



Applications of Tree-Maps to hierarchical biological data

Patrick McConnell*, Kimberly Johnson and Simon Lin

Duke Bioinformatics Shared Resource, Duke University Medical Center,
Durham NC 27710, USA

Received on January 16, 2002; revised on April 10, 2002; accepted on April 13, 2002

ABSTRACT

Summary: A brief overview of Tree-Maps provides the basis for understanding two new implementations of Tree-Map methods. TreeMapClusterView provides a new way to view microarray gene expression data, and GenePlacer provides a view of gene ontology annotation data. We also discuss the benefits of Tree-Maps to visualize complex hierarchies in functional genomics.

Availability: Java class files are freely available at <http://mendel.mc.duke.edu/bioinformatics/>.

Contact: mcon012@mc.duke.edu

Supplementary information: For more information on TreeMapClusterView (see <http://mendel.mc.duke.edu/bioinformatics/software/boxclusterview/>), and for more information on GenePlacer (see <http://mendel.mc.duke.edu/bioinformatics/software/geneplacer/>).

INTRODUCTION

Hierarchical data is traditionally represented as a directed tree with the root at the top and the tree unfolding beneath with parents pointing to children. Alternative representations take advantage of some intrinsic properties of the tree, such as placing the root at the center and drawing the children around, but all of these suffer from the same problem: as the size and complexity of the tree increase, it becomes impractical to display in limited space. The Tree-Map algorithm was developed to efficiently render large trees consisting of millions of nodes, such as a UNIX file system (Shneiderman, 1992).

Tree-Maps have a simple recursive definition: a parent node is drawn as a box and each of its children are drawn as boxes within it. In this way, a tree, no matter how large, can be rendered in a predefined space. Even when fine structure of underlying nodes cannot be displayed in detail, the size of the node can indicate the number of children it contains.

Since its birth, many other applications of Tree-Maps have been prepared, two of which are visualizing stock portfolios (Jungmeister and Turo, 1992) and visualizing

the analytical hierarchy process in decision-making (Asahi *et al.*, 1995). Its success in these areas suggests that the Tree-Map algorithm could have beneficial applications in functional genomics where large and complex hierarchies exist.

APPLICATIONS

Much of bioinformatics data has a natural hierarchy, and many applications focus on extracting this hierarchy to reveal properties of the data. A notable example is microarray studies where recognition of how genes and experiments cluster provides a better understanding of their biological relatedness. This data is often rendered as a top-down or left-right rooted tree (Eisen *et al.*, 1998). However, the size and complexity of the hierarchy prevents its application to more than a few hundred genes. Therefore, we have developed TreeMapClusterView to render microarray clusters in a Tree-Map. This Java program reads in cluster data conforming to the output of Eisen's Cluster program (Eisen *et al.*, 1998), as well as the corresponding expression data. In addition to nested boxes indicating hierarchy, colors are used to indicate the homogeneity of the node, and the size of a box is proportional to the number of genes beneath it. This visualization allows the user to explore the hierarchy without losing the global view (see <http://mendel.mc.duke.edu/bioinformatics/software/boxclusterview/figure1.jpg>).

The Gene Ontology (GO) is another naturally hierarchical data structure (The Gene Ontology Consortium, 2000), but can be difficult to view in its entirety with the standard tree-rendering methods (see the Gene Ontology Consortium's AmiGO <http://www.godatabase.org/cgi-bin/go.cgi>). To discover if a group of genes share any annotation in the GO, we have developed GenePlacer in Java. The program reads in a set of yeast genes (defined by their ORF), and maps each gene to a set of GO terms. Each node in the Tree-Map is a GO term, and its size is directly correlated to the number of genes mapped to it and its descendants. Thus, larger nodes represent terms that are more represented by the set of genes. This program can be used in two ways. First, we can evaluate a group of genes

*To whom correspondence should be addressed.

identified by statistical methods, and see if they have correlated biological relatedness. Alternatively, we can select a group of genes from the entire microarray gene chip with certain biological functions, such as 'respiration', and then study their expression profiles (see <http://mendel.mc.duke.edu/bioinformatics/software/geneplacer/figure2.jpg>).

CONCLUSIONS

Based on our experience with microarray data and ontological data, we found that the TreeMapClusterView and GenePlacer applications of Tree-Maps provide a better global view of relationships in a hierarchy. Although boxes cannot be infinitely drawn within each other on a computer screen, Tree-Maps can handle this obstacle much more gracefully than traditional tree visualizations because each parent encodes some of its descendants information. So, even when children are clipped from view, their properties remain. Furthermore, zooming capabilities are implemented so that a user can quickly drill down to the leaves. In addition to its hierarchical semantics, color and size can be used to convey further information. Our implementation of Tree-Maps in

TreeMapClusterView and GenePlacer presents a useful alternative technique to explore large and complex hierarchies in functional genomics. With this technique, we can reveal interesting relationships that may otherwise have gone unnoticed.

REFERENCES

- Asahi,T., Turo,D. and Shneiderman,B. (1995) Using treemaps to visualize the analytic hierarchy process. *Information Systems Research*, **6**, 357–375.
- Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
- Jungmeister,W.-A. and Turo,D. (1992) Adapting treemaps to stock portfolio visualization. *University of Maryland, Center for Automation Research technical report*, CAR-TR-648 (also CS-TR-2996, SRC-TR-92-120).
- Shneiderman,B. (1992) Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, **11**, 92–99.
- The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genet.*, **25**, 25–29.