Linking Representation with Meaning

D.J. Duke*

School of Computing, University of Leeds, UK

Abstract

The purpose of visualization is not just to depict data, but to gain or present insight into the domain represented in data. However in visualization systems, this link between features in the data and the meaning of those features is often missing or implicit. It is assumed that the user, through looking at the output, will close the loop between representation and insight. An alternative is to view visualization tools as interfaces between data and insight, and to enrich this interface with capabilities linked to users' conceptual models of the data. Preliminary work has been carried out to develop such an interface as a modular component that can be installed in a pipelined architecture. This poster expands the motivation for this work, and describes the initial implementation carried out within the Visualization Toolkit (VTK).

CR Categories: I.3.8 [Computer Graphics]: Applications; I.3.m [Computer Graphics]: Miscellaneous;

Keywords: Ontology, Visualization, Pipeline, Database

1 Motivation

Forty years ago Hamming suggested that the purpose of computing is "insight, not numbers". At a time when graphics has become a commodity, it is important to remember that the purpose of visualization is insight, not pictures. This holds whether the visualization is performed for presentation, analysis or exploration [1].

Our understanding of insight comes in part through work in cognitive science on theories of categorization and inference. Theories are intended to explain observations relative to some given background knowledge; revisions may be made in the light of new observation, or to accommodate an explanation that is 'better' in some accepted sense [3]. Fundamental to any theory is an ontology; this provides the vocabulary for talking about observations, the criteria by which a given phenomenon may be assigned to a category, and the organization of categories within some form of abstraction hierarchy. Revision may involve the formation of new categories to capture newly observed distinctions, an activity nicely articulated in work on geographic visualization [4, 2]. This poster reports initial work on making explicit the link between data depiction and the ontology on which that depiction is grounded. The long-term aim is to provide a better basis for interacting with visual representation, and for the insight from machine and human analyses of data.

2 Ontologies in Visualization

Explicit links between categories and representation can be found in domain-specific visualization tools; for example, molecular viewers such as RasMol [5] allow the user to choose views that have some relation to concepts in the domain (protein structure): ball-and-stick (reflecting atoms and bonds), space-filling (contact surfaces), and sheet-and-helix (tertiary structure) to name but three. Figure 1 shows protein visualization operating at these three different levels of categorical abstraction. Category formation is also explicitly

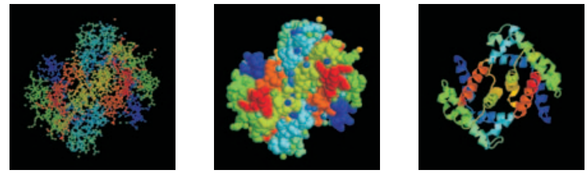

Figure 1: Multiple representations of Haemoglobin

supported through clustering techniques, well known applications being in multivariate analysis and graph visualization. The assumption is that statistical regularities within the data are meaningful in terms of structure within the underlying domain. However, while such categories are sometimes identified explicitly through 'best exemplars' drawn from the data, closing the loop from hypothesized category back to domain knowledge is not supported within most visualization systems.

3 Modular Support for Ontologies

Modular visualization tools offer flexibility in exploring and adapting representations for data; they support the reuse of generic algorithms, and provide for novel combinations of technique. However, support for linking ontology and representation is limited:

- modules are available that 'hard wire' a model of the data, for example a protein structure representation that provides access to specific levels of structure information; or
- modules that provide an abstract analytic capability, for example various forms of clustering over multivariate data.

This work sets out to explore how ontology might be introduced into a modular, pipelined visualization tool through a filter that relates the data in a pipeline to a explicit model of the domain from which the data is derived. By making the connection explicit, it should be possible to:

1. interact with categories as objects, i.e. to provide operations linked to the categories in the ontology rather than the low-level geometric structures representing those categories;
2. adapt and blend depiction style; different parts of the data may be depicted by different kinds of representation, depending on the focus and level of abstraction at which the user is working.
3. modify the mapping between ontology and data, supporting simple conjectures about data.

Figure 2 shows the conceptual organization of a pipeline with ontology support. A component called a 'compositor' takes a dataset as input, and distributes sections of that dataset to parts of the pipeline responsible for processing particular categories of representation. Output from these pipelines is re-integrated in the display. Operation of the compositor is driven by an ontology and an internal mapping between categories and data set structure. The mapping between dataset and ontology can be constructed in two ways:

*e-mail:djd@comp.leeds.ac.uk

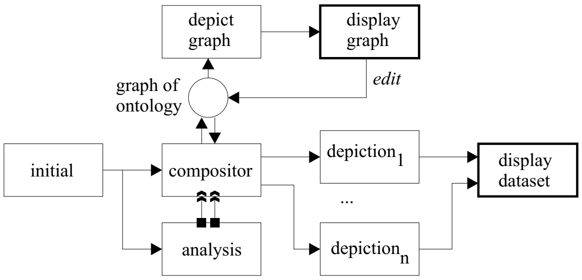

Figure 2: Ontology in the pipeline

1. through a module that analyses the data and uses the interface of the compositor to construct an ontology; or
2. by direct manipulation of the displayed ontology.

In the work presented here the ontology itself had been viewed as a kind of dataset, specifically as a directed acyclic graph, with nodes representing categories and edges indicating when one category is a subtype of another. This has the advantage that the ontology itself can be visualized using existing general-purpose filters. In the longer term, the model of ontology would need extending to include domain-specific relationships between categories, but this does not present any fundamental problem.

4 Preliminary Work: vtkCompositor

An implementation of a compositor has been developed as an extension to VTK [7]. Exploration of this approach has been done in the context of graph visualization using a library developed by the author [8]; at present, modules for building an ontology just implement a simple form of graph clustering. Figure 3 shows the contents of two windows in a prototype application. On the left is the dataset, in this case a graph obtained from a finite state protocol simulation. The window on the right is intended to show the category hierarchy depicted in the data. Here, the categories have been induced by a simple graph clustering technique, resulting in a flat hierarchy with around 20 constituents. Edges of the FSM graph that connect nodes within a common category share the same colour; edges whose start and end nodes lie in different parts of the ontology are in gray.

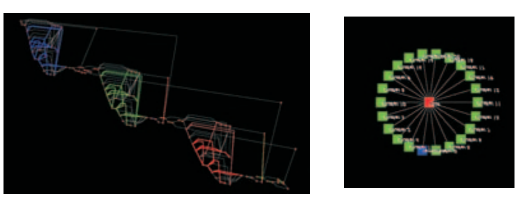

Figure 3: Graph and Ontology

Interaction with the dataset and ontology is currently limited to geometric operations; parts of the dataset representing a particular concept in the ontology can be manipulated either directly via the view of the dataset, or via the glyph representing that concept in the ontology graph window. It is also possible to enable/disable the display of selected concepts. Figure 4 shows the contents of the graph window after particular regions in the graph have been pulled out; by dragging and superimposing these subgraphs it is possible to confirm for example that they are structurally isomorphic. This suggests that these sections of the state machine perform a similar function. This function could be captured by a generic category (let us say “recovery” for the sake of example), and then the instances could then be re-organized as subtypes within the ontology.

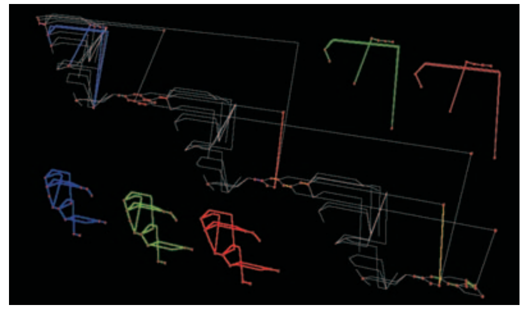

Figure 4: Extracting Components of the Graph

Functionality to edit the ontology graph is not yet available but may be provided by a generic graph editor being developed as an MSc project at the University of Leeds.

5 Conclusion

At some point in the visualization process, attention must be paid to the meaning of the data, wheher it is in the assumptions encoded into domain-specific applications, or the interpretation of an expert observing the output. The work reported here is intended to support a systematic approach to linking data with an explicit model of its meaning, allowing the user to inspect, interact with, and ultimately test the assumptions on which the depiction of the data is grounded. Such an approach is timely for two reasons. First, it reflects similar concerns in other communities, in particular the separation between data and meaning that underpins the semantic web. Second, visualization is not the only way of obtaining insight into data; there has been much interest in combining human and machine driven analyses of data into what Thomas [6] has termed ‘visual analytics’. One step in this direction is to make explicit the models that both human and software agent might be using to express the analysis of that data, so the understanding can be built on common ground.

Acknowledgement

This work is supported by the UK EPSRC through an Advanced Research Fellowship, AF/001275.

References

[1] D. Bergeron. Visualization reference models (panel position statement). In G.M. Nielson and D. Bergeron, editors, *Proceedings of IEEE Visualization*. IEEE Computer Society Press, 1993.

[2] M. Gahegan, X. Dai, J. Macgill, and S. Oswal. From concepts to data and back again: Connecting mental spaces with data and analysis methods. In *Proceedings of the 7th International Conference on GeoComputation*. GeoComputation CD-ROM, 2003. http://www.geocomputation.org/2003/.

[3] J.R. Josephson and S.G. Josephson. *Abductive Inference*. Cambridge University Press, 1994.

[4] A.M. MacEachren. *How Maps Work*. The Guildford Press, 1995.

[5] RasMol Home Page. http://openrasmol.org, 2004.

[6] J. Thomas. New frontiers for visualization. In *Joint Eurographics-IEEE TCVG Symposium on Data Visualization*. Eurographics Association, 2004.

[7] Visualization Toolkit. www.vtk.org, 2001.

[8] VTK Graph Library. www.comp.leeds.ac.uk/djd/graphs/, 2004.