

Text visualization

Lucas Rizoli
CPSC 533C, November 2006

Web-pages
Email
Instant messages
Digitized books, articles

Why use text in visualization?



Reading is fast



[from <http://en.wikipedia.org/wiki/Image:Statue-Of-Liberty.jpg>]

Text can be a dense representation
Text can be inexact

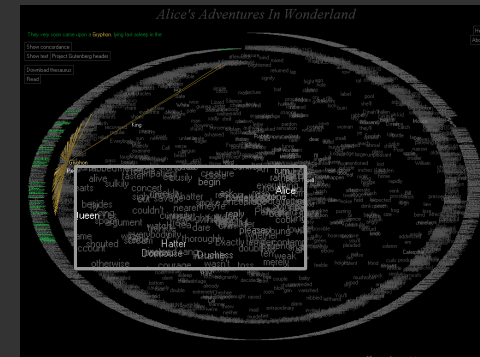
Fast
Dense
Inexact

Difficulties of using text

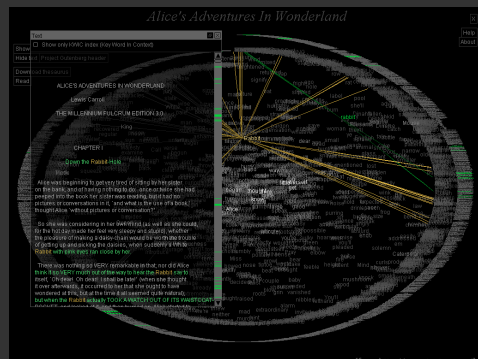
Space
Arrangement
Orientation
Legibility
Meaning



[from <http://www.thoughtbank.org/mindmappers/>]

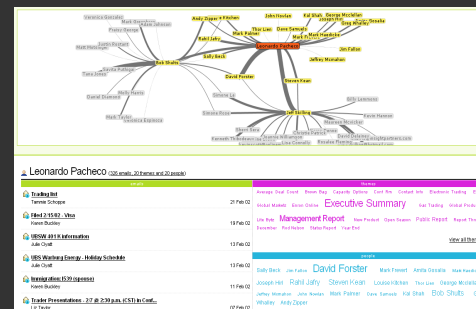


[from <http://www.textarc.org/>]

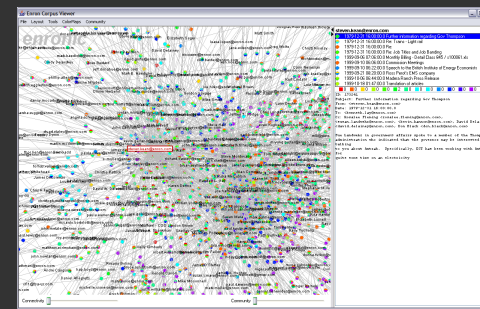


[from <http://www.textarc.org/>]

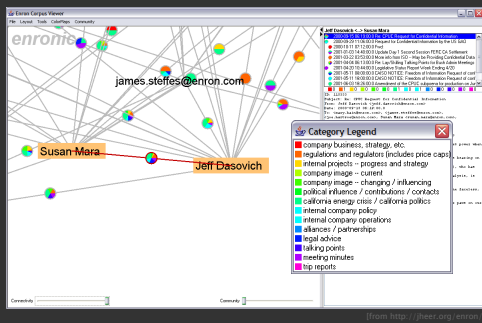
Index
Searching
Explicit in data



[from <http://enron.transparency.vicorio.com/>]

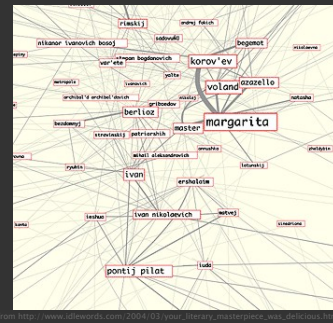


[from <http://jeffr.org/enron/>]



[from <http://jheer.org/enron/>]

17



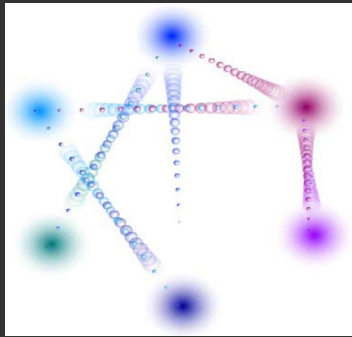
[from http://www.dbravada.com/2004/03/your_library_masterpiece_was_deleted.html]

18

Graph
Analyzing
Derived from data
Human supervision of automated processes

Reliance on meta-data
Says little about content

19



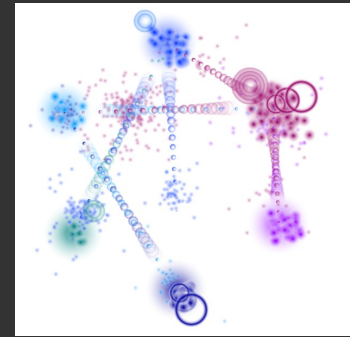
[from [Tab, A., & Carpendale, M. S. T. \(2002\)](http://www.dbravada.com/2004/03/your_library_masterpiece_was_deleted.html)]

21



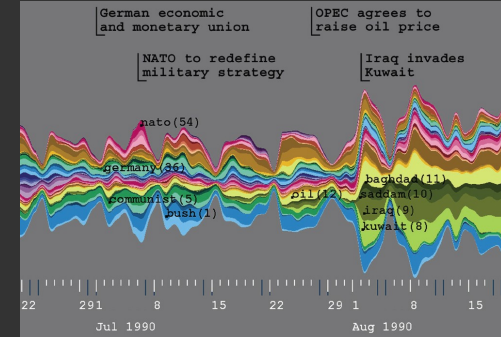
[from [Tab, A., & Carpendale, M. S. T. \(2002\)](http://www.dbravada.com/2004/03/your_library_masterpiece_was_deleted.html)]

22



[from [Tab, A., & Carpendale, M. S. T. \(2002\)](http://www.dbravada.com/2004/03/your_library_masterpiece_was_deleted.html)]

23



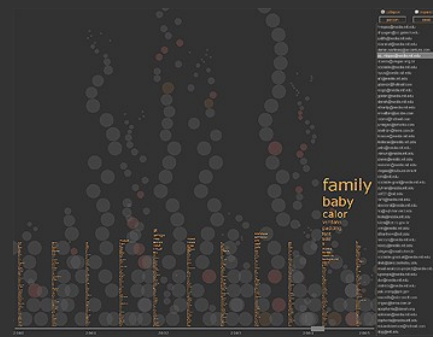
[from [Haars, S., Hertzler, E., Whitney, P., & Nowell, L. \(2002\)](http://www.dbravada.com/2004/03/your_library_masterpiece_was_deleted.html)]

20



[from [Viegas, F. B., Golden, S., & Donath, J. \(2006\)](http://www.dbravada.com/2004/03/your_library_masterpiece_was_deleted.html)]

25



[from <http://alumni.media.mit.edu/~viegas/projects/thewall/study/index.html>]

26



[from [Viegas, F. B., Golden, S., & Donath, J. \(2006\)](http://www.dbravada.com/2004/03/your_library_masterpiece_was_deleted.html)]

27

Trouble pre-processing data
Many assumptions made

Finding meaning in text is difficult

Adjusting for word frequency
Full semantic processing

Unique visual representation
Exploration
Derived from data
Increasingly semantic
Greater reliance on human users

Take-home lessons

29

30

31

32

Text in visualization

Fast, dense, inexact
Complicated to apply

33

Visualizing text

Range of levels and methods
Meta-data adds structure
Pre-processing is hard, important

34

- * Ceglowski, M. (2004). *Your Literary Masterpiece Was Delicious*. Retrieved November 6, 2006 from http://www.idlewords.com/2004/03/your_literary_masterpiece_was_delicious.htm
- * Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). ThemeRiver: Visualizing Thematic Changes in Large Document Collections. *Visualization and Computer Graphics, IEEE Transactions*, 8, 9-20.
- * Heer, J. (2004). *Exploring Enron: Visualizing ANLP Results*. Retrieved November 6, 2006 from <http://jheer.org/enron/w/>
- * Paley, W. B. (2002). TextArc: Showing Word Frequency and Distribution in Text. In Wong, P. C., & Keith Andrews (Eds.), *Proceedings of the IEEE Symposium on Information Visualization (Infovis '02) Poster Compendium*. Los Alamitos, CA, USA: IEEE Press.
- * Paley, W. B. (n.d.). *TextArc.org Home*. Retrieved November 6, 2006 from <http://textarc.org/>
- * Tat, A., & Carpendale, M. S. T. (2002). Visualizing Human Dialog. *Proceedings of the IEEE Conference on Information Visualization (Infovis '02)*. London, UK: IEEE Press.
- * Trampoline Systems (n.d.). *Trampoline Enron Explorer*. Retrieved November 6, 2006 from <http://enron.trampolinesystems.com>
- * Viégas, F. B., Colder, S., & Donath, J. (2006). Visualizing Email Content: Portraying Relationships from Conversational Histories. *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '06)*. Montréal, Québec, Canada: ACM Press.

35