

Exploring Emily Dickinson Letters

Catherine Plaisant, Nitin Madnani, Matt Kirschenbaum, Martha Nell Smith, Tanya Clement, Greg Lord



this project is part of the larger

nora project

It brings together humanists, computer scientists and library and information specialists

Institutions

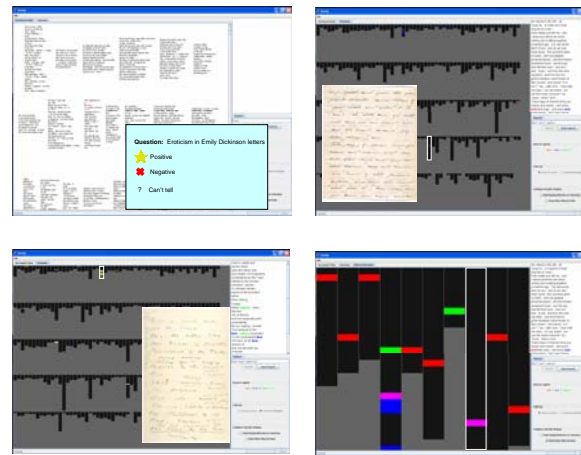
University of Illinois, Urbana-Champaign
 School of Library and Information Sciences
 National Center for Supercomputing Applications
 University of Maryland
 Human-Computer Interaction Laboratory
 Maryland Institute for Technology in the Humanities
 University of Georgia
 Department of English
 University of Virginia
 Department of Computer Science

18th and 19th Century British and American Literature

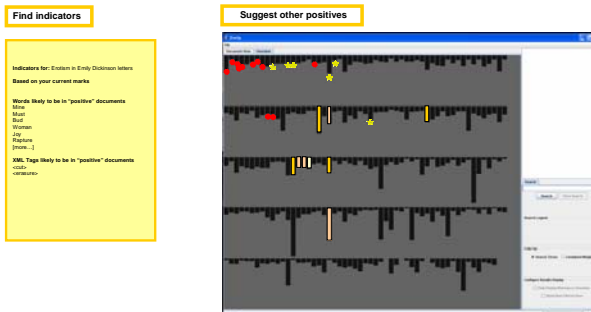
Guiding Principles

Web-deliverable services
 Visualizations as interface
 State of the art text mining
 Structured digital library data

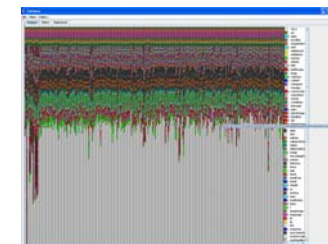
1 Emily: early exploratory prototype



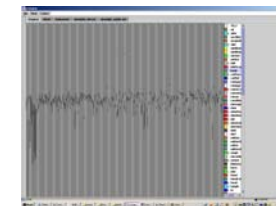
1. Identify question
2. Manual exploration
Search for indicators
(e.g. keywords, form of speech, tags, etc.)
3. Mark positives and negatives
4. Request suggested indicators
(naïve bayesian analysis)
5. Request suggestions
of other possible positives
6. Review and annotate



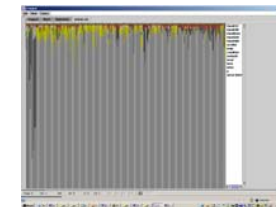
2 Using Compus [Fekete, 2000] to visualize the tagging structure of the collection <http://www.lri.fr/~fekete/compus/>



Overview of entire collection. The letters are ordered by time. The top corresponds to the header tags, i.e. the bulk of the XML document, which varies little between documents.



Filtering to show only the coverage of the <body> tag



After extracting new tags from attributes, we can reveal the names of persons that wrote on the letter, e.g. here yellow indicates the hand of Susan.

3 Possible next steps



Integration with InfoVis toolkit
(<http://ivtk.sourceforge.net>)

Cluster analysis (e.g. Ramsay et al. 2004)

Support for this project is provided by
The Andrew Mellon Foundation

