Typograph: Multiscale Spatial Exploration of Text Documents

Alex Endert, Russ Burtner, Nick Cramer, Ralph Perko, Shawn Hampton, Kristin Cook
Pacific Northwest National Laboratory
Challenge

How to visualize and explore large, text data?

- How to present a meaningful overview?
- How to maintain context during exploration?
- How to gradually increase detail upon demand?
Term-Centric Spatializations

[Cui et al., 2010]

[Kohonen et al., http://websom.hut.fi/websom/]
Generating a multi-scale spatialization that embeds multiple levels of detail (terms, phrases, snippets, and documents) in a single spatial layout

Preserving similarity through relative distances between information (i.e., similarity conveyed through relative proximity)
Generating the Visualization

Results:
- Terms: a single word
- Phrases: Two or more words in sequence
- Snippets: An extracted portion of text
- Articles: Full text document
How to present a meaningful overview?
How to maintain context during exploration?
How to gradually increase details upon demand?

- Zoom decreases level of aggregation within specified area
- Transition from document-centric to multiscale spatialization
- Maintain relative distances/locations to terms
Open Questions

▶ Data Coverage
  ▪ Visualize the data -> Visualize a model abstraction of the data?
  ▪ How much of the original data is \textit{represented}?

▶ Is an “overview” in the traditional sense meaningful?
  ▪ What does an overview of all the (English) text ever recorded look like?
  ▪ \textit{What does this mean for the Visualization Mantra?}
Thanks!

Alex Endert, Russ Burtner, Nick Cramer, Ralph Perko, Shawn Hampton, Kristin Cook

Pacific Northwest National Laboratory