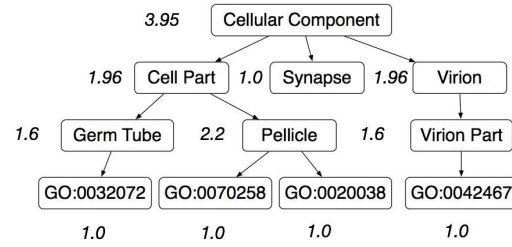


Figure 1: The sample hierarchy tree from Gene Ontology

its SDOI value as:

$$SDOI(v) = SSDOI(v) \times DSDOI(v) \quad (1)$$

The SSDOI value of the node is calculated by the importance of its corresponding resource in the RDF data. For a node representing an instance, we define its importance as 1. For a node representing class, its importance is determined by the multiplication of itself and the total importance of its children. The DSDOI value of the node reflects the user's attention on it. Initially, the value is set to 1. It will then be updated according to two kinds of user interaction: (1) expanding a node indicates the user is more interested in that node and increases its DSDOI value, and (2) collapsing an expanded node shows the user loses his/her interest on it and decreases its DSDOI value. Therefore, when the user expands or collapses a node, its SDOI value will be updated based on the dynamically computed DSDOI value.

3.3 Workflow

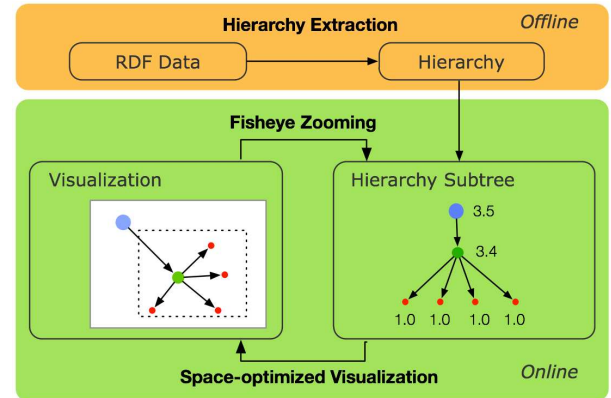


Figure 2: The workflow of ZoomRDF

Figure 2 illustrates the workflow of ZoomRDF. During the offline step, *Hierarchy Extraction* extracts the hierarchy tree from the input RDF data. At the same time, the SDOI value of each node is calculated and is attached to the node on the hierarchy tree. When the online exploration starts, *Space-optimized Visualization* generates the layout for a selected portion of the hierarchy tree according to the SDOI values of its nodes. After the user interacts with the hierarchy tree, such as expands a node on the screen, *Fisheye Zooming* rearranges the layout and leaves free space for the new resources to be displayed. Accordingly, the SDOI value of the selected node will be updated, and the space-optimized visualization

will help render the updated part of the layout. This way, we tightly integrate the space-optimized visualization with the fisheye zooming, which helps the user understand the RDF data by interactive exploration.

3.4 Space-optimized Visualization

ZoomRDF adapts the space-optimized tree visualization [4] to fully utilize the screen space for showing the relations explicitly. It helps users easier to understand the global structure of the RDF data. The main idea of the visualization is using enclosures to partition the display space into a collection of polygon areas and assign them to the nodes of the tree according to their SDOI values. This ensures one hundred percent space utilization. More importantly, this gives the resources which have higher SSDOI values more space to display, in other words, the resources which are more useful to the user look more obviously in ZoomRDF.

3.5 Fisheye Zooming

Traditional navigation approaches on a hierarchy tree employ a linear scaling to provide the detailed view for a node. When viewing details of the node, the user has to decrease the visualization scale to see the global context of the whole hierarchy tree. On the contrary, the fisheye zooming emphasizes certain details while still preserving the complete global context.

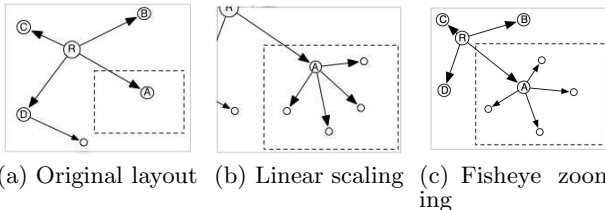


Figure 3: Linear scaling vs. fisheye zooming

Figure 3(a) is the original layout generated for the hierarchy tree of an example RDF data snippet. When the user expands the node A, the linear scaling displays the details of A clearly (see Figure 3(b)) while losing some global context information, such as the siblings of A in the example. Instead, fisheye-based zooming [5] preserves the global context including A's parent and all siblings appeared in the original layout (see Figure 3(c)). Meanwhile, it shows the details of A as clearly as the linear scaling does.

The fisheye zooming process of ZoomRDF first distorts the layout using an *fish-eye transformation*, which pushes the nodes away from the user selected node within the screen. Second, it deletes the less interested nodes, i.e., the nodes with lower SDOI values, from the areas of high node density. Finally, it applies the space-optimized visualization on the free space (left from the fisheye transformation) for the newly added children of the user selected node.

4. DEMONSTRATION SCENARIOS

We implement ZoomRDF using Adobe Flex³ and Python in a client/server fashion. The client side focuses on data visualization and user interaction while the server maintains the hierarchy tree for the Gene Ontology and sends back the selected portion to the client on demand.

³<http://www.adobe.com/products/flex/>

To test the usability, we apply ZoomRDF in two different scenarios, i.e., RDF data exploration and RDF data editing. Then we explain how it improves user experience and help accomplish tasks in both scenarios

4.1 RDF Data Exploration

RDF data exploration helps users understand the underlying knowledge by exploring the content of the RDF data. Traditional approaches cannot present the global structure of the RDF data and the local details of some resources to the user simultaneously. Unlike these approaches, ZoomRDF integrates the space-optimized visualization with the fisheye zooming so that both global and local information are organized in a unified view. Thus, the user does not need to switch to another view for the global structure when he or she is looking at information of some individual resources. Additionally, since ZoomRDF keeps the global context almost unchanged when viewing different portions of the RDF data, the whole data exploration process is smooth and offers good user experience.

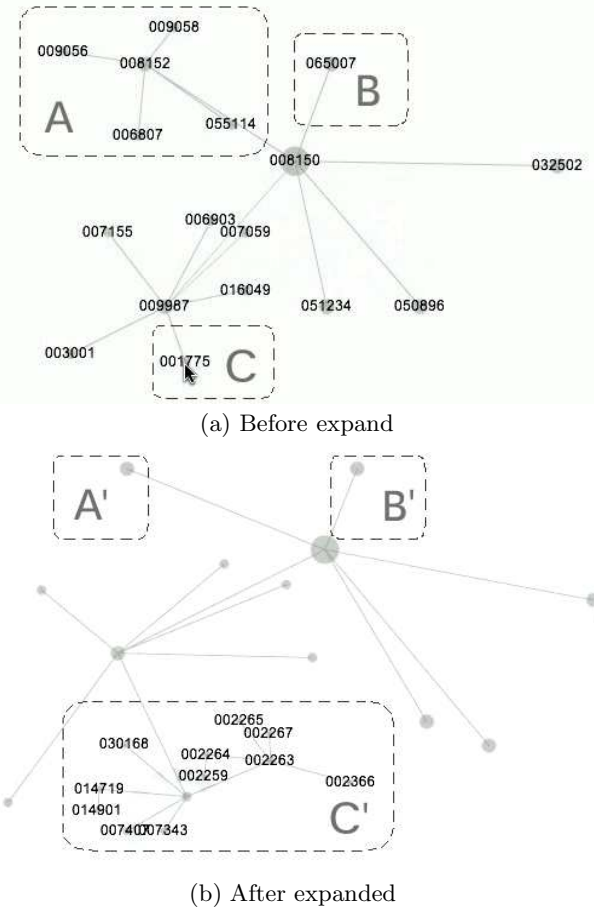


Figure 4: Preserved global details in exploration

Figure 4 gives an example of data exploration on the Gene Ontology. Figure 4(a) shows the initial layout which contains a number of gene terms and their IDs. The node in the rectangle C is the one that will be expanded by the user. As shown in Figure 4(b), after the expansion, the fisheye zooming rearranges the layout, and the children of

some nodes (such as the nodes in the rectangles A and B) are collapsed automatically so that more space is left to the expanded node and its children. Although the fisheye zooming distorts the layout and collapses some nodes, the nodes in the rectangles A' and B' have been preserved to keep the same global structure of the data to the user (see Figure 4(b)). Thus, the user is guided by the global context information, and he or she could directly expand the node in rectangle A' or B' to see its details. Meanwhile, the children of the expanded node has been displayed at the bottom of the screen (indicated by rectangle C' in Figure 4(b)) using the space-optimized visualization. As the number of newly added nodes on the screen is determined by the available unused space, some children of the selected node have been expanded as well, which improves the utilization of the display space.

4.2 RDF Data Editing

RDF data editing is a service for users to find and modify certain kind of resources in the RDF data, which is quite common in applications like semantic data annotating and ontology engineering. When dealing with large RDF data, it always takes a lot of time and effort for users to locate a given resource to be edited.

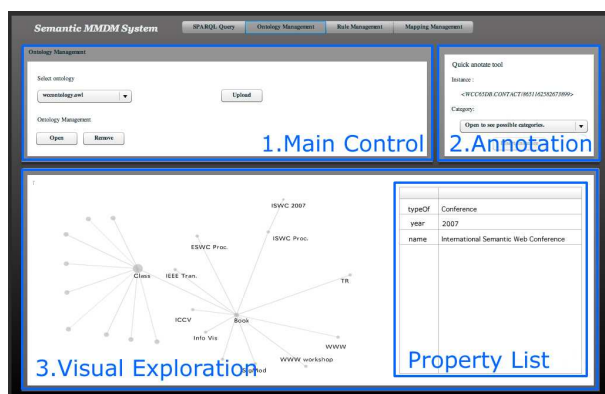


Figure 5: The user interface for Semantic Master Data Management System

To solve the above problem, we apply ZoomRDF to RDF data editing, which provides feedbacks visually to highlight potential nodes that users are looking for. In particular, we integrate ZoomRDF into an ontology engineering environment that is used for insurance and clinical data analysis, namely, the Semantic Master Data Management System [9]. Figure 5 shows its main components: (1) the *main control* panel on the top left used to open and upload a RDF data, (2) the *annotation* panel on the top right for annotating resources, and (3) the *visual exploration* panel in the center used to explore the data visually using ZoomRDF. Notice that we also display the properties of a selected resource in a list (see the *property list* in Figure 5) which is a complementary way to ZoomRDF for RDF data visualization. In this way, the system presents the structure of the RDF data visually in a holistic view.

ZoomRDF guides users to narrow the search space in the RDF data through its hierarchy tree, and finally to find the target resource without too many user interactions. More specifically, we list its benefits as follows: (1) the space-

optimized visualization enables more resources to be displayed on the screen, which provides more information to users, (2) the fisheye zooming preserves the global context when displaying the local details so that users are aware of their orientation during interaction, and (3) the user interest model assigns more space to the selected node as its SDOI value is increased by boosting the corresponding DS-DOI value during user interaction. Since ZoomRDF leaves more display space for the visited nodes, the corresponding path from the root looks more obvious. As a result, the user find it easier to find the target resource by navigating through the path.

5. CONCLUSIONS

In this demonstration, we presented ZoomRDF, a framework for large scale RDF data visualization and navigation. By employing a space-optimized visualization algorithm, ZoomRDF is able to display the data as a node-link diagram in a fully utilized space. The fisheye zooming concept adopted for this framework provides the global context of an ontology as well as the local details of selected elements in a unified view. This allows faster navigation to targeted nodes. Furthermore, a SDOI data model is proposed for the framework to consider both the importance of data elements and the user interactions. This allows resources potentially more useful to the user to be presented more obviously.

6. REFERENCES

- [1] C. Fluit, M. Sabou, and F. van Harmelen. Ontology-based information visualization. In *Visualizing the Semantic Web*, pages 36–48. 2002.
- [2] S. Goyal and R. Westenthaler. Rdf gravity (rdf graph visualization tool), May 2004. <http://semweb.saltzburgresearch.at/apps/rdf-gravity/>.
- [3] S. Mazzocchi and P. Ciccarese. Welkin, May 2005. <http://simile.mit.edu/welkin/>.
- [4] Q. V. Nguyen and M. L. Huang. Space-optimized tree: a connection+enclosure approach for the visualization of large hierarchies. *IV*, (1):3–15, 2003.
- [5] M. Sarkar and M. Brown. Graphical fisheye views of graphs. In *SIGCHI*, pages 83–91, 1992.
- [6] M. Sintek. Ontoviz tab: Visualizing protege ontologies, May 2003. <http://protegewiki.stanford.edu/index.php/OntoViz>.
- [7] M.-A. D. Storey, N. F. Noy, M. A. Musen, C. Best, R. W. Fergerson, and N. Ernst. Jambalaya: an interactive environment for exploring ontologies. In *the 7th international conference on Intelligent user interfaces*, pages 239–239, 2002.
- [8] K. Tu, M. Xiong, L. Zhang, H. Zhu, J. Zhang, and Y. Yu. Towards imaging large-scale ontologies for quick understanding and analysis. In *ISWC*, pages 702–715, 2005.
- [9] X. Wang, X. Sun, F. Cao, L. Ma, N. Kanellos, K. Zhang, Y. Pan, and Y. Yu. Smdm: Enhancing enterprise-wide master data management using semantic web technologies. In *Demonstrations of VLDB*, 2009.
- [10] M. Xiong, Y. Chen, H. Zheng, and Y. Yu. Towards Quick Understanding and Analysis of Large-Scale Ontologies. In *ASWC*, page 84, 2006.