SemaSun: Visualization of Semantic Knowledge based on an improved Sunburst Visualization Metaphor

Christian Stab, Matthias Breyer, Kawa Nazemi, Dirk Burkhardt, Cristian Hofmann, Dieter W. Fellner
Fraunhofer Institute for Computer Graphics Research
Fraunhoferstrasse 5, 64283 Darmstadt, Germany
{christian.stab, matthias.breyer, kawa.nazemi, dirk.burkhardt, cristian.hofmann, d.fellner}@igd.fraunhofer.de

Abstract: Ontologies have become an established data model for conceptualizing knowledge entities and describing semantic relationships between them. They are used to model the concepts of specific domains and are widespread in the areas of the semantic web, digital libraries and multimedia database management. To gain the most possible benefit from this data model, it is important to offer adequate visualizations, so that users can easily acquire the knowledge. Most ontology visualization techniques are based on hierarchical or graph-based visualization metaphors. This may result in information-loss, visual clutter, cognitive overload or context-loss. In this paper we describe a new approach of ontology visualization technique called *SemaSun* that is based on the sunburst visualization metaphor. We improved this metaphor, which is naturally designed for displaying hierarchical data, to the tasks of displaying multiple inheritance and semantic relations. This approach also offers incremental ontology exploring to reduce the cognitive load without losing the informational context.

Introduction

Ontologies have become an established data model for conceptualizing knowledge entities and describing semantic relationships between them. They are used to model the concepts of specific domains and are widespread in the areas of the semantic web, digital libraries and multimedia database management (Katifori et al. 2007). However, ontologies may become very large and complex what makes it difficult for the common user to understand the underlying knowledge space (Kriglstein and Motschnig-Pitrik 2002). To alleviate ontology exploration and knowledge acquirement, visualizations are needed, so that common users can gain the most benefit from this kind of data models.

For representing ontologies, an adequate visualization must be able to display the hierarchical inheritance structure, multiple inheritance and semantic relations between ontology entities (Katifori et al. 2007). For this reason tree-based visualizations like indented lists are not sufficient for visualizing this kind of data model, because they are limited to the visualization of hierarchical data. Thus, most of the common ontology visualizations are based on graph visualization methods. In contrast to tree-based visualizations these node-link diagrams are able to display ontologies without information loss. However graph-based visualizations become unwieldy if the visualized information become larger and make inefficient use of screen space (Plaisant et al. 2002). It is also complicated to maintain the context of the information while the user navigates through the knowledge space and to reduce the cognitive overload of the user.

In this paper we describe a new approach of ontology visualization technique called *SemaSun* that is based on the sunburst visualization metaphor. We improved this metaphor, which is naturally designed for displaying hierarchical data, to the tasks of displaying multiple inheritance and semantic relations. Our approach also offers incremental ontology exploring to reduce the cognitive load of the users without losing the informational context.

In the following section we provide an overview of existing ontology visualizations and discuss their advantages and drawbacks. We introduce our approach of ontology visualization in the section "SemaSun – Sunburst Visualization for Ontologies" and give a detailed description of the SemaSun Visualization and its advantages compared to other existing ontology visualizations. We conclude this paper with a case study and a conclusion of our work.

Related Work

Nowadays there are many different approaches for visualizing ontologies. In this section we present an overview of different ontology visualization techniques and discuss the advantages and drawbacks. We start with a short technical introduction to the fundamentals of ontologies and their structure. After that we categorize some visualizations according to the ontology visualization survey from Katifori et al. (Katifori et al. 2007) and present representatives for every category.

Fundamental description of Ontologies

For ontology visualization the three most important data elements are *concepts*, *individuals* and *relations*. These elements can contain further properties which describe various features and attributes. Concepts of an ontology represent abstract models of entities in the domain of interest (Gruber 1993), for example the concept *person*. These concepts are defined as terminological statements in the schema, which tends to be more permanent. A concept can inherit properties from other concepts using *subclass-of* relations. This inheritance structure constitutes a (overlapping) hierarchy describing the domain of interest as generalized concepts which become more and more specific downwards this hierarchy, like *'male person' subclass-of 'person'*. Next to this inheritance structure semantic relations are defined in the schema to model contextual references.

On the instance level of the ontology individuals are defined which instantiate concepts and specify inherited properties. The individuals represent real world objects and are the actual data of the modeled domain, for example *Bob* and *Joe* are two instantiations of the concept *male person*. Furthermore semantic relations between individuals are specified on the instance level to model a concrete relationship between two individuals, like *'Bob' is-brother-of 'Joe'*. Each relation has a direction, a type and a label. Especially this label is important for the field of ontology visualization, so the user can more easily understand the semantics of that relationship (Studer et al. 2006).

To ensure a reasonable schema design ontology-experts work on the process of conceptualizing the domain of interest. But besides the design for reliable reasoning, ontologies are designated to be used as databases for applications domain-experts and common users interact with. To gain the most benefit for these common users an ontology visualization technique must be able to impart the multiple inheritance, the concept hierarchy and the semantic relations between ontology entities (Katifori et al. 2007).

Indented List Visualizations

Indented Lists are tree-based visualizations that offer a Windows Explorer like tree view of an ontology. Because of their familiarity to the common user, indented lists are easy to use and allow high performance in ontology exploring (Katifori et al. 2006). They are used in most of the ontology management systems like the Protégé Class Browser (Noy et al. 2000), OntoRama (Eklund et al. 2002) and Kaon (Kaon, http://kaon.semanticweb.org). Tree-based visualizations provide a clear view of entity labels and the concept hierarchy. However, this kind of visualization has several drawbacks in the task of ontology visualization. Indented lists are only applicable for representing the hierarchical part of the ontology. Thus the representation of semantic relations and multiple inheritance is not feasible. Furthermore, only a limited part of the ontology can be displayed at once. The top-down layout results in poor space-filling causing the need for scrolling during ontology exploration (Parsia et al. 2005). For this reason indented lists are not very applicable for imparting the general structure of the ontology.

Node-link Visualizations

Node-link visualizations represent the concepts and individuals as nodes and relations as edges. In contrast to indented lists, the representation of multiple inheritance and semantic relations is feasible, by interconnecting a child with edges to all its parents. For this reason this visualization technique is used for many different ontology visualizations, like OntoViz (Sintek 2003), IsaViz (Pietriga 2004), OntoTrack (Liebig and Noppens 2004), OntoSphere (Bosca et al. 2005) and WSMOViz (Kerrigan 2006). Node-link visualizations are well suited for imparting an overview of the entire ontology structure. Nevertheless they make inefficient use of screen space (Plaisant et al. 2002). For large datasets this leads to an insufficient presentation of the whole structure of the ontology which results in context-loss and the need for scrolling. Also the visualization of many relations may result in confusing diagrams with overlapping edge labels.

Zoomable Visualizations

In zoomable visualizations the hierarchy of the ontology is represented by nesting nodes of lower levels inside their parents. Usually the user is able to zoom into the child nodes in order to gather information from items at deeper levels. This visualization technique is used for the visualization plug-in Jambalaya (Storey et al. 2001) for the Protégé ontology tool, CropCircles (Parsia et al. 2005) and SemaSpace (Bhatti and Weber 2009) (Bhatti 2008). Zoomable ontology visualizations provide a clear overview of the ontology's hierarchy. The user is able to request details-on-demand for items of interest by zooming into the desired entity which reduces the cognitive load of the user. But on the other hand the context of the selected element is lost and in some cases it is difficult to recognize the parent node of the zoomed entity or to identify its level in the hierarchy. Relations between the ontology elements are usually visualized as directed, labeled links and are displayed by default (Jambalaya) or on-demand (CropCircles). If the ontology contains many relations this type of relation visualization ends in visual clutter and overlapping labels what makes it difficult for the user to acquire the needed information.

Space-Filling Visualizations

Space-filling visualizations are based on the concept of using the whole screen space by subdividing the available space for a node among its children (Katifori et al. 2007). The best known representative of space-filling visualizations is the Treemap visualization proposed by Shneiderman (Shneiderman 1992). It uses a 2D approach of space-filling to represent hierarchies, in which each node is a rectangle and has been applied by Baehrecke et al. for visualizing ontologies (Baehrecke et al. 2004). Treemaps are efficient when users are interested in the leaf nodes and provide a good overview if the hierarchy is trivial (Katifori et al. 2007). If the hierarchy becomes larger and deeper, significant cognitive effort is needed in order to understand the hierarchical structure of the visualized information (van Ham and van Wijk 2002), so this type of visualization does not offer an efficient way to impart knowledge from complex ontologies (Nazemi et al. 2009). Another space-filling approach for visualizing ontologies is SeMap, proposed from Nazemi et al. (Nazemi et al. 2009). SeMap allows the incremental exploring of the ontology's hierarchy which reduces the cognitive load of the user. The exploration starts with the root node and the user can expand a single path of entities of interest. The drawback of this approach is that the user is only able to examine one path of the hierarchy at once. The main drawback of both visualizations is that they are only applicable for visualizing the hierarchy of ontologies and are not feasible for an appropriate visualization of semantic relations and multiple inheritance.

Context, Focus and Distortion Visualizations

Context, Focus and Distortion Visualizations are based on the concept of distorting the view of a visualized graph. The user is able to select a node of interest in order to focus and enlarge it. The focused node is usually centered and other nodes are placed around the focused node reduced in size. This technique is used in ontology visualizations like TGVizTab (Alani 2003) and OntoRama (Eklund et al. 2003). Both representatives are based on graph visualizations and thus they are able to represent multiple inheritance and relations between the entities of the ontology. The advantage of context, focus and distortion visualizations is that an entity of interest is visualized without losing its context. The drawback of this visualization technique is that the position of nodes alters when the user selects a new node and thus it is complicated for the user to keep track of the visualized ontology structure and to understand the complete inheritance structure.

SemaSun – Sunburst Visualization for Ontologies

Our approach of ontology visualization follows the *Visual Information-Seeking Mantra* proposed by Shneiderman (Shneiderman 1996). First the user is able to gather an overview of the entire ontology to understand the complete structure of the knowledge space. After gaining an overview by exploring the inheritance structure, the user is able to focus on entities of interest and to request details of these entities on demand. So the user is able to easily explore the ontology and to understand the knowledge space of the ontology. In the following sections we describe the details of the SemaSun Visualization and how our approach tackles the problems of information loss, visual clutter, cognitive overload and context-loss.

Incremental ontology exploration

To reduce the cognitive overload of the user, the SemaSun visualization does not show the entire structure of the ontology as initial state. At the startup of the visualization, the root concept of the ontology is shown (Figure 1a) and the user is able to incrementally explore the hierarchy of the ontology by expanding entities of interest (Figure 1b). The radial layout of the sunburst visualization offers thereby the expansion of multiple paths (Figure 1c) so users are able to gather an overview of the whole inheritance structure and are not limited to the exploration of a single path.

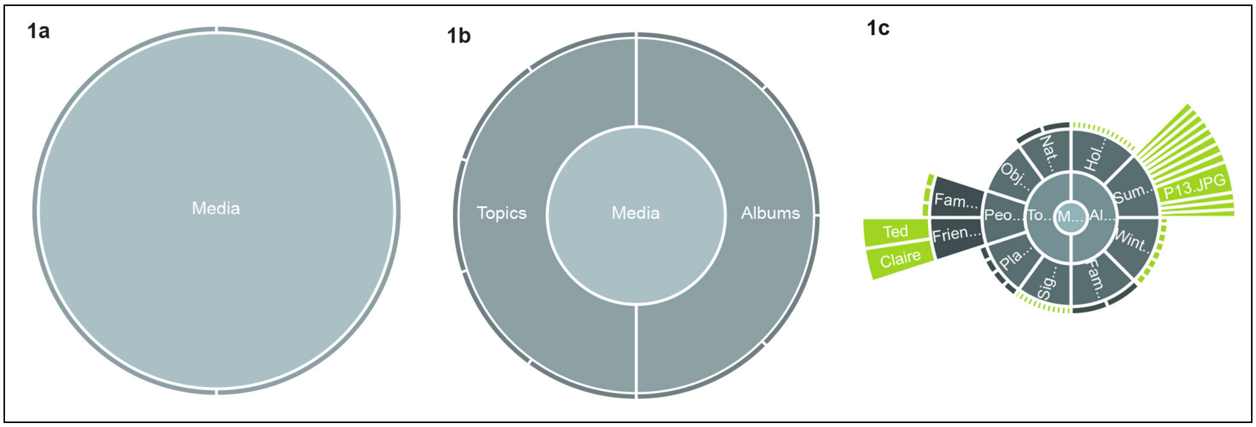

Figure 1: Incremental Ontology Exploration without losing contextual information.

To maintain the informational context and to avoid the need for scrolling, entities that users visited earlier move closer to the center and are reduce in their size (Figure 1c). If the user moves the cursor over an entity whose label is not completely visible the size and the angle of the entity is adjusted so that the whole label is visible. This distortion technique is especially important if the slices become thin or the user wants to explore small elements near the center. To advise users of the existence of child nodes, an expandable mark is shown around entities that contain child nodes. The number of child nodes is denoted as the number of arcs around the parent node (Figure 1). To distinguish different entities concepts and individuals are visualized with different colors (Figure 1c).

Multiple Focus Support

During the exploration process, users may find entities they are interested in. In order to not loose these relevant entities during the further exploration of the knowledge space, we integrate multiple focus support to the SemaSun visualization.

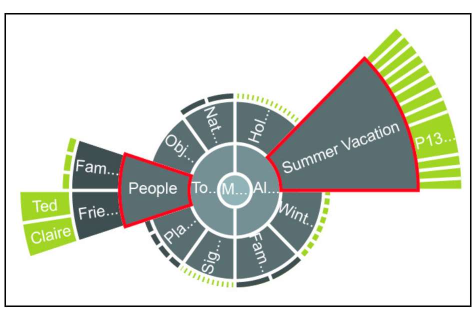

Figure 2: Multiple Focus Support.

With this feature users are able to mark entities of interest, so that the labels of these focused entities are always visible and are not reduced in their size by the distortion effect. Figure 2 shows an example of the multiple focus feature, where the user has marked the two concepts *People* and *Summer Vacation*. In the following sections we describe how the multiple focus feature is used for visualizing multiple inheritance and semantic relations between knowledge entities.

Support of Multiple Inheritance

One characteristic of ontologies is multiple inheritance. In cases where the ontology contains entities with multiple parents, tree-based visualizations are not sufficient for visualizing the whole knowledge of the ontology. To avoid information loss, we improved the sunburst visualization metaphor that is naturally designed for tree-based data, for the task of visualizing entities with multiple parents. For that purpose an entity that inherits properties from multiple parents is duplicated in the hierarchy among all its parents. If users navigate to such an entity all duplicates of this entity are focused by the multiple focus feature, so that the labels of all duplicates are visible. In order to better notify the user of the repeated occurrences, the color of all duplicates is highlighted and matched (Figure 3). In cases where a duplicate of an entity is collapsed, the expandable mark of the related path is highlighted (Figure 3) so the user can easily be aware of the existence of the invisible duplicate.

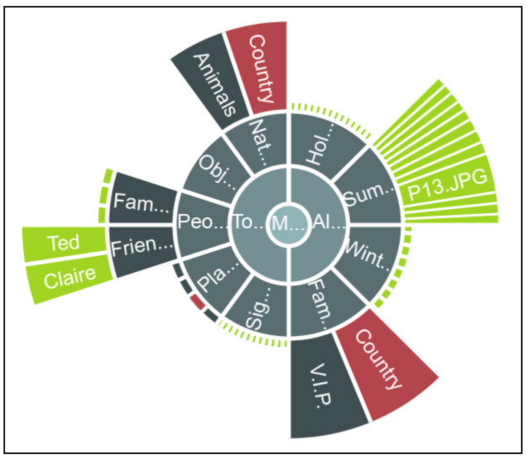

Figure 3: Visualizing Multiple Inheritance with the SemaSun visualization

Visualization of semantic relations

Beside the multiple inheritance, semantic relations are an important characteristic of ontologies. As mentioned before these links between ontology entities are not presentable with visualizations for hierarchical data. For this reason we improved the sunburst visualization for visualizing semantic relations. To reduce the cognitive load of the user our approach of ontology visualization does not show semantic relations by default, rather the user is able to request these information for an entity of interest on demand. On request of the relation information for a specific entity, all visible entities not semantically related to the requested one will be reduced in their alpha values. Semantically related entities are highlighted using the multiple focus feature, so these entities are adjusted in their sizes and angles to ensure their label readability. This visual awareness supports the user to clearly perceive the semantic relations without losing the informational context. Additionally to the linked elements, each semantic relation contains a link name which should be displayed, so that the user is able to understand the type of the specific relation. These link names are shown as a curved text inside of the entities and the direction of the relation is denoted with a triangle aside the link name. So overlapping labels are avoided and the user is able to understand the relations between the highlighted entities. Figure 4 shows an example of the SemaSun visualization displaying the relations of an image file. The entities which are semantically related with this image are highlighted and the relation names are shown inside of the highlighted nodes.

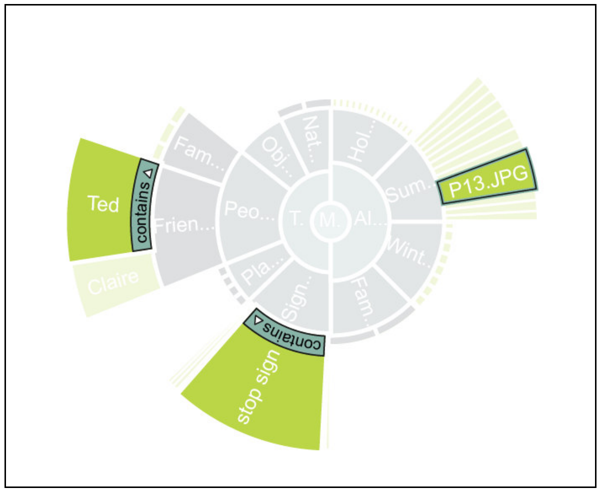

Figure 4: Visualization of semantic relations with SemaSun

Case Study

The described SemaSun visualization was developed as a part of the *Core-Technology-Cluster* (*CTC*) of the THESEUS Program (Theseus 09), a 60-month program partially funded by the German Federal Ministry of Economics and Technology. The partners in the THESEUS Program research under the device "*New Technologies for the Internet of Services*" heterogeneous technologies for gathering and offering semantic information on web. The program itself consists of twelve projects, divided in *THESEUS Use Cases* and the *THESEUS Core-Technology-Clusters*. Where the six Core-Technology-Clusters are lead by research institutions and focus on fundamental research areas, the THESEUS Use Cases are lead by enterprise institutions and bridge the gap between fundamental research and industrial dissemination. Different enterprise partners focus on their usage scenario of the different areas of information processing. For example the Siemens Corporation investigates the processing of medical-related information. In this THESEUS Use Case (Medico) different usage scenarios identify different user groups: There are medical doctors, who use the information of the patient's clinical and medical records to find similar cases and provide the adequate care for them. On the other hand you have the patients themselves, who should be able to understand about their disease and find for example groups or a community with similar ailments.

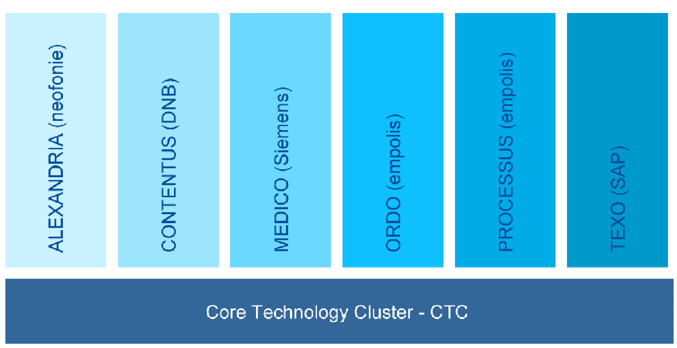

Figure 5: Structure of the THESEUS Program

Another example for a THESEUS Use Case is THESEUS Texo, which is conducted by SAP Research Germany and investigates different models and techniques for providing, engineering, disseminating and using web-based services. The different user roles e.g. service-engineers, domain-experts or service-consumers handle the same information in different ways. Complex information structures defined as ontologies makes it necessary to provide here a best fit of visualizing the information structures for the different user-groups and their precognitions.

A third example for a THESEUS Use Case is Contentus, lead by the German National Library. Here you find the same heterogeneity of users. There are domain-experts, who have the required knowledge in a specific scientific domain, e.g. experts for German Literature, but are not experts in using and processing complex ontology-based information-systems. Of Course you find in Contentus the average user too, who just explores knowledge domains and expects a very simple to use visualization and user interface. Figure 5 shows the general structure of the THESEUS Program, the six Use Cases and their leading partner.

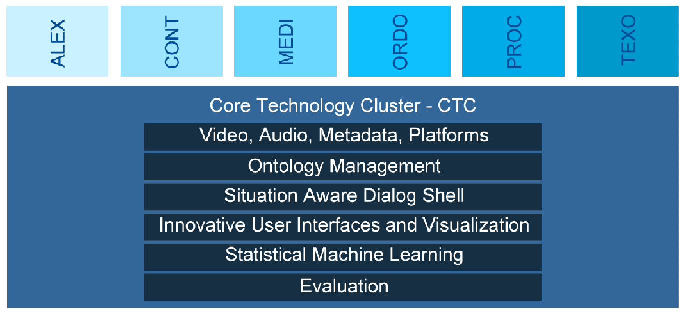

Figure 6: THESEUS Core-Technology-Clusters

Beyond the THESEUS Use Cases there are six THESEUS Core-Technology-Clusters (CTCs) investigating different fundamental research questions regarding semantic information processing. The CTC are mainly conducted by research institutes. The CTC for Ontology Management, lead by "Forschungszentrum Informatik", investigates for example managing, reasoning, editing and inferencing ontologies. The CTC Situation Aware Dialog Shell investigates different questions regarding context-aware information processing. Figure 6 shows the different Core-Technology-Clusters in the THESEUS Program:

The SemaSun visualization was developed as a CTC-component of the THESEUS Core-Technology-Cluster Innovative User Interfaces and Visualizations and is used in different THESEUS Use Cases for exploring knowledge domains within the specific scenarios of the Use Cases. SemaSun was tested in different Use Cases in their specific semantic knowledge domain. The first tests in real scenarios, where users were able to choose between different visualizations at the beginning of their knowledge exploration, pointed out a higher frequency of usage compared to other provided ontology visualizations, especially for the average user.

Discussion

SemaSun is a visualization concept for visualizing semantic knowledge, so that the user is able to easily understand the modeled knowledge. The used sunburst visualization metaphor provides a clear view of the ontology's concept hierarchy and provides the exploration of multiple paths without losing the informational context. With the integration of the incremental navigation design the cognitive overload is reduced and the users are able to focus on knowledge entities of interest. The extension for the visualization of multiple inheritance and semantic relations using the multiple focus feature allow the displaying of ontologies with the sunburst visualization metaphor without losing information. In particular our approach for visualizing semantic relations is a novel way of imparting semantic knowledge. Instead of connecting two graphical objects with edges, we used highlighting techniques and the multiple focus feature to illustrate semantic relationships between ontology entities. Therefore the name of a semantic relation is shown as curved text inside the graphical representation of the entity. In the case of multiple relations between two entities it is conceivable to display these relations with multiple curved texts side by side or one upon each other. However if the number of the relations becomes larger, the space inside of the graphical representation of an entity may not be sufficient. In this case the following possibilities could be considered: It is conceivable to visualize the names of multiple relations as a tooltip or in an external text field. Another possibility could be the visualization of relations with different colors and to integrate a legend for their meaning. Filtering methods to select relations of interest are also conceivable.

Currently we conduct several evaluations to enhance the usage of the SemaSun visualization. Particularly our evaluations are focused on the automatic adaption of the SemaSun visualization to the needs of the user for example the automatic expansion of specific paths of the ontology's hierarchy based on user's behavior analysis. However the first usage tests in the THESEUS Use Cases regarding the user acceptance show great promises for this kind of new ontology visualization.

Conclusion

In this paper we presented a novel approach of ontology visualization which is based on the sunburst visualization metaphor. We improved this visualization metaphor, which is naturally designed for displaying hierarchical data, to the tasks of displaying multiple inheritance and semantic relations. Thereby the sunburst visualization is capable for displaying ontologies without information loss. To reduce the cognitive overload of the users we integrated incremental ontology exploration, so users are able to focus on entities of interest and to request information on demand. The radial layout of the sunburst visualization offers thereby the expansion of multiple paths and maintains the context while the user navigates through the knowledge space.

References

Alani, H. (2003). *TGVizTab: An Ontology Visualization Extension for Protégé* Knowledge Capture (K-Cap'03), Workshop on Visualization Information in Knowledge Engineering.

Baehrecke, E., Dang, N., Babaria, K. and Shneiderman, B. (2004). *Visualization and analysis of microarray and gene ontology data with treemaps.* BMC Bioinformatics, 5, 1-12

Bhatti, N. (2008). *Web Based Semantic Visualization to Explore Knowledge Spaces - An Approach for Learning by Exploring*. Proceedings of ED-Media 2008, 312-317

Bhatti, N. & Weber, S. H. (2009). *Semantic Visualization to Support Knowledge Discovery in Multi-Relational Service Communities*. Handbook of Research on Social Dimensions of Semantic Technologies and Web Services, 23

Bosca, A.; Bonino, D. & Pellegrino, P. (2005). *OntoSphere: more than a 3D ontology visualization tool*. Proceedings of SWAP 2005, the 2nd Italian Semantic Web Workshop, CEUR Workshop Proceedings

Eklund, P., Roberts, N. and Green, S. (2002). *OntoRama: Browsing RDF ontologies using a hyperbolic-style* browser. Cyber Worlds, 2002. Proceedings First International Symposium on, 405-411

Gruber, T. (1993). *A Translation Approach to Portable Ontology Specifications*. Knowledge Acquisition, Vol. 5, Issue 2, 199–220

Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C. and Giannopoulou, E. (2007). *Ontology Visualization Methods - A Survey*. ACM Computing Surveys (39) 10-43

Katifori, A., Torou, E., Halatsis, C., Lepouras, G. and Vassilakis, C. (2006). *A Comparative Study of Four Ontology Visualization Techniques in Protégé: Experiment Setup and Preliminary Results*. IV '06: Proceedings of the conference on Information Visualization, IEEE Computer Society, 417-423

Kerrigan, M. (2006). *WSMOViz: An Ontology Visualization Approach for WSMO*. Information Visualization, IV 2006. Tenth International Conference on, 411-416

Kriglstein, S. and Motschnig-Pitrik, R.(2008). *Knoocks: New Visualization Approach for Ontologies*. Proceedings of the 2008 12th International Conference Information Visualization, IEEE Computer Society.

Liebig, T., Noppens, O. (2004). *OntoTrack: Combining Browsing and Editing with Reasoning and Explaining for OWL Lite Ontologies*. Proceedings of the 3rd International Semantic Web Conference (ISWC2004), Springer, 244-257

Nazemi, K., Breyer, M. and Hornung, C. (2009).*SeMap: A Concept for the Visualization of Semantics as Maps* UAHCI '09: Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Part III, Springer, 83-91

Noy, N. F., Fergerson, R. W. and Musen, M. A. (2000). *The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility*. EKAW '00: Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management, Springer, 17-32

Parsia, B., Wang, T., and Golbeck, J. (2005). *Visualizing web ontologies with CropCircles*. End User Semantic Web Interaction WS @ ISWC2005

Pietriga, E. (2004). *IsaViz: A Visual Authoring Tool for RDF*. http://www.w3.org/2001/11/IsaViz

Plaisant, C., Grosjean, J., and Bederson, B. (2002). *SpaceTree: supporting exploration in large node link tree, design evolution and empirical evaluation*. Information Visualization. INFOVIS 2002. IEEE Symposium, 57-64

Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. ACM Trans. Graph., ACM, 11, 92-99

Shneiderman, B. (1996). The *Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. Proceedings of the 1996 IEEE Symposium on Visual Languages, IEEE Computer Society, 336-343

Sintek, M. (2003). *OntoViz tab: Visualizing Protégé ontologies*. http://protegewiki.stanford.edu/index.php/OntoViz

Storey, M., Musen, M., Silva, J., Best, C., Ernst, N., Fergerson, R. and Noy, N. (2001). *Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protégé*. Workshop on Interactive Tools for Knowledge Capture (K-CAP-2001)

Studer, R., Davies, J., Warren, P. (2006). *Semantic Web Technologies – Trends and Research in Ontology-Based Systems*. John Wiley & Sons.

Theseus Program Home Page 2009: http://theseus-programm.de

Van Ham, F. and van Wijk, J.(2002) *Beamtrees: compact visualization of large hierarchies*. Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on, 93-100