

# Communicating the Tree of Life: a Project Proposal

## ***The Team***

Denise Green  
Rebecca Shapley

## ***Problem Statement/Idea***

For our SIMS Master's Project we will conduct needs assessment and usability evaluations in support of the NSF Tree of Life initiative.

The NSF Tree of Life initiative is funding biologists' work to identify and resolve the less-well known parts of the evolutionary divergence process that relates all organisms on earth. These relationships are often shown as a branching tree, earning the nickname the "Tree of Life." Existing and future results from this taxpayer-funded project should be useful in a variety of professional and educational contexts.

With the understanding that the Tree of Life data will be presented through an interactive online application, our goals are to

- a) identify important tasks that members of the target audiences will perform with the application;
- b) define what qualities should be present in an effective visualization and interaction environment for the Tree of Life data;
- c) conduct the first round of evaluations on the usability and effectiveness of three different interactive approaches to visualizing this data;
- d) document resulting preliminary requirements for the online Tree of Life application.

The work for this project will be conducted within the context of Dr. Nancy van House's SIMS 214 course on Needs and Usability Assessment, Rebecca's continuing GSR position, and the SIMS 298 Directed Group Study final project course.

## ***Context/Prior Work/Motivation***

Our project builds on the work Rebecca has been conducting as a Graduate Student Researcher under the supervision of Dr. Brent Mishler, Director of the University and Jepson Herbarium at UC Berkeley. See [www.sims.berkeley.edu/~rebecca/cipres/](http://www.sims.berkeley.edu/~rebecca/cipres/) for details.

During the Fall 2004 semester, Rebecca

- a) identified likely audience categories (users) of the Tree of Life data;
- b) conducted two interviews with non-academic members of the audiences;
- c) reviewed existing approaches to visualizing tree-structured data;

- d) developed a preliminary classification for those visualizations;
- e) and based on heuristic evaluations, identified three visualization approaches deserving further testing with target audience members: hyperbolic trees, treeMaps, and UCB CS grad student Jeff Heer's Degree of Interest trees.

Feedback from this work has been quite positive, with particular enthusiasm for hearing about the results from interviews. The work proposed here for the SIMS Master's project continues and furthers these activities.

The GSR position is funded by the CIPRES grant, which is funded in turn by the NSF Tree of Life initiative. CIPRES brings biologists and computer scientists together from a variety of universities to develop the next generation of software tools and algorithms for conducting the phylogenetics research that produces the Tree of Life data. Fortunately, a CIPRES meeting is being held the last weekend in January in San Diego, allowing both project team members to attend, get immersed in the subject area, and conduct interviews with members of the target audiences.

The Tree of Life data is a special instance of tree-structured data, and has its own history of visualizations and interactions among biologists. Tree-structured data, often seen as a special case of graphs, is a common focus of work in fields such as Computer Science, Human-Computer Interaction research, and Information Visualization. Developers of classification systems such as librarians and website navigation designers also work with branching data structures. The Tree of Life data benefits from all of this work, and brings additional interesting challenges:

- a) Unlike most graphs, some nodes are much more important than others. How can the data be effectively summarized to present only important branch points dynamically?
- b) Because nodes/branches are named by specialists, for many target audience members the name of branch doesn't connote anything about the other nodes included on it. Good website navigation design avoids this problem by giving branches names that pass user-testing as being evocative of the branch contents. Without second-guessing the scientists' work, what approach can be taken here?
- c) The data set is huge, unbalanced, and changes over time.
- d) The structure of the Tree of Life is itself a product; for various reasons, people need a memorable, take-away mental impression of the Tree of Life.

We hope to be able to say some interesting and practical things on these points, to deepen our understanding of the needs assessment process, and to enjoy working with a highly-visual project.

## ***Tentative Project Schedule***

A tentative schedule for the major activities of the Master's project is;

January	<ul style="list-style-type: none"><li>• Complete Project Proposal</li><li>• Establish Dr. Nancy van House as advisor</li><li>• Attend CIPRES conference in San Diego</li><li>• Begin interviews</li></ul>
February 1 to February 28	<ul style="list-style-type: none"><li>• Establish project website</li><li>• Conduct task analysis through interviews, survey with target audience members.</li><li>• Review literature on usability studies with tree-visualizations.</li><li>• Define effectiveness of interactions with and visualizations of the Tree of Life data, and more specifically, with our test data, the Green Plants branch.</li></ul>
February 28 to March 18	<ul style="list-style-type: none"><li>• Prepare software for usability evaluations of the three visualization approaches (hyperbolic trees, treeMaps, and degree-of-interest trees) with the Green Plants data.</li><li>• Develop and run preliminary tests of the evaluation procedures.</li><li>• Establish evaluation appointments with target audience members.</li></ul>
March 18 to March 28	SPRING BREAK
March 28 to April 8	Conduct usability evaluations with target audience members.
April 8 to April 26	<ul style="list-style-type: none"><li>• Complete requirements recommendations</li><li>• Complete the project website</li><li>• Complete any written documentation required for the Master's project.</li></ul>
April 26 to May 10	<ul style="list-style-type: none"><li>• Prepare project poster</li><li>• Prepare project presentations</li><li>• Present poster, presentations</li></ul>

## ***Final Project Deliverables***

We plan to create the following deliverables:

1. A project website that includes
  - a. Interview questions
  - b. Transcripts of interviews with target audience members
  - c. Survey questions and results for the task analysis
  - d. Task analysis report

- e. Annotated bibliography of existing literature on user testing conducted with tree-structured data
  - f. Our definitions of effective interactions with a Tree of Life application
  - g. Documentation on the procedure, tools, and results of our usability evaluation
  - h. Our report giving recommendations on the application requirements for the future online application that presents the Tree of Life data, directed at the CIPRES project and in general to future software developers in this area.
2. Three pre-existing software packages that have been adjusted as needed for use in our usability evaluation, and the Green Plant data files for each of them.
  3. Written reports for the Master's project.

It is likely that the results will be appropriate to compile into a paper and submit it for publication.