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

Paul Craig[a] and Jessie Kennedy[a*]

[a] Napier University, School of Computing, Edinburgh, UK

ABSTRACT

An increasingly common approach being taken by taxonomists to define the relationships between taxa in alternative hierarchical classifications is to use a set-based notation which states relationship between two taxa from alternative classifications. Textual recording of these relationships is cumbersome and difficult for taxonomists to manage. While text based GUI tools are beginning to appear which ease the process, these have several limitations. Interactive visual tools offer greater potential to allow taxonomists to explore the taxa in these hierarchies and specify such relationships.

This paper describes the Concept Relationship Editor, an interactive visualisation tool designed to support the assertion of relationships between 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: Information Visualisation, Hierarchy Visualization, Focus + Context

1. INTRODUCTION

The science of Linnean 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 Linnean 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[2-4]. 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 assert relationships between them is a significant challenge. For the efficient assertion 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[5-11] and a number of visualisations that also allow the user to explore relationships between or within hierarchies[5, 8, 9, 12, 13] there are none that display different varieties of relationship and allow a user to assert relationships. The one visualisation that does display multiple relationship types between hierarchies[14] 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.

2. REQUIREMENTS ANALYSIS

Early requirements analysis was performed with taxonomists working on the Science Environment for Ecological Knowledge (SEEK) project[†] and already involved, or likely to become involved, in the process of adding relationships between existing classifications. These taxonomists were familiar with the use of interactive visual interfaces to explore taxonomic data but when faced with the task of adding relationships between classifications they were compelled to resort to more traditional methods: either manually editing the cells of a spreadsheet to edit relationships or using a text based GUI tool with no representation of the classification hierarchy[15]. The taxonomists agreed that a more efficient way to edit relationships was needed and saw the development of an interactive visual tool as a likely solution. The five main requirements resulting from requirements analysis are summarised as follows;

1. Taxa must be clearly ordered by rank in the visualisation.

2. Navigation of the classification must be simple and intuitive.

3. Multiple groups of similarly classified taxa (i.e. sibling taxa with the same parent) must be easily identifiable at the same time in either of two classifications. Taxa can be identified by their name, rank and position in the hierarchy.

4. Minimal interaction should be required to add and edit relationships.

5. Existing relationships that might affect the addition of new relationships should be clearly identifiable and their representation should not hinder the identification of taxa with new relationships pending.

The first of these requirements, to have taxa clearly ordered by rank, is related to the essential structure of the data. Taxonomic classification data is organized into strict ranks such as family, genus, species etc. These ranks form a discrete ordered system that can be applied to all biological taxonomies so that any two taxa sharing the same rank are said to reside at the same level. This is regardless of their placement and absolute depth in the hierarchy. User testing of an earlier system showed that if the structure of ranks was not preserved in a visualization of taxonomic data a taxonomist would find it very difficult to use[16]. Taxonomies do not, however, use ranks consistently and taxa often have children at different ranks even within the same hierarchy. Where this would cause a visualisation of taxonomic data to have redundant screen space in order to effectively communicate the rank structure of the data, this is preferable to destroying that structure and conserving screen space.

The second user requirement is for simple and intuitive navigation. Here the taxonomists required not only to be able to navigate using the hierarchical structure of the classification but also to be able to search for taxa by name. Given the size and variety of classifications and the fact that the taxonomist may be relating taxa to less familiar classifications, it was judged that this might be the most efficient way to locate some taxa. For navigating the hierarchical structure of classifications, the order of ranks is fixed and the convention is for sibling taxa to be ordered alphabetically by their scientific names.

[†] http://seek.ecoinformatics.org/

The third requirement is for multiple groups of similarly classified taxa to be identifiable in either of two classifications at the same time. The taxonomists involved in requirements analysis felt they would be likely to focus on one taxonomy and relate that taxonomy to only one other taxonomy at a time. The need to have more than one group of similarly classified taxa identifiable at the same time can be explained by dissecting the likely process of adding a group of relationships. First the user will select a group of similarly classified taxa within a hierarchy. Next they will step through these taxa relating them to taxa in another hierarchy. The majority of these taxa are likely to be similarly classified and require a group to be selected in the second hierarchy. If, however, as will often be the case, the related taxa are not all similarly classified the user will have to select two or more groups in the second hierarchy in order to add relationships.

If the interface from which relationships are added does not allow more than one group of taxa to be selected at the same time the user is likely to have to switch between groups while adding relationships. For larger hierarchies this may require a significant amount of additional, otherwise unnecessary, interaction to navigate the hierarchy. If taxon identities within groups are not immediately apparent (i.e. with their names visible) then the workload of the user is increased further as interaction is required to identify each related taxon before a relationship is added. To increase the accessibility of groups of taxa further, the users also wanted to have taxa names vertically stacked so they could be scanned efficiently.

The fourth user requirement is that relationship addition should entail minimal interaction. Here, the users requested that once the relevant taxon groupings are selected, adding relationships between any two taxa should be a simple drag-and-drop action. This would allow rapid efficient addition of a number of relationships that are already obvious to the user (perhaps specifying the congruence of taxa with the same name).

The fifth requirement relates to how relationships are displayed stating that existing relationships should remain visible so the user does not attempt to duplicate work by adding relationships that have already been added. As the synonymy relationships between taxa come in a number of varieties (congruence, containment, overlap etc) it is important to identify the relationship type as well as the taxa involved. The assertion that relationship representations should not hinder the identification of taxa reasserts requirement three which states that groups of taxa should be identifiable (those being the groups of taxon the user is likely to relate). This requirement means that the representations of relationships cannot occlude the representations of taxa or, at the very least, the representations of groups of taxa selected to have relationships added.

3. 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 interface comprises of a menu and three panels; the toolbar panel, text panel and classification panel. The menu allows the user to open and save sets of related classifications in Taxonomic Concept Schema format[17], switch between modes of navigation (see section 0) and change settings such as the colour coding of classifications.

The toolbar panel contains buttons that: allow the user to create a file with a screenshot of the interface in its current state, change the navigation mode and select a synonymy relationship type for relationship addition or editing.

The interface's text panel is divided on the horizontal plane for each classification displayed in the main classification panel. The upper-most combo-boxes allow the user to select which two classifications from the currently loaded data-set are to be displayed. The combo-boxes below these allow taxa to be selected from a scrollable list or by typing their scientific name. When the user is not interacting with either combo-box they display the full names of the last taxon selected in either classification. The lower-most boxes in the text navigation panel display the name of any brushed taxon in either classification.

The classification panel 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 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 by assigning them appropriate amounts of horizontal and vertical space for horizontal or vertical text. 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 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 prioritised taxa, the lens mode or scroll mode take effect to make taxon labels otherwise accessible. In lens mode a fish-eye lens[18] 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 to pre-empt the users' actions.
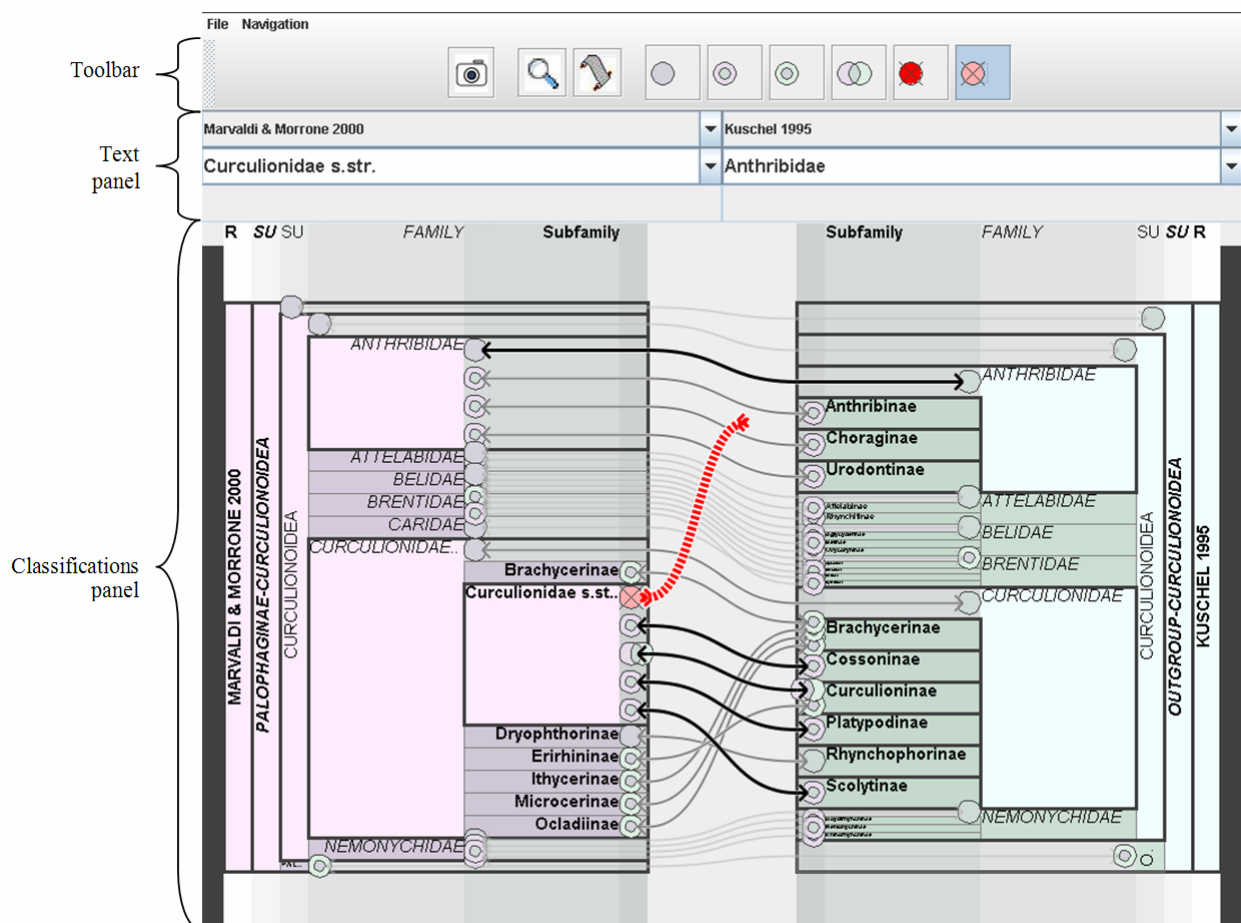

Fig. 1. Screenshot of the Concept Relationship Editor interface.

3.1 Design Rationale

Most of the decisions made during the design of the Concept Relationship Editor evolved directly from the user requirements. The first of these was the choice of layout style for hierarchical classifications. Here, our users told us of the need to make groups of taxa easily recognisable so that relationships can be efficiently added between them. Some form of user interaction is required in order to add a synonymy relationships and it is therefore more appropriate to have

selected taxa with fixed labels rather than rely on additional user interaction to reveal taxon names. This means that taxa need to have space assigned to them in the visualisation and rules out the nesting of taxon representations[6, 11] which would occlude the names of taxa with children in the hierarchy. The other options for representing a hierarchical structure are to use lines for parent-child relationships in a node link layout[5, 7, 8] or have an adjacency layout where child nodes are placed next to their parent[10, 16, 19, 20]. While node-link layouts have been used effectively to communicate structure in other visualisations, adjacency removes the need to have space for lines between nodes making it ultimately more space-efficient. For our system, with the large number of nodes that need to be labelled in order to add relationships efficiently, a more space-efficient adjacency type layout is preferable. Existing adjacency layouts are either top-down[10, 16], centre-out[19] or radial[20] (see figure 2). Of these different layouts the top-down representation with its rectangular node spaces is the most suitable for labelling. It is however limited due to the fact that lower level nodes are assigned progressively less horizontal space. The vast majority of text labels with normal orientation are likely to be wider than longer. If node spaces are made larger to fit text at lower levels it means that node-spaces at higher levels also need to be made larger by the sum of the amount all their children are enlarged. This is likely to have the effect of making upper level nodes, even those with a relatively small number of descendents, massive. A better layout for labelling nodes is to have the top-down layout turned on its side. This means that lower level nodes can be assigned extra horizontal space without it having an impact on the space assigned to their ancestors. To save horizontal space, the smaller numbers of ancestors, which would already have to be taller for their descendents to have adequate height, can have their labels rotated. While it would also be possible to rotate lower levels in a top-down layout to have a similar effect on the label to space ratio, with a sideways layout more labels can maintain a normal text orientation which is easier to read.
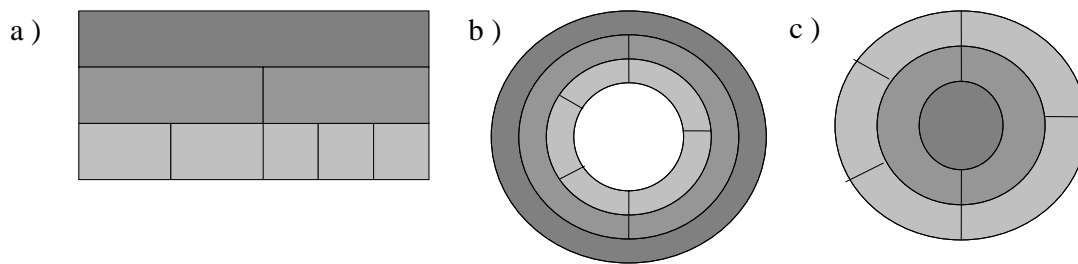

Fig. 2. Adjacency layouts for hierarchical data (lower ranks are shaded lighter): a) top-down, b) center-out and c) radial.

Irrespective of the layout style with larger hierarchies such as taxonomic classifications it is not possible to display all nodes with readable labels without exceeding typical screen space. This makes it necessary to employ some method to properly utilise the existing screen space according to the user's selection. Here, the existing options are: focus + context, and overview and detail. Overview and detail techniques[10, 11, 21, 22] initially display an overview of the data. When the user selects on area of interest from the overview, this is displayed in increasing detail. Focus + context techniques[5, 7, 8, 18] display the entire data but show selected areas of interest with more detail. Often areas of interest are enlarged and surrounding areas have progressively less detail as their distance from to the area of interest increases.

For editing relationships between hierarchies there is a need have multiple groups of similarly classified taxa to be identifiable in either classification at the same time (req. 3). These groups may have very different positions in a hierarchy which makes a basic overview and detail strategy unfeasible as it only allows one area of interest to be visible at a time. While different areas of interest might be included in multiple linked views[23, 24] this would have the undesirable effect of dividing screen space and breaking up the representation of a hierarchical structure. A focus + context technique would not require the division of screen space if the user is allowed to select multiple areas of interest at the same time.

A focus + context solution also helps satisfy our users' requirement to make relationships that might effect the addition of new relationships visible (req. 5). Here, multiple taxa can be selected for focus with other taxa laid out around these with different priorities for their sizing. Taxa without focus may have a small amount of space (and often

not enough space for a readable label) but they are still given a position on the screen so that that position can be used to display relationships with other focused taxa.

For hierarchical data there are two types of focus + context techniques that can be applied. The common approach is to distort the layout itself by assigning different relative spacing to nodes according to the user's selection. This has been demonstrated to allow progressive navigation through a hierarchy[7, 8, 12]. Alternatively focus + context can act as post-process geometrically distorting the visualisation of a fixed layout at a focus point under the mouse cursor. This technique is more commonly applied to unstructured data such as lists of options in a file menu[18].

While a focus + context layout approach is preferable for navigation in our visualisation, the user also has the option to apply geometric distortion (in lens mode) when several long lists of selected taxa cannot be labelled adequately within the screen space. Here the effect is similar to that of fish-eye menus[18] but the magnification of the lens changes so as to be just enough to give brushed nodes enough vertical space to have a readable labels. To ensure magnification does not change too abruptly between nodes, when the mouse cursor is near the edges of a node the algorithm to determine magnification takes into account the magnification required for either neighbouring node.

As an alternative to the geometric distortion of the lens mode the user can switch to scroll mode. In this mode all selected nodes and all nodes that might need labels are given adequate space. If the layout cannot fit the screen space the user can navigate across the hierarchy (up and down in the visualisation) using a vertical scrollbar. This means that the visualisation operates partly as an overview and detail technique as well as a having features of a focus + context view. The disadvantage is that not all nodes have a space on the screen at the same time. This means that not all relationships can be shown completely and if the user needs to access several large lists of nodes to add relationships these may not all fit on the screen at the same time either. The advantage is that relevant node names remaining on the screen are always visible and can be scanned to add relationships quickly. Without comprehensive user testing it is difficult to determine which mode is preferable to the user although early testing indicates it is likely to depend on the stage of adding relationships they are at, the nature of sections within hierarchies (number of nodes, levels etc), the structure of the hierarchies themselves and personal preferences. In any case the user has the option to quickly switch between lens and scroll mode.

3.2 Layout Algorithm for Classifications

The layout of nodes within the adjacency layout of a classification as based around a list of selected focus taxa. Normally a selected taxon is added directly to this list. If a user selects a taxon that is already in the list then that taxon is removed and replaced by its parent. If the new selected taxon is the ancestor of a focused taxon then it is added and all its descendents are removed. If a newly selected taxon is a descendent of a focus taxon then the new taxon is added and the ancestor removed. Thus, a focus taxon cannot be the ancestor or descendent of another focus taxon. Bar this, the manner in which taxa are added and removed is similar to the way nodes are expanded and collapsed in an indented file browser. In the application the expanding and collapsing mechanism is very similar (the difference being that in the Concept Relationship Editor expanded children do not stay expanded when their parent is collapsed).

Taxa directly below focus taxa in the hierarchy are considered focus children, siblings of a focus taxon are considered focus siblings, ancestors of the focus taxon are considered focus ancestors and the siblings of focus taxon are considered focus ancestor siblings. An example of this for a single focus taxon is shown in figure 3a. When there are more than one focus taxon, focus ancestor siblings that are also focus siblings are considered only as focus siblings. For an example of this see figure 3b.

In the next stage of the algorithm heights are assigned to focus taxon children, focus siblings and focus ancestor siblings. These, together with focus ancestors, are what might be termed as secondary focus nodes, being the taxa a user is most likely to select to navigate the hierarchy of the classification. A user would select focus children to drill-down into the hierarchy, select focus ancestors to move up the hierarchy and select focus siblings or focus ancestor siblings to move across the hierarchy. As focus taxon children, focus siblings and focus ancestors are adjacent to focus ancestors, the height assigned to these taxa can be used to calculate the height of focus ancestors. The adjacency of taxa also means that the sum of all focus taxon children, focus sibling and focus ancestor heights is equal to the height of the entire layout.
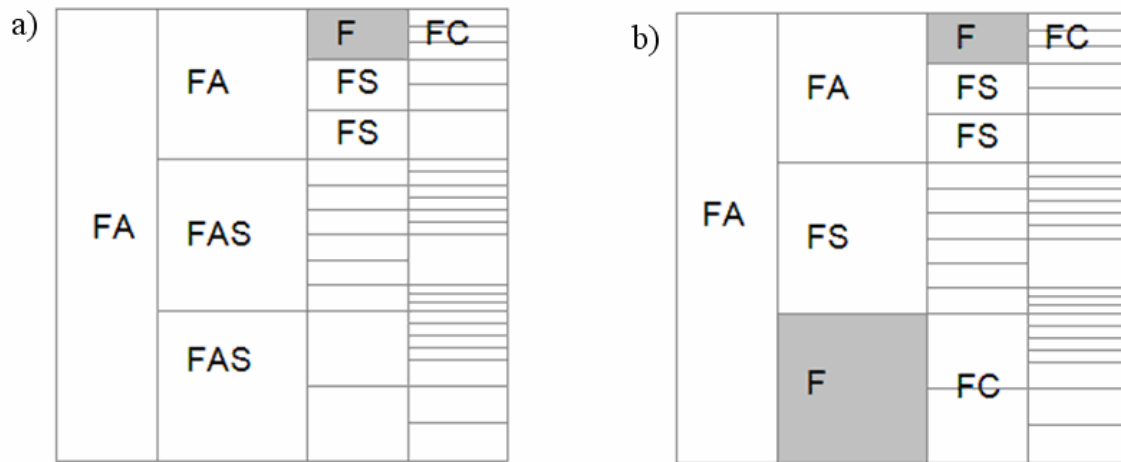
Fig. 3. Classification of nodes in the layout for different selections:. a) a single focus node selected and b) two focus nodes selected. F – focus, FC - focus children, FS - focus siblings, FAS- focus ancestor siblings, FA - focus ancestors

The algorithm attempts to assign enough space so that taxa can have an easily readable label in the visualisation. Labels are deemed to be easily readable if they exceed a specified height threshold. If there is enough height for all focus children, siblings and ancestor siblings to have space for their labels to be readable then the space is distributed evenly amongst them. When there is not enough space for all these nodes, priority is given first to focus children, then to focus siblings and lastly to focus ancestor siblings. This prioritisation attempts to offer the user the best options for navigation with a limited screen space.

If only enough space is available for focus children and focus siblings then these are given the height of the specified threshold with the remaining height distributed evenly amongst the focus ancestor siblings. If only enough space is available for focus children then focus children are given the height of the specified threshold and the remaining height is distributed amongst the focus siblings and focus ancestors. Here twice as much space is assigned to each sibling than there is to each focus ancestor sibling. If there isn't enough vertical space for focus children the height is distributed amongst the focus children, focus siblings and focus ancestor siblings. Here each focus child is given twice as much height as a focus sibling and four times as much as a focus ancestor sibling. After focus children, focus siblings and focus ancestor siblings have been assigned heights, the height of focus ancestors and focus taxa can be determined by the numbers of focus children, focus siblings and focus ancestor siblings they have as descendents (see figure 3). All other taxa are assigned an equal proportion of the height assigned to their parent.

The width of taxon ranks is also determined by the user's selection. Here, focus taxa are given priority up to a maximum width or, if there is enough vertical space for their names to be displayed vertically, the threshold for text height. Focus children are given the next level of priority after focus nodes and successive levels below that successively lower levels of priority still. Ranks below focus taxa where focus taxa have no descendents (i.e. where a branch of the classification has no taxa at a particular rank) and ranks above the focus taxa are given minimal priority. This assigns them a width the same as the threshold for text height. Figures 4-6 show some examples of the layout for a classification with larger numbers of taxa. In each of these, focused taxa and their ancestors have darkened borders that include their descendents to help distinguish groups of taxa adjacent at the same rank but with different parents.

Fig. 4. The adjacency layout with a single taxon focused.

Fig. 5. The adjacency layout with two taxa focused. Ancestor siblings are condensed to save space.
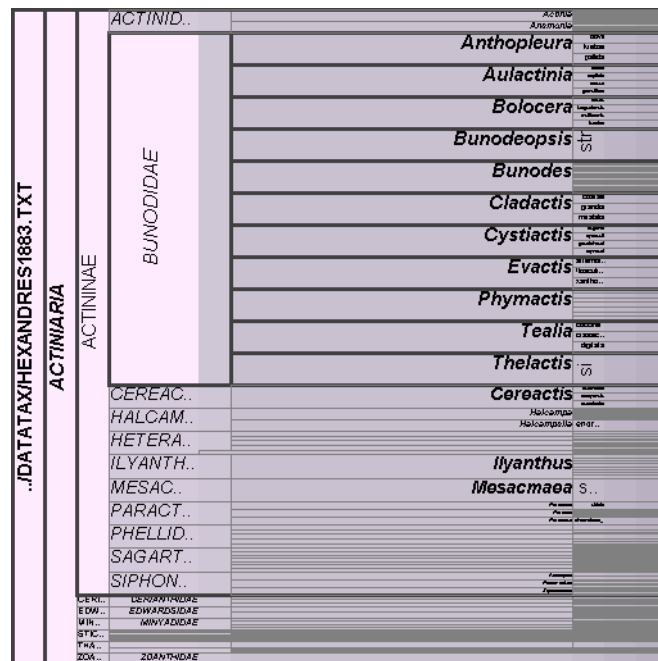
Fig. 6. The adjacency layout with one taxon, with a large number of child taxa, focused. Here the focused taxon has its label rotated to save horizontal space.

3.3 Editing Relationships

To add a relationship between two taxa, after their names are made visible through appropriate navigation of the hierarchy, a user first selects the type of relationship by clicking on a symbolic representation of that relationship type in the toolbar. These symbols are derived from Venn diagrams and colour coded according to the colourings of each classification in the main display. The left-hand-side classification is coloured purple and the right-hand-side classification green. The symbol for two taxa from either classification being congruent is a circle with the two colours mixed. The symbol for an overlap between taxa is two circles of either colour overlapping with their colours mixed in the intersecting area. For a taxon containing a taxon from the other classification, a circle is placed inside another with the outer circles colour the same as the containing taxon's classification and the inner circle the same as the contained taxon's classification. The 'does not contain' and 'is not congruent relationships' are represented by reddish hued circles overlaid with 'x' symbols to emphasise their negative assertions. While the symbols utilise colour, they do not require differentiation between red and green for proper interpretation and can still be used by the majority of colour-blind users[25].

The next step toward adding a relationship between taxa is to select a first taxon, in either classification, and drag the cursor without releasing the mouse button onto the related taxon in the other classification. During the drag a curved line is displayed from the first taxon to either the taxon under the mouse pointer or the mouse pointer itself if it is not over another taxon. Each classification view assigns space for the relationship being added and the transition between views with and without space assigned is smoothed with animation. This allows the user to view what a relationship will look like before the relationship is finally created by releasing the mouse button.

Once relationships are added the curved lines and symbols included in their representation remain part of the visualisation but those not relating either focus taxa or the children of focus taxa have their lines made transparent. This helps to minimise clutter in the display while allowing a user to keep track of which relationships they have already added. Individual relationships are highlighted by brushing their line or symbol. To edit a relationship's type the user

selects its line or symbol and clicks on a new relationship symbol in the toolbar. Similarly a relationship can be deleted by selecting it and pressing delete.

3.4 Fisheye lens and Scroll mode

The algorithm to layout taxa attempts to size focus taxa and secondary focus taxa so that groups of can be adequately labelled. If, however, there are too many selected taxa for this to be possible the user has two options. Firstly, the default option is for a fish eye lens[18] type effect (see figure 7) is applied onto the vertical axis of the visualisation so that the taxon under the mouse cursor is magnified to the threshold height for readable text. When the cursor is moved between taxon labels with a different height the magnification factor changes to accommodate the new taxon label. In an early prototype fisheye distortion was also used on the horizontal axes to make taxon labels wider but users of the tool found the effect to be too erratic to be of practical use.

The other option is for the user to switch to scroll mode. In this mode all focus children, focus siblings and focus ancestor siblings are assigned the threshold height for text and the space for the layout is allowed to extend beyond the screen space. To navigate across the hierarchy in this state the user has the option to operate a scroll bar on the outer side of each classification. When, however, a new focus taxon is selected scrolling to that taxon is automatic. When a relationship is selected both classifications automatically scroll to their related taxa.
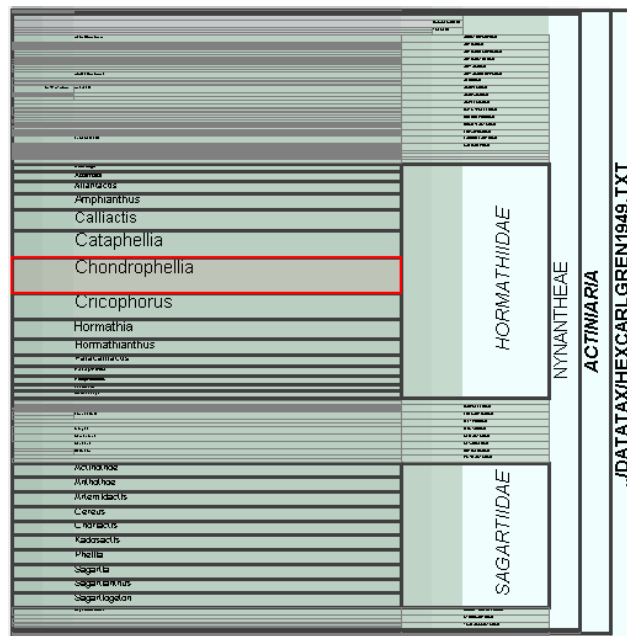

Fig. 7. Lens mode with the lens applied to a group of the horizontally stacked labels of similarly classified taxa.

3.5 Animation

In either scroll mode or lens mode animation[22, 26] is used to smooth the transition between different implementations of the layout algorithm for different user selections. This helps the user by giving them a period of time to comprehend how their actions affect the display, allowing them to watch it gradually change from one state to another rather than immediately presenting them with a new view where a number of features (including the height and width of taxa, the size of labels, and the colour of selections) have changed instantaneously.

This animation involves a regular change of each taxon's coordinates by a proportion of the displacement required, rather than by an absolute value. While this means that taxa with different displacements required move at different speeds, it also ensures that taxa remain nested during animation so the user can maintain context without falsely interpreting a change in structure.

4. EVALUATION

Although there has been no formal evaluation of the software an early prototype was presented to taxonomists at the SEEK all hands meeting late last year (2006). Here we demonstrated the tool to a number of taxonomists and allowed them to practice navigating classifications and explore relationships with a view to the process of adding relationships (the functionality allowing them to do this was not yet implemented at that time). The general consensus was that the tool provided an effective interface for navigating the hierarchy of taxonomic classifications and the layout would be more appropriate for efficiently adding relationships than that of other hierarchy visualisations. Envisioning the process of adding relationships, the taxonomists judged that the 'scroll mode' would tend to allow for relationships to be added more efficiently between smaller groups of similarly classified taxa. They also considered the 'lens mode' to be useful in providing context and allowing for more direct comparison of relationships throughout the classifications. During a more recent informal evaluation with a smaller group of taxonomists the tool included the functionality to drag-and-drop to assert relationships and the taxonomists found this process trivial.

A more formal evaluation of the software is imminent. This is planned to be performed in three stages. The first stage is to be a task and scenario based evaluation designed to uncover usability issues. The second stage will be devised to evaluate different options for minor display elements (symbols to encode relationship types, styles of lines to use, curvature/ non-curvature of lines etc). The final evaluation will involve allowing taxonomists to mark up a new data-set using the technique.

5. CONCLUSION AND FURTHER WORK

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. This constituted a significant improvement over the techniques currently employed.

The general technique of the Concept Relationship Editor would also allow the exploration and editing of relationships between other types of hierarchy such as gene classification or file structure data. While the technique was designed for data with a strict rank structure, there is nothing to prevent data without ranks also being used, in which case the depth of a node in the hierarchy would be used to assign a nominal rank.

A possibility for further developments is to extend the technique to include intra-hierarchy relationships. This would involve representing a single hierarchy on both sides of the main panel and having each relationship displayed twice. A similar strategy might be employed to display directed acyclic graphs. Here a minimum spanning tree algorithm could be used to extract a hierarchy. Those relationships from the original graph that are not extracted by the algorithm could be displayed as intra-hierarchy relationships.

ACKNOWLEDGMENTS

REFERENCES

1.   Kukla, R. and T. Paterson. *Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. .* in *2nd International Workshop on Data Integration in the Life Sciences* 2005. San Diego, USA.
2.   Berendsohn, W.G., *The concept of "potential taxa" in databases.* Taxon, 1995. **22**: p. 207-212.
3.   Zhong, Y., et al., *Data model and Comparison and Query Methods for Interacting Classifications in a Taxonomic Database.* Taxon, 1996. **45**(2): p. 223-241.
4.   Koperski, M., et al., *Referenzliste der Moose Deutschlands.* Schriftenreihe für Vegetationskunde. Vol. 34. 2000, Münster-Hiltrup: LV Druck im Landwirtschaftsverlag GmbH. 519.
5.   Jeong, C.-S. and A. Pang. *Reconfigurable Disc Trees for Visualizing Large Hierarchical Information Space.* in *IEEE InfoVis '98.* 1998. Research Triangle, North Carolina, USA: Computer Society Press.

6. Johnson, B. and B. Shneiderman. *Treemaps: A Space-Filling approach to the visualization of hierarchical information structures*. in *IEEE Visualization '91*. 1991. San Diego, California, USA: IEEE Computer Society Press.

7. Lamping, J., R. Rao, and P. Pirolli. *A focus + context technique based on hyperbolic geometry for visualizing large hierarchies*. in *ACM CHI '95*. 1995. Denver, Colorado, USA: ACM Press.

8. Munzner, T., *Exploring Large Graphs in 3D Hyperbolic Space*. IEEE Computer Graphics & Applications, 1998. **18**(4): p. 18-23.

9. Neumann, P., S. Schlechtweg, and S. Carpendale. *ArcTrees: Visualizing Relations in Hierarchical Data*. in *EUROVIS 2005: Eurographics / IEEE VGTC Symposium on Visualization*. 2005.

10. Sifer, M., *Filter co-ordinations for exploring multi-dimensional data*. Journal of Visual Languages & Computing 2005. **17**(2): p. 107-125.

11. Wang, W., et al. *Visualization of large hierarchical data by circle packing*. in *SIGCHI conference on Human Factors in computing systems*. 2006. Montréal, Québec, Canada: ACM Press.

12. Munzner, T., et al., *TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with Guaranteed Visibility*. ACM Trans. Graph., 2003. **22**(3): p. 453-462.

13. 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): p. 741 - 748.

14. Graham, M. and J. Kennedy, *Extending taxonomic visualisation to incorporate synonymy and structural markers*. Information Visualization, 2005. **4**(3): p. 206-223.

15. Liu, X. *Concept Mapper*. 2006 [cited; Available from: http://seek.ecoinformatics.org/Wiki.jsp?page=ConceptMapper.

16. Graham, M., J.B. Kennedy, and C. Hand. *A Comparison of Set-Based and Graph-Based Visualisations of Overlapping Classification Hierarchies*. in *AVI 2000*. 2000. Palermo, Italy: ACM Press.

17. Kennedy, J.B., R. Kukla, and T. Paterson, *Scientific Names Are Ambiguous as Identifiers for Biological Taxa: Their Context and Definition Are Required for Accurate Data Integration*, in *Data Integration in the Life Sciences*. 2005, Springer: Berlin / Heidelberg. p. 80-95.

18. Bederson, B.B. *Fisheye Menus*. in *ACM UIST 2000*. 2000. San Diego, California, USA: ACM Press.

19. Stasko, J. and E. Zhang. *Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations*. in *IEEE Symposium on InfoVis 2000*. 2000. Salt Lake City, Utah, USA: IEEE Computer Society Press.

20. 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. **12**(5): p. 741 - 748.

21. Lee, B., et al., *TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts* IEEE Transactions on Visualization and Computer Graphics, 2006. **12**(6): p. 1414-1426.

22. Graham, M. and J. Kennedy. *Combining linking & focusing techniques for a multiple hierarchy visualisation*. in *IV 2001*. 2001. London, UK: IEEE Computer Society Press.

23. Roberts, J.C. *On Encouraging Multiple Views for Visualization*. in *IEEE International Conference on Information Visualization '98*. 1998. London, UK: IEEE Computer Society Press.

24. Craig, P. and J.B. Kennedy. *Coordinated Graph and Scatter-Plot Views for the Visual Exploration of Microarray Time-Series Data*. in *I2003 IEEE Symposium on Information Visualization*. 2003. Seattle WA: IEEE Computer Society.

25. Brewer, C.A. *Guidelines for Use of the Perceptual Dimensions of Color for Mapping and Visualization*. in *International Society for Optical Engineering (SPIE)*. 1994. San Jose.

26. Craig, P., J.B. Kennedy, and A. Cumming, *Animated Interval Scatter-plot Views for the Exploratory Analysis of Large Scale Microarray Time-course Data*. Information Visualisation, 2005. **4**(3).