

Concept Relationship Editor: A visual interface to support the creation of relationships between taxonomic classifications

Paul Craig^{*}, Jessie Kennedy[†]

School of Computing, Napier University

ABSTRACT

This paper describes the Concept Relationship Editor, an interactive visualisation tool designed to support the specification of relationships between hierarchical taxonomic classifications. The tool operates using an interactive space-filling adjacency layout which allows users to expand multiple lists of taxa with common parents so they can explore and add relationships between two classifications. Whenever selected lists contain too many items for them to be legible within the restrictions of available screen space the user can alleviate the problem by either operating in 'lens mode' or 'scroll mode'. In 'lens mode' the layout is configured so that both of the classifications and all the relationships are completely visible on-screen. Here a fish-eye lens type distortion effect is applied under the cursor to allow taxa names with less assigned space to be made legible. In 'scroll mode' the layout assigns sufficient space for the labels of all expanded taxa lists to be legible and scroll bars can be used to navigate across the hierarchy of either classification. While the 'lens mode' provides context and allows for more direct comparison of relationships throughout the classifications, 'scroll mode' tends to allow for relationships to be added more efficiently between smaller groups of similarly classified taxa.

Keywords: Interaction, Applications of InfoVis, Hierarchy visualization, Focus+context

1 INTRODUCTION

The science of Linnaean taxonomy classifies specimens according to shared characteristics, into hierarchies of taxonomic concepts (taxa) of differing ranks. A typical classification includes anything from around 200 to 1,500 taxa spread over about four to fifteen ranks. These taxa are given scientific names e.g. *Apium graveolens* L. for the species commonly known as celery. Over time the relative importance of particular characteristics used in differentiating organisms changes, resulting in alternative classifications being proposed. This along with the rules of Linnaean taxonomy which control the naming of taxa has the effect of the same taxonomic name having differing definitions. These definitions are described in field guides used by biologists for identifying organisms.

When comparing data containing the scientific names of organisms recorded by biologists using different field guides there is often ambiguity over what a particular name means; this can cause errors in data analyses [1]. To resolve this problem, an increasingly important task for taxonomists and ecologists is to understand the relationships between alternative taxa so they can be taken into account when comparing data sets. An increasingly common approach being taken to define the relationships between

taxa in alternative classifications is to use a set-based notation. This states the relationship between taxon pairs [2]. These relationships may be defined when a taxonomist is undertaking a new revision of a classification by relating the new taxa to previous taxa or by a third party taxonomists examining existing classifications.

Allowing a taxonomist to visualise the taxonomies in a familiar manner, to explore them and easily create relationships between them is a significant challenge. For the efficient creation of relationships, taxonomists must not only be able to effectively navigate the hierarchical structure of classifications but also be able to have different groups of similarly classified taxa clearly identifiable. The fact that groups of interrelated taxa are often dispersed differently in different classifications makes it necessary to have multiple different groups of taxa from each classification visible at the same time. The layout for individual classifications must also be such that when relationships are added they do not impede the display of taxa names and the information needed to add further relationships.

While there are many effective visualisation techniques for large scale hierarchical data [3-9] and a number of visualisations that also allow the user to explore relationships between or within hierarchies [3, 6, 7, 10], there are none that display different varieties of relationship and allow a user to create relationships. The one visualisation that does display multiple relationship types between hierarchies [11] is primarily designed to identify overlap between hierarchies with the same name. Synonymy relationships are only displayed when they exist between a selected taxon and the limited visible portions of each other hierarchy. When displayed these relationships also have a tendency to occlude taxa names so the interface is not suitable to be adapted for the addition of new relationships.

1.1 Concept Relationship Editor

A screenshot of the Concept Relationship Editor interface for exploring and editing relationships between taxonomic classifications is shown in figure 1. The main component is the classification panel. This has representations of each selected classification on each side of the display. Taxa are represented using labelled rectangles with adjacency used to indicate positions in the hierarchy. Relationships between taxa are represented using curved lines between the related taxa. These have symbols at either end to indicate the relationship type. Relationships not attached to a selected taxon are greyed out and relationships in the process of being formed (as in figure 1) are represented using a red dotted line.

Each classification is colour-coded, with taxa currently selected for focus (together with their ancestors in the hierarchy) coloured a lighter hue than their classification. The overall layout of the classifications attempts to make, where possible, the labelled names of taxa readable. Here priority is given first to selected taxa, then the children of selected taxa, then the siblings of selected taxa and finally the siblings of ancestors of selected taxa. These nodes require priority for navigation and relationship addition. When a user selects a new taxon, priority is

*e-mail: p.craig@napier.ac.uk

†e-mail: j.kennedy@napier.ac.uk

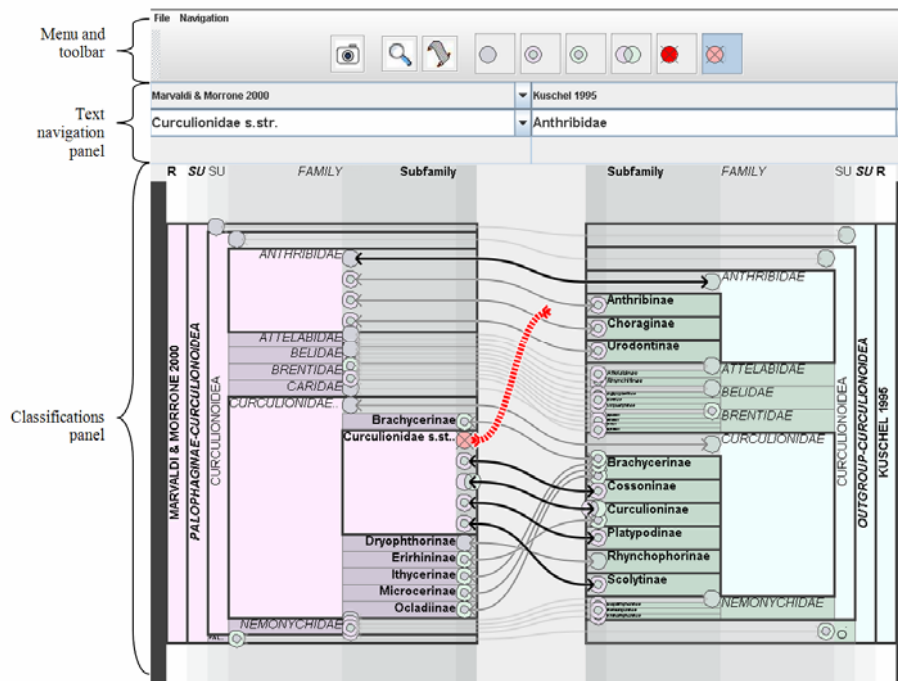


Figure 1. Screenshot of the Concept Relationship Editor interface.

appropriately reassigned and the layout changes. During the transition, animation is used to provide visual cues and aid reinterpretation.

When the layout algorithm cannot assign sufficient space to adequately display the labels of the prioritized taxa, the lens mode or scroll mode take effect to make taxon labels otherwise accessible. In lens mode a fish-eye lens [12] type effect is applied to the vertical axis under the mouse cursor. Here, the magnification is calculated to be just enough to expand the taxon directly under the cursor and make its label readable. Alternatively, in scroll mode, priority nodes for navigation are all sized above a threshold for readability and the user may be required to operate a scroll-bar to determine which taxa occupy the screen space. Whenever possible, scrolling is automated pre-empt the users actions .

2 CONCLUSION

We have developed an information visualisation technique for exploring and editing relationships between taxonomic classifications. The technique applies a novel focus + context adjacency layout for hierarchies that allows multiple groups of nodes to be expanded with their labels readable in horizontally stacked lists. The technique can be differentiated from other visualisations in that it allows users to edit and view different types of relationships between two hierarchies. During an informal evaluation of the interface users found it easy to navigate the hierarchy of classifications and add relationships between taxa.

REFERENCES

- [1] Kukla R and Paterson T. Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. . 2nd International Workshop on Data Integration in the Life Sciences 2005 (San Diego, USA); 80-95.
- [2] Berendsohn WG. The concept of "potential taxa" in databases. *Taxon* 1995; 22: 207-212
- [3] Jeong C-S and Pang A. Reconfigurable Disc Trees for Visualizing Large Hierarchical Information Space. *IEEE InfoVis '98 1998* (Research Triangle, North Carolina, USA), Computer Society Press; 19-25.
- [4] Johnson B and Shneiderman B. Treemaps: A Space-Filling approach to the visualization of hierarchical information structures. *IEEE Visualization '91 1991* (San Diego, California, USA), IEEE Computer Society Press; 284-291.
- [5] Lamping J, Rao R, and Pirolli P. A focus + context technique based on hyperbolic geometry for visualizing large hierarchies. *ACM CHI '95 1995* (Denver, Colorado, USA), ACM Press; 401-408.
- [6] Munzner T. Exploring Large Graphs in 3D Hyperbolic Space. *IEEE Computer Graphics & Applications 1998*; 18(4): 18-23
- [7] Neumann P, Schlechtweg S, and Carpendale S. ArcTrees: Visualizing Relations in Hierarchical Data. *EUROVIS 2005: Eurographics / IEEE VGTC Symposium on Visualization 2005*; 53-60.
- [8] Sifer M. Filter co-ordinations for exploring multi-dimensional data. *Journal of Visual Languages & Computing 2005*; 17(2): 107-125
- [9] Wang W, Wang H, Dai G, and Wang H. Visualization of large hierarchical data by circle packing. *SIGCHI conference on Human Factors in computing systems 2006* (Montréal, Québec, Canada), ACM Press; 517-250.
- [10] Holten D. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. *IEEE Transactions on Visualization and Computer Graphics (TVCG; Proceedings of Vis / InfoVis 2006) 2006*; 20(5): 741 - 748
- [11] Graham M and Kennedy J. Extending taxonomic visualisation to incorporate synonymy and structural markers. *Information Visualization 2005*; 4(3): 206-223
- [12] Bederson BB. Fisheye Menus. *ACM UIST 2000 2000* (San Diego, California, USA), ACM Press; 217-226.