Serendip
Topic Model-Driven Visual Exploration of Text Corpora

Eric Alexander, Department of Computer Science, University of Wisconsin-Madison
Joe Kohlmann, Department of Computer Science, University of Wisconsin-Madison
Robin Valenza, Department of English, University of Wisconsin-Madison
Michael Witmer, Folger Shakespeare Library in Washington, D.C.
Michael Gleicher, Member, IEEE, Department of Computer Science, University of Wisconsin-Madison

Three Views

- CorpusViewer: Re-orderable matrix
- TextViewer: Examination within one document
- RankViewer: Examine specific words

Features combatting scale

- Ordering
- Aggregation
- Annotation
- Assigning colors
- Details on demand

How are the factors for Serendipity implemented?

- Multiple access points: 3 Views, Ordered to user's liking
- Highlighting adjacencies: Through ordering and visualization
- Flexible pathways: Jumping between views
- Curiosity and playfulness: Interaction & Discovery

2 use cases

- Vis Abstracts
  1127 abstracts
  From SciVis, InfoVis, and visualization papers from 2007-2013
  each 30 to 389 words
- Early Modern Literature
  1060 digitized texts
  From English literature published between 1530 to 1799
  each few hundred words to few hundred pages

Promoting Serendipity

- Providing multiple access points
- Highlighting adjacencies
- Offering flexible pathways for exploration
- Enticing curiosity and playfulness

Discussion

- Strengths
  - Suitable for any document and corpus size
  - Three layers (whole corpus, single document, single words)
  - Simple but effective visualizations
  - Easily accessible (Online tool; though topic modelling part still due)

- Weaknesses
  - Not for quick exploration
  - Sceptical about serendipitous discoveries

Prepare

- System
- Data
- Tasks
- Encode
- Reduce
- Manipulate
- Scale

- Why
- What
- How
- Serendip
- Document collection (Text, Metadata, Topics)
- Explore document collections
- Facilitate serendipitous discoveries
- Reorderable matrix, line graph, bar graph
- Aggregation, Ordering
- Order, Color, Annotate, Details-on-Demand
- Up to ~1,000 Documents (various length)

Formal evaluation still due

To confirm serendipitous discoveries across multiple scales of data and abstraction
- Problem: How to evaluate serendipity?
- Long-term user studies needed