Presentation Topic: Group 5 - Chapter 15

Andrew Struthers

February 2023

Statements

1. The 6 visualization softwares presented are Scagnostics, VisDB, Hierarchical Clustering Explorer, PivotGraph, InterRing, Constellation.

2. **Scagnostics** is an idiom for scatter plot matrices. This is a way to categorize the shape of point distributions in each scatter plot. We can classify and associate data based off of the nine scagnostic measures.

3. **VisDB** treats an entire database as a very large table of data, then displays the data based off of a specific query that matches some subsets of the items. The visualization is determined by a series of attributes, like relevance to the query.

4. **Hierarchical Clustering Explorer (HCE)** is a systematic exploration of a multidimensional table with an associated hierarchical clustering. HCE uses interactive aggregation and filtering for the salable display of multidimensional tables. This visualization works on data with a multidimensional table with categorical and quantitative data, and shows how the categorical and quantitative data relate.

5. **PivotGraph** allows you to group nodes and links into a "roll-up" according to categorical attribute values. This allows us to compress many nodes of a graph down to very few notes, typically one node per each categorical attribute. The number of connections between two nodes in the typical graph gets compressed down and encoded by thicker or thinner lines to represent more or less connection displayed in the original graph.

6. **InterRing** is a way to radially visualize data, especially hierarchical data. This visualization only lets us visualize three different categorical attributes. The visualization uses color to encode the relationship of the parent nodes in the hierarchy to the children below it, by creating a gradient of color that impacts all children nodes.

7. **Constellation** supports browsing a complex, multilevel, linguistic network. This visualization has a very small audience, specifically computational linguistics who are trying to develop software involving linguistics. The high level Constellation spacial layout is based on a curvilinear grid, where the paths flow along a vertical column with the ordered word.

The horizontal order of the paths are determined by the plausibility attribute, where the left is the most plausible and the right is the least. This visualization helps show which words are associated with others, by vertically ordering and horizontally ordering the word relationships with other related words.

8. Case studies are a great way to break down and compare the visualization methods taught throughout the textbook

Questions

1. How are the quantitative attributes of a Scagnostic analysis calculated? For example, if you are viewing outliers, how far away does a datapoint need to be before a Scagnostic analysis considers it an outlier.

2. How does VisDB calculate relevance? How does the visualization actually provide any analysis? What does a database query look like in this case, and how does the query tie to the visualization?

3. Could the Constellation data visualization method be extended beyond linguistic uses and be adapted to show other relationships based off of the "closeness" of other nodes? For example, visualizing a family tree or the chances one person on a social media site knows (or is "related to") another user?