

Combining Computational Analyses and Interactive Visualization to Enhance Information Retrieval

Carsten Görg, Jaeyeon Kihm, Jaegul Choo,
Zhicheng Liu, Sivasailam Muthiah,
Haesun Park, **John Stasko**

Information Interfaces Research Group
School of Interactive Computing
Georgia Institute of Technology

HCIR '10 Workshop

Focus: Exploratory Search

- Document collections (< 10k)
- Sense-making, understanding, browsing, foraging, triage, analysis, ...



Jigsaw

Visualization for Investigative Analysis across Document Collections

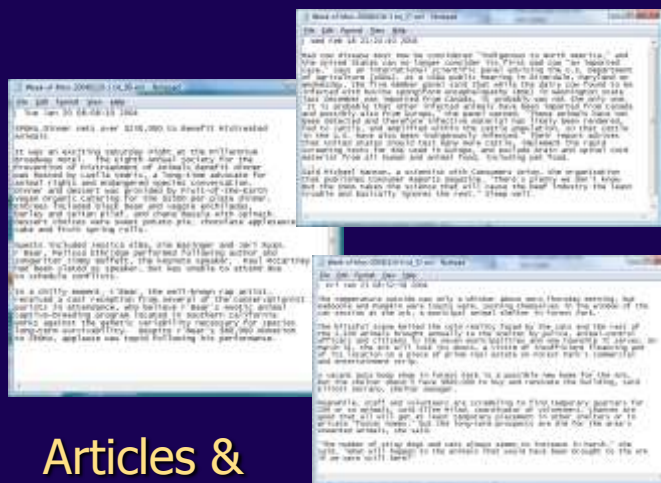
- Law enforcement & intelligence community
- Fraud (finance, accounting, banking)
- Academic research
- Consumer shopping

“Putting the pieces together”

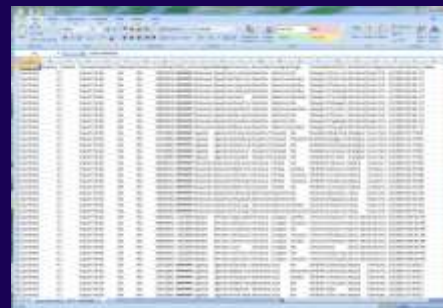


Problem Addressed

Help “investigators” explore, analyze and understand large document collections



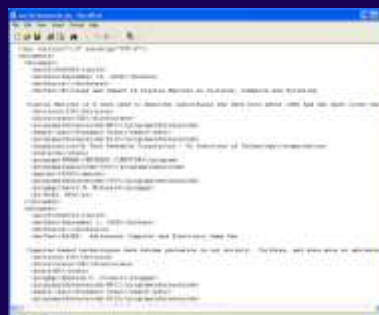
Articles & reports



Spreadsheets



Blogs



XML documents



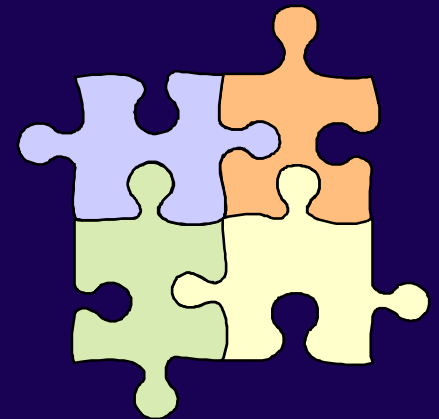
Our Focus

- Entities within the documents
 - Person, place, organization, phone number, date, email address, etc.
- Entities relate/connect to each other to make a larger “story”
- Connection definition:
 - Two entities are connected if they appear in a document together
 - The more documents they appear in together, the stronger the connection



Jigsaw

- Multiple visualizations (views) of documents, entities, & their connections
- Views are highly interactive and coordinated
- User actions generate events that are transmitted to and (possibly) reflected in other views



Computational Analyses

- Document summarization
- Document similarity
- Document clustering by content
 - Text or entities
- Sentiment analysis



System Views

The collage displays several distinct data visualization and analysis tools:

- Top Left:** A search interface for "JESAW" (Jawa Iqbal) with a color legend for categories like place, organization, person, date, country, and time. It includes a search bar and checkboxes for "Text search" and "Bibliographic search".
- Top Right:** A hierarchical tree view of a network graph, showing nodes grouped into categories like "Africa", "Asia", "Europe", and "North America".
- Middle Left:** A search results window for "Arrow" with a list of documents and a detailed view of a document from May 14, 2004, discussing Canadian immigration and environmental issues.
- Middle Center:** A large circular network graph with numerous nodes and edges, representing a complex system.
- Middle Right:** A smaller network graph with nodes and edges, possibly representing a different dataset or a zoomed-in view.
- Bottom Left:** A circular network graph with nodes arranged in a ring and connected by lines.
- Bottom Center:** A data table with multiple columns and rows, likely representing a dataset or a list of records.
- Bottom Right:** A network graph with nodes and edges, showing connections between various entities.



More Pixels Help





Paper's Example

- Car reviews from edmunds.com
 - Text: Review body text
 - Entities: Car features, make, model, numeric ratings (ride, power, overall), ...
 - 178 Reviews
 - Computational analyses took ~1 minute



Demo

- NSF grant info
 - Text: Title + abstract
 - Entities: PI, co-PI, organization, PM, keywords, amount, program, ...
- NSF > CISE > IIS
- 2005-2010
 - 2,070 awards
- Computational analysis took ~90 minutes



Other Domains

- Intelligence & law enforcement
 - Police cases
 - Won 2007 VAST Contest
- Academic papers, PubMed
 - All InfoVis & VAST papers
 - CHI papers
- Grants
 - NSF CISE awards from 2000-now
- Topics on the web (medical condition)
 - Autism
- Consumer reviews
 - Amazon product reviews, edmunds.com, tripadvisor.com
- Business Intelligence
 - Patents, press releases, corporate agreements, ...



To Learn More & Availability

<http://www.gvu.gatech.edu/ii/jigsaw>

Available for (free)
trial use

Send email to:
stasko@cc.gatech.edu



Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)



- Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788



End

- Thanks for your attention!
- Questions?

