Visualizing Etymology

Stanford University

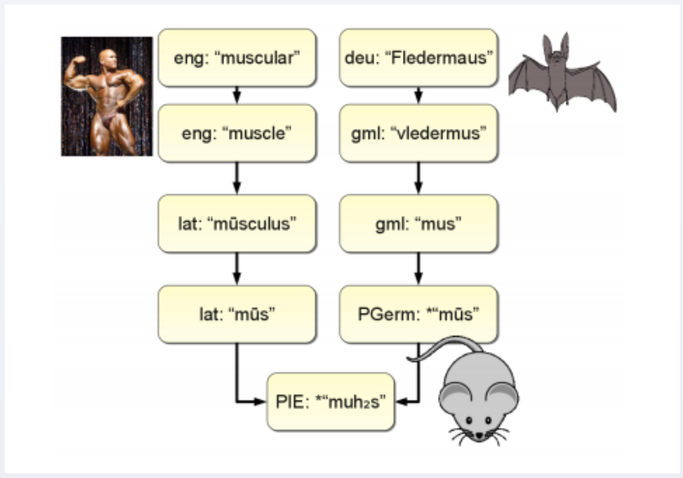
Chinmayi Dixit & Filippa Karrfelt

Stanford University

CS 448b Spring 2016

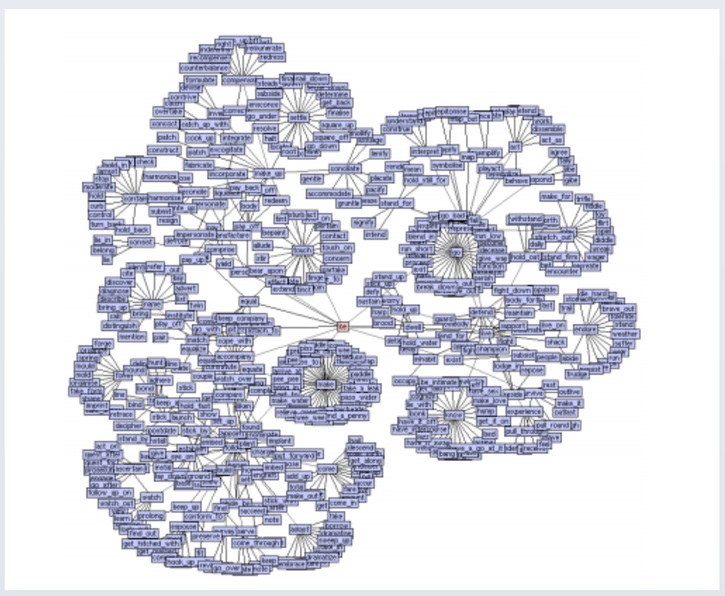
PROBLEM

There are many interfaces online for visualizing synonyms to words , but not many that show other word relations. There are extensive databases exploring etymological relations between words, but from what we have found, no visualizations that gives the user a tool for synthesizing and understanding. Our project seeks to address this problem; by visualizing etymology.



MOTIVATION

To visualize word relations is challenging since there are many elements to the data. In the visualizations available today, there are many things that could be improved to increase readability. Clutter is one thing that we have seen as a common factor among current visualizations, and something that is hard to improve because of the large data set, and its nominal properties. Word visualizations can really have a big difference in the understanding of etymology and language, and therefore we think this problem is interesting and worthwhile.



APPROACH

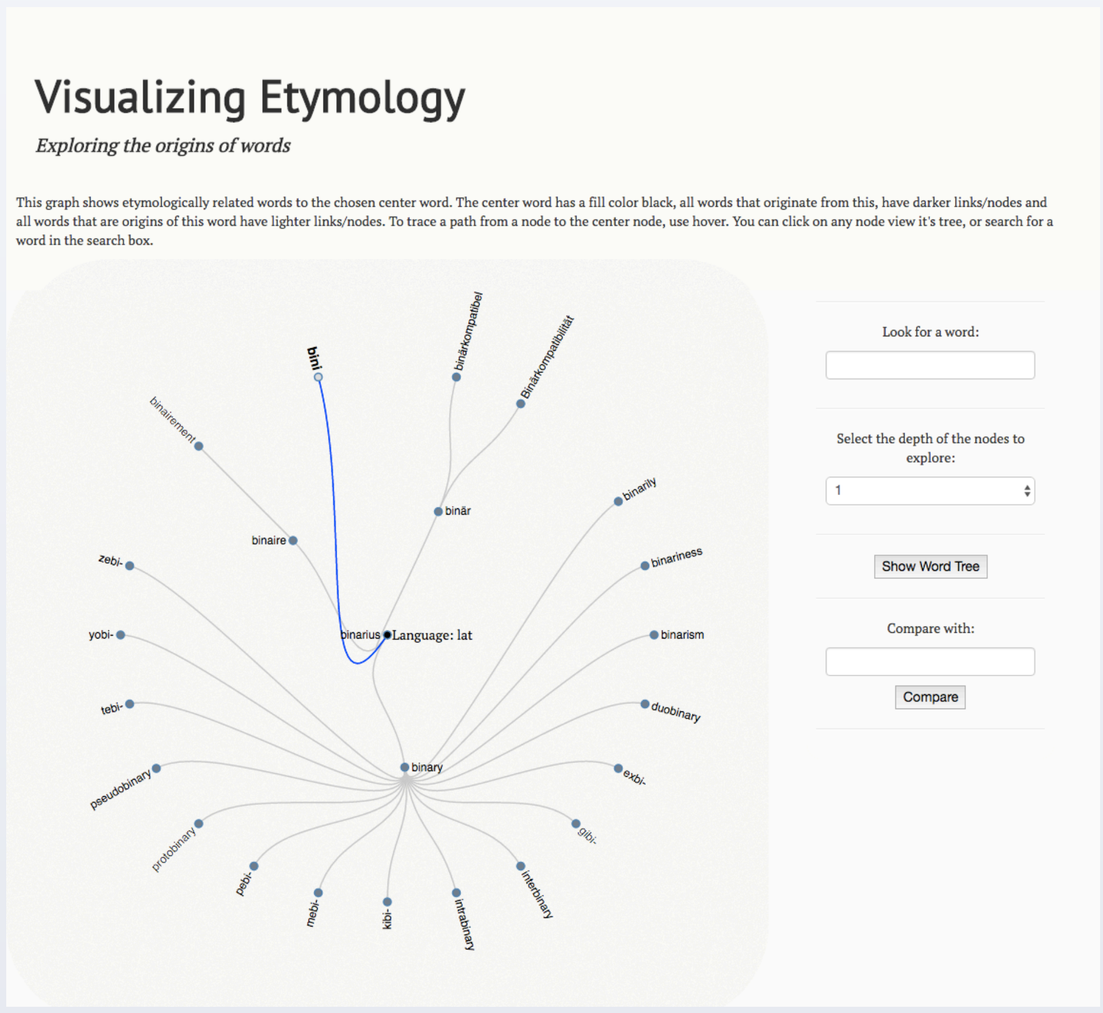
The data that we had was in csv format, but to effectively be able to handle the large data set, we wrote a script to put the data is JSON format. When finding the connections between words, we wrote the program to traverse the structure recursively, until it got to either a certain depth or to a certain word was found.

The manipulated data structure was then graphed with the d3 JavaScript library, using a combination of the bundle functionality and the tree layout.

RESULTS

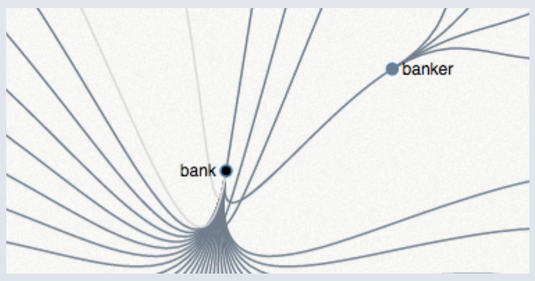
We create a radial bundled tree starting from a center node. The user can search for a new center node or choose one from on the screen. Each of the nodes have colors based on if they are a parent node or a child node. The links are colored according to this as well.

A user has the option to search for one word and explore it or choose two words and try to find a connection.



Functionality includes:

- Click a node to center around that node
- Search word with selected depth
- Compare two words
- Path highlighting on hover



- - Highlighting of center node
- - Edge bundling to avoid clutter

FUTURE WORK

There are possible extensions both when it comes to the visualization and the data. The work that can be done on the visualization is to make an algorithm to filter the related words if the word has too many connections. For the data set there is room for future work to include more etymological categories and maybe also include other components such as synonyms. This would create a more extensive visualization for that particular word.

RESEARCH POSTER PRESENTATION DESIGN © 2015
www.PosterPresentations.com