Visualizing Semantic Wiki Structure

Exploring the Hesperian Digital Library

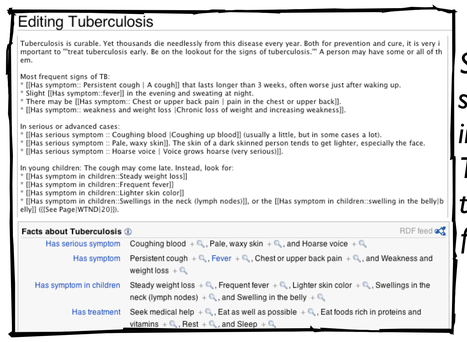
CS 294-10: Visualization
UC Berkeley ✤ Fall 2008
Matt Gedigian

Problem

Semantic wikis provide rich formal structure for content, but current software does not present this information in an accessible manner. How can such information be presented to reveal relationships and facilitate navigation?

Motivation

A semantic wiki is a wiki which utilizes an ontology to represent formalized knowledge. Current semantic wiki software leverages this representation primarily to support enhanced queries.



Semantic wikis offer a simple syntax for annotating formal properties (top). These are extracted (bottom) and made available for queries.

There are existing ontology visualization tools, but they are ill-suited for this application since they are disconnected from the article content of the wiki. Presenting a visualization alongside the wiki in the browser allows it be used as a navigation tool for someone exploring the site.

By constructing a visualization tailored to the particular characteristics of the domain, it is possible to derive useful information that is only implicit in the data set. In the case of the Hesperian Digital Library's healthcare domain, it would be useful to determine which diseases have similar causes, symptoms, or treatments. Such similarity calculations can be used to constrain the layout of a visualization.

Approach

Symptom Similarity

For each pair of diseases described on the Hesperian wiki we compare their sets of symptoms. The symptom property has string values, such as "Severe coughing" and "Mild fever" so simple natural language pre-processing is applied. After stemming and removing stop-words, we compute a score based on the number of tokens in common (Jaccard index). From these pair-wise property scores, we select the assignment that maximizes the overall similarity of the two diseases.

Layout Ordering

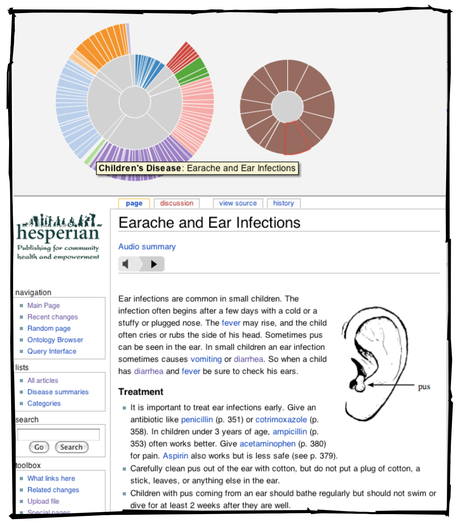
We seek a radial layout in which similar diseases are arranged close together. This objective corresponds to minimizing a cost function which incorporates the similarity scores and the radial distance between nodes. Due to the computational complexity of finding the optimal arrangement, simulated annealing is used to obtain an approximation.

Results

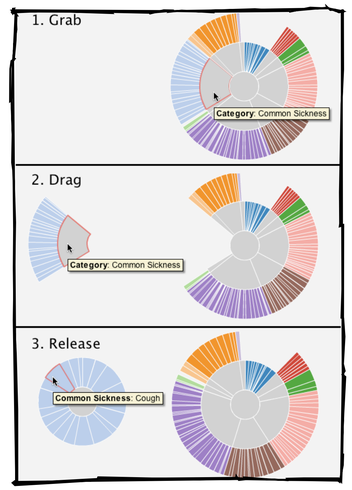
A space-filling radial layout is used to represent the hierarchical structure of diseases. The ordering is done such that diseases with similar symptoms are close together. Element sizes are scaled in proportion to the amount of information about them in the ontology.

Users can interactively manipulate the visualization. To focus in on a portion of the graph, they can remove a subtree from its parent to create a separate sunburst.

There is bidirectional communication between the visualization and the wiki. Clicking on nodes in the visualization loads the corresponding page in the wiki and navigating though the wiki highlights the corresponding node in the visualization.



Users can click nodes in the visualization to navigate to the corresponding page in wiki.



Users can drag subtrees to split them off into independent sunbursts.

Future Work

This work demonstrates the potential for integrating interactive visualizations with semantic wikis. This visualization could be improved by:

- including node labels
- incorporating search and filter controls
- allowing the user to select from different similarity metrics