

Visualizing Biodiversity with Voronoi Treemaps

Michael S. Horn, Matthew Tobiasz, and Chia Shen
Initiative in Innovative Computing at Harvard University
60 Oxford Street
Cambridge, MA 02138 USA

Email: mhorn@seas.harvard.edu, tobiasz@seas.harvard.edu, chia_shen@harvard.edu

Abstract—Introduced in 2005, the Voronoi treemap algorithm is an information visualization technique for displaying hierarchical data. Voronoi treemaps use weighted, centroidal Voronoi diagrams to create a nested tessellation of convex polygons. However, despite appealing qualities, few real world examples of Voronoi treemaps exist. In this paper, we present a multi-touch tabletop application called *Involv* that uses the Voronoi treemap algorithm to create an interactive visualization for the Encyclopedia of Life. *Involv* is the result of a year-long iterative development process and includes over 1.2 million named species organized in a nine-level hierarchy. Working in the domain of life sciences, we have encountered the need to display supplemental hierarchical data to augment information in the treemap. Thus we propose an extension of the Voronoi treemap algorithm that employs force-directed graph drawing techniques both to guide the construction of the treemap and to overlay a supplemental hierarchy.

Keywords—Voronoi treemaps; information visualization; multi-touch tabletop displays; interaction design; biodiversity

I. INTRODUCTION

In this paper, we present a multi-touch tabletop application called *Involv*, that uses the Voronoi treemap algorithm to create an interactive visualization for the Encyclopedia of Life [5]. *Involv* is an ongoing project of the Scientists' Discovery Room Lab (SDR) at the Initiative in Innovative Computing at Harvard University. At the writing of this paper, the visualization includes over 1.2 million named species organized in a nine-level hierarchy. In attempting to visualize this taxonomy of life on earth, we have encountered the need to display supplemental hierarchical information to augment our primary data source. In particular, we have found it useful to integrate phylogenetic information (i.e. information about the evolutionary history of organisms) with our existing visualization. Thus, we propose an extension of the Voronoi treemap algorithm that employs force-directed graph drawing techniques to overlay supplemental phylogenetic information on top of the treemap. This technique has the added benefit of guiding the spatial layout of regions in the treemap based on the evolutionary relatedness of the various taxonomic groups.

II. RELATED WORK

Introduced in 2005, the Voronoi treemap algorithm is an information visualization technique for displaying hierarchical datasets [1]. The Voronoi treemap is an iteration



Figure 1. The *Involv* system uses the Voronoi treemap algorithm [1] to create an interactive visualization for the Encyclopedia of Life.

on the traditional treemap algorithm [15] which recursively subdivides two-dimensional rectilinear space to produce a containment view of hierarchical data. Typically, the size of a given rectangular region corresponds to some associated weight—for example, the area of a region could correspond to the number of leaf nodes contained in that section of the hierarchy. Balzer and Deussen propose an alternative treemap algorithm that uses weighted, centroidal Voronoi diagrams to produce a nested tessellation of convex polygons [1]. They argue that the resulting Voronoi treemaps offer potential advantages for information visualization including improved clarity, better aspect ratios for treemap regions, and the ability to move beyond rectangular spaces and inscribe treemaps within arbitrary polygons. However, despite potential advantages, the information visualization community has been slow to adopt Voronoi treemaps, and few real world examples exist.

In Section IV, we outline an algorithm to overlay supple-

mental hierarchical information on top of a Voronoi treemap. There exist numerous tree comparison tools that also attempt to help users explore similarities and differences in related hierarchical data sets. For example, TreeJuxtaposer [11] uses innovative Focus+Context techniques to help scientists conduct side-by-side comparisons of two phylogenetic trees. The VisLink system [3] is a novel comparison tool that cross-references information in arbitrary visualizations using a pseudo-3D approach that bundles links between diagrams drawn on 2D planes. Holten created a technique for overlaying adjacency information on top of several styles of tree diagrams including treemaps [9]. The adjacency information is represented as bundled sets of edges. For this project, the focus was on reducing clutter in very large datasets through the use of bundling, transparency, and edge routing.

III. APPLICATION DOMAIN

Involv is an interactive visualization of the Encyclopedia of Life (EoL), an ambitious, multi-institution effort to “make available via the Internet virtually all information about life present on Earth” [5]. Launched in 2007, EoL seeks to provide an authoritative classification of life on earth using the nine-level Linnaean taxonomy: Kingdom, Phylum, Class, Order, Superfamily, Family, Genus, Species, and Infraspecies. This traditional hierarchical classification system, first proposed by Carl Linnaeus in the 18th century, organizes species into groups based on similarities in observed morphology. Of the approximate 1.75 million species that “have been formally described and given official names” [4], EoL has roughly 1.2 million entries.

The Tree of Life (ToL) project [10] is an EoL partner organization that publishes a scientifically vetted phylogenetic tree containing over 77,906 species and higher level taxonomic groups. A phylogeny is a tree structure that shows a hypothetical evolutionary history of a group of organisms, where parent nodes represent the most recent common ancestor of its children (Figure 2, right). These evolutionary relationships are determined through genetic analysis, and a large scientific effort is currently underway to resolve all life on earth within this phylogenetic tree.

Involv combines data from the Encyclopedia of Life and the Tree of Life, into a unified visualization environment. To this end, it is possible to match nodes in the taxonomy (from EoL) to nodes in the phylogeny (from ToL) based on the scientific names of organisms or other structural metrics (e.g. [11]). This matching is far from perfect—for the Involv project we have matched 32,700 nodes, or approximately 2.7% of the entries in the EoL. However, there is still a wealth of information to visualize for the user, especially at higher levels of the taxonomy. The following section describes a technique for connecting these two sources, by overlaying phylogenetic structures on top of a Voronoi treemap browser.

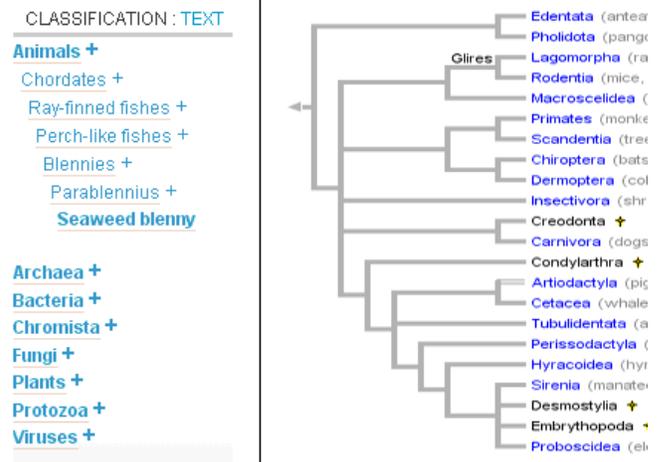


Figure 2. The Encyclopedia of Life [5] organizes species using a nine-level Linnaean taxonomy, as shown in the indented list-style browser (left); while the Tree of Life [10] groups species based on relatedness, as shown in the phylogenetic tree browser (right). Both examples were taken from web pages of their respective sources.



Figure 3. The Voronoi treemap algorithm produces a nested tessellation of convex polygons.

IV. VORONOI TREEMAP VISUALIZATION

The goal of the Involv project is to create a compelling, interactive visualization that encourages users to collaborate around a shared multi-touch display. As a touch-based application where users' visual orientation and finger size are prominent design considerations [14], we felt strongly that the Involv hierarchy browser should be based on space-filling approaches offered by treemaps. Other styles of tree visualizations (such as node-link diagrams) tend

to create long horizontally or vertically oriented lists and dedicate substantial screen real estate to representing the tree structure. For example, ToL uses horizontally-oriented cladograms (Figure 2, right). Similarly, the tree browser provided on the EoL website (Figure 2, left) uses an indentation list style similar to that offered by graphical file system browsers. These representations are, perhaps, more appropriate for single users on mouse-based platforms with a fixed orientation relative to the display. In our case, however, multiple users are gathered around a tabletop surface, and we want to optimize the use of screen space and avoid presenting lists of items that favor only a single individual.

In 2005, Balzer and Deussen proposed an innovative reinterpretation of treemaps that uses weighted, centroidal Voronoi diagrams to create a nested tessellation of convex polygons [1] (Figure 3). We selected Voronoi treemaps as a basis for the Involv project due to the organic visual aesthetic which seemed appropriate for the life sciences domain. We also hypothesize that the geometric diversity of the Voronoi tessellations would serve as a mnemonic device for users unfamiliar with the hierarchy being visualized. In other words, by breaking out of the rectilinear uniformity of standard treemaps, we hope to provide the equivalent of geographic landmarks for users as they browse.

Voronoi treemaps are constructed using a set of randomly generated seed points and an algorithm that iteratively adjusts the weight and position of each region in the map until the actual area of each cell approaches its desired area within some error threshold. In our case the size of each region corresponds to the number of named species within that section of the taxonomy. For example, the Animal Kingdom contains more named species than the Plant Kingdom, which, in turn, contains more named species than Fungi. Note, that this does not reflect the actual number of species on the planet. For example, scientists estimate that there are millions of species of bacteria, yet only a small fraction of these have been formally named.

Based on user feedback, we have adjusted the visualization so that the areas of Voronoi regions are not linearly proportional to their corresponding number of named species. Instead, we set the desired area α equal to n^α where n is the number of named species and $0.5 \leq \alpha \leq 1$. This prevents small (but not necessarily uninteresting) sections of the tree from disappearing at the standard zoom level. For example, within mammals, Order Proboscidea (elephants) contains only three named species while the Order Rodentia (rodents) contains over 2,000 named species.

A. Visualizing Related Phylogeny Trees

As described above, our treemap-based browsing system uses a nine-level Linnaean taxonomy provided by the EoL. This traditional classification system organizes species based on morphology; however this approach does not necessarily place related species into the same group. For example, the

Kingdom Protozoa includes organisms with diverse evolutionary histories that might better be variously categorized as plants, fungi, or even animals. Furthermore, several pilot users of our system indicated that they expected the spatial layout of the treemap on each level to carry meaning. *Why are elephants and rabbits drawn side-by-side here anyway?* Thus, it seemed useful to augment our treemap structure with phylogenetic information from the ToL database. To do this, we group regions spatially in the treemap based on their relatedness, overlaying a representation of the phylogenetic tree on top of the Voronoi treemap to make these relationships explicit (Figure 5). We have developed a technique to accomplish these two objectives simultaneously by combining common force-directed graph drawing techniques with the iterative Voronoi treemap algorithm. The following subsection describes the origin and implementation of this algorithm.

B. Flow Maps

Flow maps are a cartographic technique to represent the flow of quantity (such as traded goods or the migration of people) over a geographic space. Phan *et al.*[12] describe an algorithm to reduce visual clutter in flow maps and render them in an aesthetically appealing way.

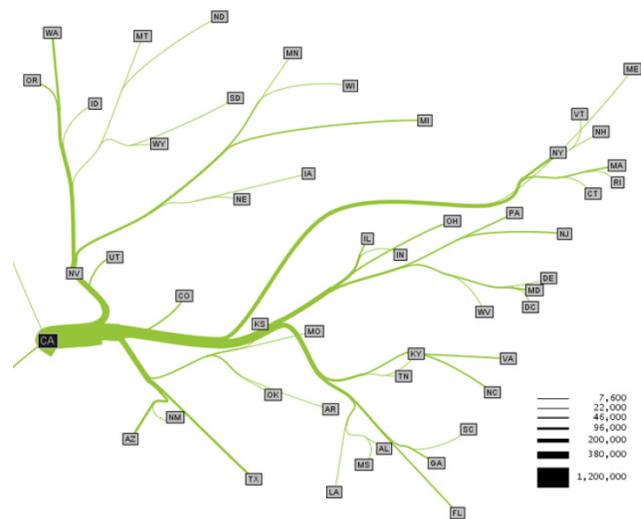


Figure 4. An example flow map rendered using the method proposed by Phan *et al.*[12]. This flow map shows migration from California across the United States from 1995-2000. *Used with permission.*

Following the best practices of cartographers, Phan *et al.*'s algorithm uses a combination of three strategies: 1) it hierarchically bundles edges from a source to each destination so that flow networks forms a binary tree structure; 2) it subtly shifts the locations of destination points so as to maintain a minimum separation distance; and 3) it uses edge routing to prevent overlap of flow lines and site labels. Figure 4 shows an example flow map generated by

this system. We propose that a flow map style rendering could be useful for superimposing supplemental hierarchical data on a Voronoi treemap. However, whereas in flow maps lines of flow are strategically bundled to be drawn over a fixed geographical space, we do the opposite. In our case the Voronoi tessellation serves as an abstract and malleable geographical space, and the structure of flow lines (i.e. the phylogenetic tree structure) is important and cannot be modified.

C. Force-Directed Graph Drawing

An advantage of Voronoi treemaps for Involv is the technique used to generate the tessellations. Rectilinear treemaps are generally constructed using deterministic layout algorithms. Voronoi treemaps, on the other hand, are constructed by generating random seed points and then by iteratively adjusting the positions of the seed points and weights until the area of each region in the map is equal to its desired area within a given error threshold. The use of random seed points and iterative updates lends this algorithm to be adapted for use with force-directed graph drawing techniques.

Force directed graph drawing algorithms apply simulated forces to a graph structure in an iterative process to produce legible layouts. The underlying assumption is that relaxed (energy-minimal) states correspond to readable graphs [2]. In general, the goal is to produce graph drawings with minimal edge crossings, evenly-spaced nodes, and edges with similar lengths. To accomplish this, these algorithms typically use physical models of charged particles for nodes (to produce even spacing) and spring forces for edges (to reduce crossings and to enforce consistent edge lengths).

D. Voronoi Flowmap Algorithm

In 1998, Gansner and North [6] proposed the use of Voronoi diagrams as a post-processing step for force-directed graph layouts. By constructing a centroidal Voronoi diagram (using node points as sites), they were able to reduce clutter in graphs with non-point nodes. We have developed a similar technique based on force-directed graph drawing that combines two simulated forces: a Voronoi *centroidal force* and a phylogenetic *tree spring force*. For each iteration of the Voronoi treemap algorithm, generator points are moved to the center of mass of their corresponding Voronoi regions. In our modified algorithm, we treat this as a force vector that pulls a generator point toward its regions center of mass. We multiply this force vector by a scalar value that we call the centroidal force coefficient c where $0 \leq c \leq 1$. By keeping generator points near the center of mass of their containing polygons, the centroidal force serves a purpose similar to the “charged partical” force (in that it tends to maintain a minimum distance between nodes of the graph).

In addition to the centroidal force, we use the phylogenetic tree structure to create *spring forces* that act on the Voronoi diagram as it is being created. In essence, the phylogenetic tree acts as a network of springs governed by Hookes Law [2]. Each spring in the network has a fixed natural length, l , and a force coefficient k (where $0 \leq k \leq 1$). This spring force works to reduce edge crossings in the phylogenetic tree and to pull related nodes into close proximity with one another. For each iteration of the algorithm, we compute the sum of these two force vectors for each generator point and then update its position accordingly.

In general, a stronger spring force (i.e. a higher k value) results in layouts with better groupings and fewer edge crossings. However, the spring forces also act to destabilize the Voronoi treemap, increasing the number of iterations necessary to achieve an acceptable layout and often times causing the algorithm to iterate indefinitely. Thus, we adjust both the c and k parameters iteratively. Specifically, k starts out with a high value (close to 1) and is gradually decreased until it reaches a value close to 0. Likewise, c is initialized to a low value and is gradually increased until it reaches a value close to 1. This allows the spring force network to arrange the nodes early in the process and then for the centroidal force to gradually take over to stabilize the diagram. As a technical note, the original Voronoi treemap algorithm adjusts weight values for each region by a factor of at most 2 every iteration. We found that using an additional factor w ($0 \leq w \leq 1$) to retard this exponential growth was helpful in achieving uncluttered layouts. We fix w with a value of 0.4. Note that this is not necessarily an optimal algorithm. It is possible that adjusting these parameter values using non-linear functions or adaptively based on performance might result in better layouts with fewer iterations.

E. Results

This force-directed approach works well in certain areas of the taxonomy. For example, Figure 5 shows the result of the algorithm run on the Order Passeriformes (perching birds). Here, the phylogenetic tree contains 61 nodes and is up to 20 levels deep. However, while the number of edge crossings is small, parts of the visualization near the bottom of the display are still difficult to interpret due to occlusions and edge crossings. Future directions for this work include employing edge routing techniques to minimize occlusion and adding interactive elements to allow users to resolve uncertainty using zoom or other inspection techniques.

In other areas of the taxonomy, our approach is less successful. For example, the Family Salticidae (jumping spiders) has a phylogeny with 203 nodes up to 12 levels deep. There are also several clusters of terminal nodes that share a common parent in the hierarchy—for example, one parent node has 52 children. The resulting phylogeny overlay is difficult to interpret due to the large number of nodes and the dense clusters of siblings that tend to clutter the display.

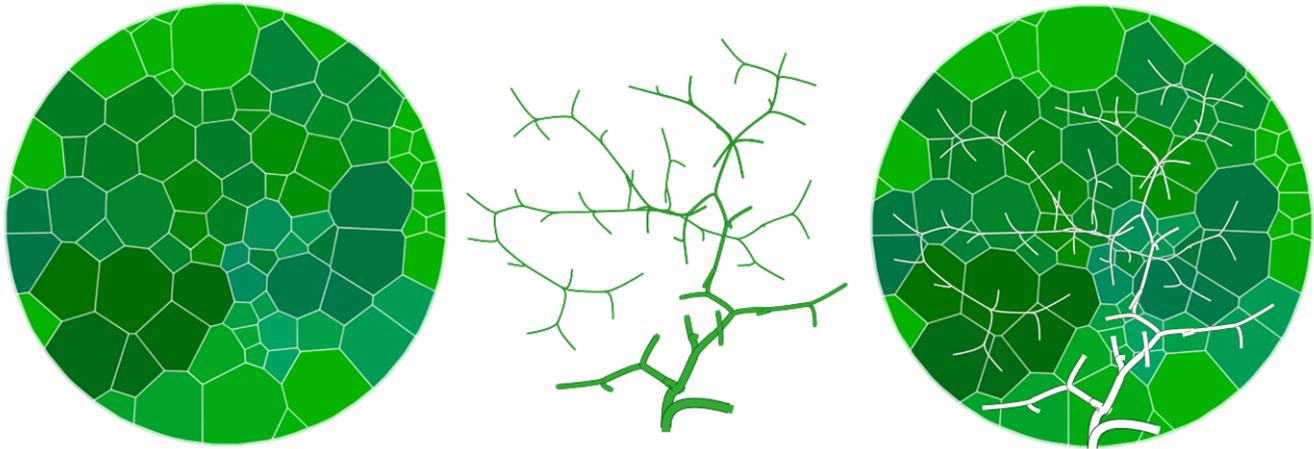


Figure 5. This figure shows the results of our Voronoi flowmap algorithm on the order Passeriformes (perching birds). The phylogeny tree contains 61 nodes and is up to 20 levels deep. The resulting visualization is shown with the Voronoi tessellation only (left), the phylogeny tree only (center), and the phylogeny tree superimposed on the Voronoi tessellation (right).

V. TABLETOP INTERACTION

Throughout the design and development of Involv, we held weekly meetings with potential users and domain experts for usability testing, as well as scientific validation of our system. At any level of the taxonomy users can rotate, zoom with two hands, and pan the current view using direct-touch gestures such as pinching or twisting with two or more fingers. This is important because it allows users to reveal the names of smaller regions (by zooming in), to bring selections of the display into arms reach, and to orient the display according to their position around table top in a group collaborative setting.

Large hierarchies with unfamiliar data can be confusing to users [8]. For example, in an informal observational session of ten researchers and staff from our lab, only three knew that the Class Mammalia is a subset of the Phylum Chordata. To help alleviate this sense of disorientation we incorporate two multi-media elements into our visualization. First, a small window displays a scrolling list of images contained in the current section of the tree structure (Figure 6). Multiple users can drag images out of this *media scroller* with their fingers onto the table surface. Once on the table, images can be rotated or scaled using finger-based gestures. Second, we added a feature to allow users to see an instant preview image by resting their finger on a region of the map or by *brushing* (sliding) their finger from region to region. For example, by touching the region labeled “Chordates”, a user might see a picture of a fish or a mammal. We designed both of these multi-media features with simultaneous multi-user interaction in mind.

To drill down into the hierarchy, users tap a labeled button in the center of the desired region with their finger. We use a target area size that is appropriate for finger touch interaction—the button is just large enough to be tapped

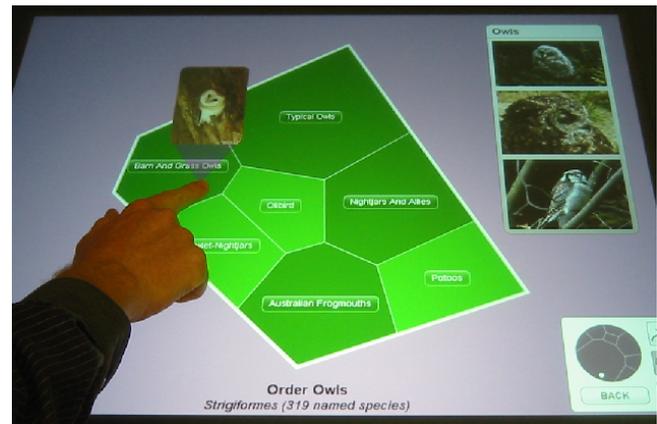


Figure 6. Involv includes two multi-media features to provide users with additional context while browsing large and potentially disorienting hierarchies. These include the media scroller (top right) and touch-activated preview images (left).

accurately with a finger tip—to prevent accidental drill-downs [13]. To *pop* back up in the hierarchy, users press a BACK button located on one of the two control panels (see Figure 1) located on opposite sides of the table for convenient access by multiple users. These control panels include a world-in-miniature view that highlights the current position in the hierarchy. To transition between levels of the hierarchy, we use zoom and pan animations to provide a sense of visual continuity between levels of the tree. This is especially important for multi-user interaction, where actions initiated by a single user must be comprehensible to the entire group. Such awareness has been shown to be crucial for group dynamics [7].

VI. CONCLUSION

In this paper we described the Involv system, a multi-touch tabletop application that uses a Voronoi treemap to display a taxonomy of life on earth provided by the Encyclopedia of Life. We also outlined an algorithm to overlay supplemental phylogenetic information on top of the treemap. This technique has the added benefit of guiding the spatial layout of regions in the treemap based on the evolutionary relatedness of the various taxonomic groups. And while this technique is preliminary, we believe that when combined with other strategies such as edge routing and interactive elements, it has the potential to become a more general and useful visualization strategy.

ACKNOWLEDGMENTS

We thank James Hanken from the Department of Organismic Evolutionary Biology at Harvard for his early discussions on the importance of visualizing phylogeny together with our initial taxonomy layout. We also thank Robert A. Lue for initiating our contact with the Encyclopedia of Life, and his valuable discussion at the early stages of this project. This INVOVL project at the SDR Lab of IIC at Harvard University has been partially funded by the Encyclopedia of Life and the Museum of Comparative Zoology at Harvard University.

REFERENCES

- [1] M. Balzer and O. Deussen. Voronoi treemaps. In *INFOVIS '05: Proceedings of the 2005 IEEE Symposium on Information Visualization*, pages 49–56, Washington, DC, USA, 2005. IEEE Computer Society.
- [2] U. Brandes. Drawing on physical analogies. In M. Kaufmann and D. Wagner, editors, *Drawing Graphs: Methods and Models*. Springer-Verlag, 2001.
- [3] C. Collins and S. Carpendale. VisLink: Revealing relationships amongst visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1192–1199, 2007.
- [4] J. E. Duffy. Biodiversity. In J. Lloyd, editor, *Encyclopedia of Earth*. National Council for Science and the Environment, 2007. Available from <http://www.eoearth.org/article/Biodiversity>. Accessed 25 April 2009.
- [5] Encyclopedia of life. Available from <http://www.eol.org>. Accessed 25 April 2009.
- [6] E. Gansner and S. North. Improved force-directed layouts. In *6th International Symposium on Graph Drawing GD'98*, pages 364–373. Springer-Verlag, 1998.
- [7] C. Gutwin and S. Greenberg. The importance of awareness for team cognition in distributed collaboration. In E. Salas and S. M. Fiore, editors, *Team Cognition: Understanding the Factors that Drive Process and Performance*, pages 177–201. APA Press, 2004.
- [8] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, 2000.
- [9] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, 2006.
- [10] D. R. Maddison and K. S. S. (eds.). The tree of life web project, 2007. Available from <http://tolweb.org>. Accessed 25 April 2009.
- [11] T. Munzner, F. Guimbretière, S. Tasiran, L. Zhang, and Y. Zhou. Treejuxtaposer: scalable tree comparison using focus+context with guaranteed visibility. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, pages 453–462, New York, NY, USA, 2003. ACM.
- [12] D. Phan, L. Xiao, R. Yeh, P. Hanrahan, and T. Winograd. Flow map layout. In *INFOVIS '05: Proceedings of the 2005 IEEE Symposium on Information Visualization*, pages 219–224, Washington, DC, USA, 2005. IEEE Computer Society.
- [13] K. Ryall, C. Forlines, C. Shen, M. R. Morris, and K. Everitt. Experiences with and observations of direct-touch tabletops. In *TABLETOP '06: Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, pages 89–96, Washington, DC, USA, 2006. IEEE Computer Society.
- [14] C. Shen, F. D. Vernier, C. Forlines, and M. Ringel. Diamond-spin: an extensible toolkit for around-the-table interaction. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 167–174, New York, NY, USA, 2004. ACM.
- [15] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11(1):92–99, 1992.