Developed by
nearly made all that were invested in the horse have
floridus
City
reticulata
Command line
cluster      word
horse
Detailed views of document summary

Representative Summarization Module
modules for experimentation and comparative performance study
• Variable summarization extent
• Support for summarization of individual documents or multiple
documents at once

*Tied to LAS Visualization Research (initial D3 experimentation)

Visualizing Summarized Text*
Case Study: Open Text Summarization (OTS)
• Analysts sift through many documents looking for info. of interest
• Reading text documents is a labor-intensive analyst activity
• A capability that effectively summarizes text could result in
considerable time savings to analysts
We desire an open design supporting various text summarization
modules for experimentation and comparative performance study

Single document summarization
Use case: analysts seek gist-level understanding of single document
• Detailed views of document summary

1963 MLK I Have A Dream Speech with 50% summarization

Weights are calculated for each word, derived from how
likely it is to exist in each cluster (topic)

Unknown documents are scored with the word/cluster
weights. The highest scoring clusters are putative categories for the
unknown document

Clustering accuracy can be assessed by calculating the max
F1 from the clusters to the articles’ category labels

Multi-document summarization
Use case: analysts rely on summarization for document triage
• Apply the same extent of summarization to all documents
• Less-detailed views of document summaries

97% summarization applied to a collection of different
documents (speeches)
Based on summaries, analysts can focus on
documents of interest.