

# Visualization and Protégé

Margaret-Anne Storey  
[mstorey@uvic.ca](mailto:mstorey@uvic.ca)

Robert Lintern  
[rlintern@uvic.ca](mailto:rlintern@uvic.ca)

Neil Ernst  
[nernst@uvic.ca](mailto:nernst@uvic.ca)

David Perrin  
[dperrin@uvic.ca](mailto:dperrin@uvic.ca)

University of Victoria,  
Victoria, BC, Canada

## Abstract

*In this talk we will explore the use of visualization as a cognitive aid for managing ontologies and knowledge representations. We discuss how we have integrated two interactive visualization tools (Jambalaya and PROMPTViz) with the Protégé system. First, we present Jambalaya, a plug-in developed to help users visualize, navigate, and ultimately understand an ontology. Secondly, we present the PROMPTViz tool which adds visualization support to the PROMPT plug-in for Protégé. PROMPTViz, using the differencing algorithms of PROMPT, graphically displays differences between versions of the same ontology. We describe our approach to developing these tools, give details on some preliminary user feedback, and propose some future directions for each tool.*

## 1. Introduction

Understanding and maintaining the structure of large knowledge bases has been a problem since the first knowledge based systems, such as rule-based expert systems, were created. Many ontologies contain very complex structures that need to be viewed at various levels of abstraction and in different contexts to be clearly understood. Consequently, knowledge modelling can be a cognitively challenging task, as the user has to clearly understand the ontology at a high level as well as manage and understand detailed information.

To enable the reuse of ontologies, analysts will often have to explore and understand ontologies they did not create, to decide how or when to use the existing ontology. Clearly, for large, complex ontologies, this is not a trivial task.

Support is also needed to enable the evolution of ontologies. A set of tasks that our research is focused on concerns how modelers of the ontology, i.e., the NCI staff, understand how the ontology is changing. In such a large and complex model, it is difficult for individuals to understand what changes from day to day.

Although visual formalisms have been used for presenting knowledge representations for decades (see [11] for a discussion on conceptual graphs and semantic networks, and [5, 7] for a discussion on concept maps), the power of computer aided information visualization has not been fully explored in this area.

What is missing in the knowledge representation community is a tool that is highly customizable and interactive, so that diverse users with varying goals can customize how they visualize ontologies from many domains. Moreover, the existing approaches are limited in their power to help a user visualize the multi-dimensional concept spaces that are typical of those found in ontologies.

Over the past few years we have been developing and integrating a sophisticated interactive visualization tool called Jambalaya[12] with Protégé[9]. We created the Jambalaya plug-

in by customizing and extending the SHriMP Views general-purpose visualization framework developed by our group at the University of Victoria [13]. SHriMP includes generic visualization techniques and navigation facilities to make better use of people's cognitive abilities when using a screen of limited size.

In parallel to the Jambalaya tool development effort, we have been doing research in how visualization support can be used to support the task of comparing ontologies. Our PROMPTViz tool adds advanced visualization to the PROMPT[8] plug-in for Protégé.

The advanced visualization techniques found in our tools, such as animated zooming and panning, sophisticated layout algorithms, and advanced filters, may be needed to help users browse and understand these large, complex multi-dimensional spaces. However, more empirical work is needed in this area as there has been insufficient research effort spent on trying to understand the requirements for visualizing ontologies.

## 2. Approach

The recent additions and interest of Protégé users of various new visualization tools, including TGViz [1], OntoViz [10], Jambalaya and PROMPTViz, indicates that there is a perceived need for visualization support in Protégé. However, the existence of the tools are not sufficient to tell us why they were created or which information understanding or navigation problems they are used to help solve. The following two questions particularly interest us:

- Which tasks could benefit from visualization support?
- What visualization techniques are most appropriate for these tasks?

In order to help us answer these questions we have so far carried out the following activities:

- a detailed review of background research in knowledge engineering that focuses on the human computer interaction aspects of knowledge engineering tools;
- a survey posted on the Protégé mailing list to elicit feedback on visualization needs by the Protégé community;
- contextual inquiries and interviews of knowledge engineers and domain experts in the field; and
- a formal user study of PROMPTViz.

In addition to these activities, we have also had many experiences using Protégé and Jambalaya as we have created and maintained ontologies for other projects using these tools. Consequently, we have significant insight into how a visualization tool can be useful as well as understanding some of the challenges that designing such a tool can pose. Our biggest design challenge has been satisfying the diverse set of users that may want to use the tool and the varying requests for different

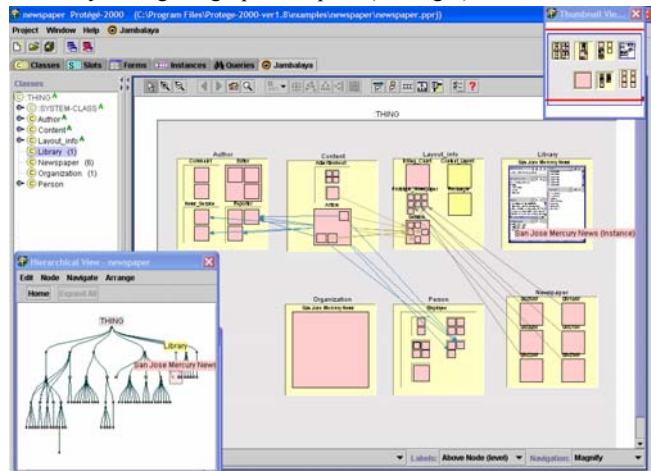
features. The biggest usability challenge appears to be learning how to effectively use the many features that Jambalaya offers.

In order to gain other perspectives, we designed a survey and deployed it using the Protégé mailing list. See [4] for a copy of the full survey. In short, the domains for which ontologies are developed are very broad and the size of ontologies being edited varies greatly as well. The respondents indicated that they believe visualization could help with a variety of tasks

User observations and contextual inquiry techniques have revealed further insights concerning how knowledge engineering is conducted in the field. For example, domain experts and modellers frequently desired the ability to see what effect their addition of a new concept might have on the assertions the model was making; the standard indented collapsible list did not provide sufficient context for them to understand this, however.

### 3. Jambalaya

First we describe the Jambalaya tool, a visualization plug-in for Protégé. We present concepts and relationships in Jambalaya using the graph metaphor (see Fig.1).



**Figure 1: The Jambalaya plug-in for Protégé. The newspaper ontology is displayed using a tree view (lower left) and a nested view (center right).**

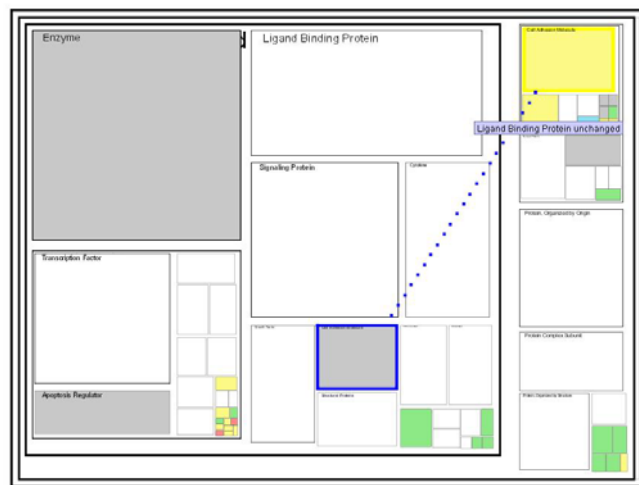
Classes and instances are represented as nodes in a graph; different types may be distinguished using different color hues. Directed edges (arcs) are used to show relationships between concepts and instances, such as is-a relationships between classes in the concept hierarchy, instance-of relationships relating instances to classes, and slot relationships between classes and instances.

In addition to using arcs to show relationships as connections (arcs) between classes and instances, the user may show a relationship type using containment. For example, rather than drawing arcs between nodes to show the is-a relationship, we can instead nest subclasses within their superclasses. Furthermore, it may also be advantageous to use other user defined slot types for nesting nodes. For example, in an anatomy ontology, the user defined part-of relationship may be a more appropriate relationship for nesting nodes than the is-a relationship.

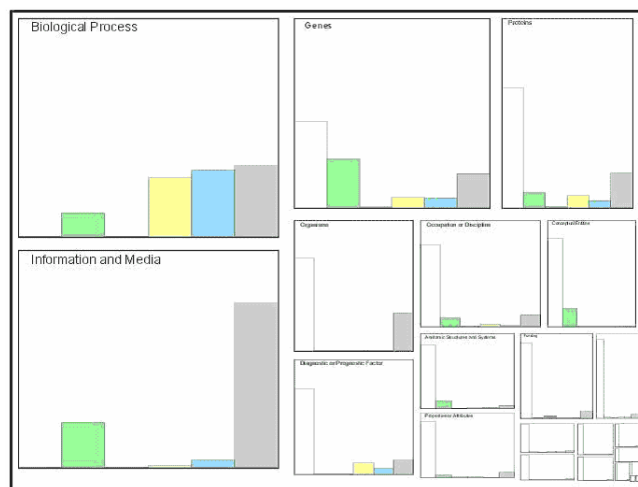
### 4. PROMPTViz

PROMPTViz builds on the diff functionality of the

PROMPT tab for Protégé by leveraging the difference table generated by PROMPT to create a treemap display showing a merge of the two versions of the ontology (see Fig. 2 and Fig. 3). Our new view aims to address the perceived deficiency of PROMPT in providing an overview of the changes it has discovered.



**Figure 2: PROMPTViz in solid colouring mode, with a dotted arc showing the reclassification of a concept. The yellow node represents the old position of the concept.**



**Figure 3: PROMPTViz in histogram colouring mode, with the treemap sized so classes with the most number of changed descendants are drawn largest.**

The treemap layout generates a space-constrained visualization of hierarchical structures and is very effective in showing attributes of leaf nodes using size and color coding.” (<http://www.cs.umd.edu/hcil/treemap/>) The treemap layout, combined with a zoomable user interface (ZUI) [3], presents an initial overview that allows users to drill in and discover more details about the changes. The space occupied by a node in a treemap can be based on any numerical value. For example, the nodes could be sized by the number of descendants they have or by the number of changes.

An additional benefit of the treemap layout is that we can paint the nodes to provide additional information about the changes to the node or its descendants. PROMPTViz does this

in two ways. One is to paint the node a solid colour (see Fig. 2) representing the type of change that has happened to the class; the other way is to paint nodes with a histogram (see Fig. 3), where each bar represents the percentage of descendents classified as each change type, an idea inspired by Baker and Eick's SeeSys system [2].

PROMPTViz also has the ability to draw arcs to show relationships between concepts and to depict the movement of concepts within the is-a hierarchy. For example, in Fig. 2, the dotted arc depicts the change of superclass for Cell Adhesion Molecule from Membrane Protein to Protein Organized by Function.

## 5. User Evaluation

### 5.1 Evaluating Jambalaya's adoption

Although Jambalaya sees on average 3-4 downloads per day from our website, we were interested in understanding how many of these downloads were from people who actually adopted the tool in daily use. From informal and formal studies, we knew the power and complexity of Jambalaya – its wide array of tools – was also a major stumbling block for users. Furthermore, from our survey we understood that ontology modelling tasks vary from domain to domain. We therefore undertook to construct and evaluate a customizable version of Jambalaya to evaluate how the ability to select and enhance the tool's features would affect the use of the tool. A qualitative analysis was conducted using five expert users from outside the group. The analysis concluded that the formal customization model used was too complex, and that an informal model, which placed more burden on the developers to understand what cognitive support was required, would be preferable.

### 5.2 PROMPTViz User Evaluation

Our evaluation of PROMPTViz was conducted to help focus our continued development effort, to refine the useful features, and to re-evaluate those that were not useful. To this end, we recruited 8 graduate students from the faculty of computer science at the University of Victoria to participate in the study.

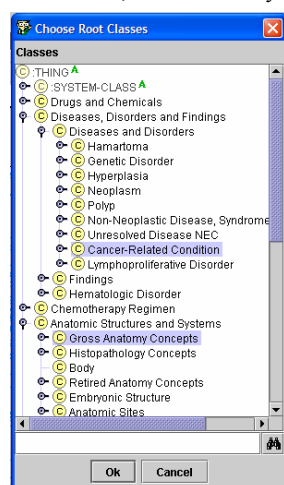
Although the number of participants in this exploratory study is too small for us to make any statistical proofs, it was reassuring to have all the users confirm that the addition of PROMPTViz made answering questions about the general differences between two versions of a large ontology easier than could be done with PROMPT by itself.

The most significant usability problems found by users were the lack of tooltips and labels in the histogram mode, and the general confusion surrounding the interpretation of arcs.

Overall, the user study suggests that the cognitive support PROMPTViz adds to PROMPT is important to the users we evaluated.

## 6. A Case Study

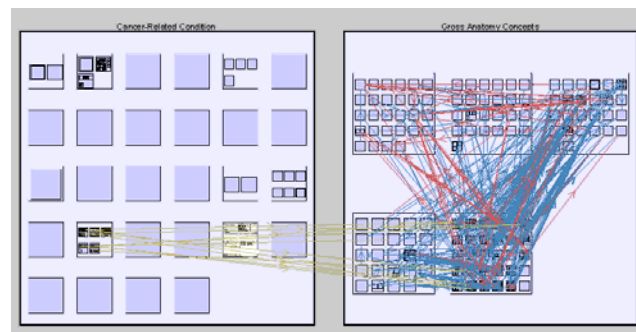
In our presentation, we will give a detailed demonstration of how Jambalaya and PROMPTViz can be applied to the NCI Thesaurus. We will highlight some of the challenges we face in developing these



**Figure 4: Selecting a subset of the NCI Thesaurus to visualize.**

advanced user interfaces. For example, we will discuss how we deal with issues of scale. Figures 4 and 5 show how we can use the concept of “working sets” to reduce the amount of information being visualized.

We will also demonstrate how we can make use of customization and end user programming facilities to improve how the NCI Thesaurus is visualized based on our knowledge about that domain.



**Figure 5: This Jambalaya view shows the selected information based on the working set selected in Fig.4.**

## 7. Ongoing and Future Work

We have explored the use of visualization as a cognitive aid for managing ontologies and knowledge representations. We described the Jambalaya and PROMPTViz visualization tools and how they have been integrated with Protégé.

We make an important observation that there is no “single view” that will appeal to all the users and tasks requiring visualization. Therefore, many different ways to customize and extend the visualizations are needed. Such mechanisms need to be easy to use and should map to the “world view” of the user in question. We are working towards extending the customization model introduced in Section 5.1 to be more intuitive and reduce the burden on the user of learning the many features of Jambalaya.

We have more recently been investigating the use of the PROMPTViz interface in conjunction with the Protégé OWL plug-in. This combination allows users to visualize the results of classifying OWL ontologies using a description logic classifier such as RACER [6]. Future work includes making improvements to PROMPTViz that will match better with the unique needs of ontology classification. For example, visualization could help in understanding not only which concepts were reclassified or newly inferred, but also why such changes occurred. An interface which animated selected concepts as they move to new locations, for example, would be a powerful way of debugging the results of unexpected or misunderstood classifications.

We recognize that there is the need for more empirical work, and our efforts are ongoing. Our tools should be refined and updated further as more is learned about visualization and knowledge engineering tasks.

## 8. Discussion

At the end of our presentation, we will lead a discussion on how visualization can be helpful (or not) for representing and managing ontologies. Moreover, we also need to consider non-functional visualization requirements such as scalability.

## 9. Acknowledgements

PROMPTViz and Jambalaya are supported by a grant from the National Cancer Institute's Center for Bioinformatics.

The Protégé resource is supported, in part, by grant P41 LM007885 from the National Library of Medicine. We acknowledge the helpful and effective support from the Stanford team during the development of our tools.

Nasir Rather and other members of the CHISEL group have also been instrumental in developing these visualization tools.

## 10. References

1. Alani, H., TGVizTab: An Ontology Visualisation Extension for Protege. in *Knowledge Capture 03 - Workshop on Visualizing Information in Knowledge Engineering*, (Sanibel Island, FL, 2003), ACM, 2-7.
2. Baker, M.J., Stephen G. Eick Space-filling Software Visualization. *Journal of Visual Languages and Computing*, 6 (2). 119-133.
3. Bederson, B.B. and Hollan, J.D., Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. in *UIST'94, ACM Symposium on User Interface Software and Technology*, (Marina del Rey, CA, 1994), 17-26.
4. Ernst, N.A. and Storey, M.-A. A Preliminary Analysis of Visualization Requirements in Knowledge Engineering Tools, University of Victoria, Victoria, 2003, 6.
5. Gaines, B. and Shaw, M.L.G. Concept Maps as Hypermedia Components. *Int. Journal of Human-Computer Studies: Special Issue on Knowledge-Based Hypermedia*, 43 (3). 323-361.
6. Haarslev, V. and Moeller, R., High Performance Reasoning with Very Large Knowledge Bases: A Practical Case Study. in *Int. Jt. Conf. on Artificial Intelligence*, (2001).
7. Kremer, R., Visual Languages for Knowledge Representation. in *11th Workshop on Knowledge Acquisition, Modelling, and Management (KAW '98)*, (Banff, Alberta, 1998).
8. Noy, N.F. and Musen, M.A. The PROMPT Suite: Interactive Tools For Ontology Merging and Mapping. *International Journal of Human-Computer Studies*, 59 (6). 983-1024.
9. Protege-2000, The Protege-2000 website, Stanford Medical Informatics, <http://protege.stanford.edu>
10. Sintek, M., OntoViz Tab: Visualizing Protege Ontologies, <http://protege.stanford.edu/plugins/ontoviz/ontoviz.html>
11. Sowa, J.F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole, Pacific Grove, CA, 2000.
12. Storey, M.-A.D., Musen, M.A., Silva, J., Best, C., Ernst, N., Ferguson, R. and Noy, N.F., Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protege. in *Workshop on Interactive Tools for Knowledge Capture, K-CAP-2001*, (Victoria, B.C. Canada, 2001).
13. Storey, M.-A.D., Wong, K., Fracchia, F. and Müller, H., On Integrating Visualization Techniques for Effective Software Exploration. in *InfoVis '97*, (Phoenix, AZ, 1997), 38-45.