Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw

Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Haesun Park, and John Stasko

University of Colorado, Adobe, Cornell University, Georgia Tech, Georgia Tech, Georgia Tech
Text documents are everywhere…

Academic Papers

Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw

- Carolyn Gao, Member, IEEE, Chunyi Lu, Jaypee Khem, Jagic Choi, Student Member, IEEE, Hanan Park, Member, IEEE, and John Saltz, Senior Member, IEEE

Product Reviews

2010 Hyundai Genesis Sedan - Consumer Reviews

Average Consumer Rating: 4.5 out of 5 stars

- Great value and excellent performance
- Very roomy for a standard model!
- Comfortable, no complaints there
- Value: 5 out of 5 stars
- Part-to-Drive: 5 out of 5 stars
- Interior Design: 5 out of 5 stars
- Build-Quality: 5 out of 5 stars

Health Forums

A safe and friendly online community where you can share experiences, get support and discuss autism

Police Reports
Investigative Analysis

• Assist analysts (police/intelligence, lawyers, reporters, scientists) confronting “information overload” problems while doing exploratory data analysis

• Given a collection of textual documents, assist analysts in
  1. Information foraging
  2. Sense-making
Two Types of Scenarios

1. Targeted analysis scenarios
2. Open ended, strategic analysis scenarios

Hourglass-shaped investigations

Top: lots of data
Middle: interesting lead
Bottom: lots of data

narrow scope of investigation
broaden scope of investigation
Text Analytic Approaches

Text Mining & Computational Analysis

Interactive Text Visualization

contribution of this paper

Integration

Visual Text Analytics
Evaluation Study

4 Settings

Paper

Desktop

Entity

Jigsaw

4 Strategies

1. Overview, filter & detail
   - Summarization
   - Doc Metrics
   - Doc Similarity
   - Recommend related docs

2. Build from detail

3. Hit the keyword
   - Doc Clustering

4. Find a clue and follow the trail

Computational support for hourglass-shaped investigation

**Top:** lots of data

- narrow scope of investigation

**Middle:** interesting lead

- broaden scope of investigation

**Bottom:** lots of data

- Document clustering
- Sentiment analysis
- Document similarity
- Recommending related documents
Computational Text Analyses

• Document Summarization
  (one sentence summaries, word clouds, and keyword summaries)

• Document Similarity
  (based on text or entities, cosine similarity)

• Document Clustering
  (based on text or entities, k-means clustering)

• Sentiment Analysis
  (hierarchical classification using Lingpipe and dictionary-based approach)
Document Summarization

- word clouds (multiple docs)
- one sentence summaries (single doc)
- keyword summaries (multiple docs)
Document Summarization

word clouds (multiple docs)

one sentence summaries (single doc)

keyword summaries (multiple docs)
Document Summarization

- **word clouds** (multiple docs)
- **one sentence summaries** (single doc)
- **keyword summaries** (multiple docs)
Document Clustering
Document Similarity & Document Metrics

Documents ordered and colored by similarity to selected reference document.

Documents ordered by date and colored by sentiment.
1. InfoVis and VAST papers from 1995 to 2011
2. Car reviews

Videos available in IEEE DL
Lessons learned

1. Make different computational analysis results available throughout the system in a variety of different contexts and views
2. Flexibly allow analysis output also to be used as input
3. Integrate different, independent computational analysis measures through interactive visualization to extend functionality and power
4. Provide computational support for both analysis directions: narrowing down as well as widening the scope of an investigation
5. Expose algorithm parameters in an interactive user accessible way
Contributions

1. Methods for fluidly integrating computational text analysis and visualization approaches

2. Illustrating of the utility through example usage scenarios

3. Lessons learned toward the design and construction of visual analytics systems for document exploration
Acknowledgments

IIS-091788
CCF-0808863

2009-ST-061-CI0001
Get your own copy (incl. InfoVis/VAST paper dataset)

www.jigsaw-analytics.net

Plug-in architecture for your visualizations and algorithms