

# Visual Analytics 2

---



CS 7450 - Information Visualization  
March 31, 2011  
John Stasko

## Agenda

---

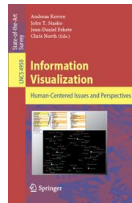


- Last time
  - Overview of what the term means and how it relates to information visualization
  - Some example VA research projects
- Today
  - Specific example, Jigsaw, helping investigative analysis
  - Related systems

## VA Definition



- Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets



Keim et al, chapter in  
*Information Visualization:  
Human-Centered  
Issues and Perspectives*, 2008

## Application Area



- Investigative & Intelligence Analysis
  - Gather information from various sources then analyze and reason about what you find and know
  - Analyze situations, understand the particulars, anticipate what may happen



# Definitions



- Thinking<sup>1</sup> - or reasoning - involves objectively connecting present beliefs with evidence in order to believe something else
- Critical Thinking<sup>1</sup> is a deliberate meta-cognitive(thinking about thinking) thinking act whereby a person reflects on the quality of the reasoning process simultaneously while reasoning to a conclusion.
- Intelligence<sup>1</sup> is a specialized form of knowledge, an activity, and an organization. As knowledge, intelligence informs leaders, uniquely aiding their judgment and decision-making. ...

1. *Critical Thinking and Intelligence Analysis: David Moore*

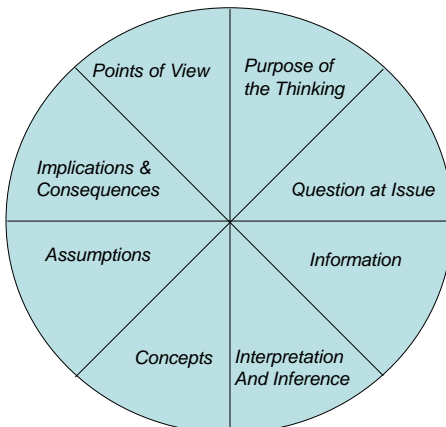
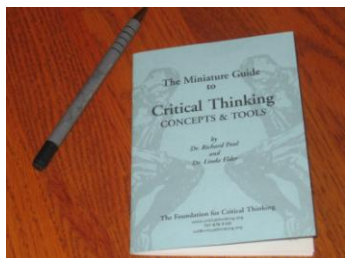


# Critical Thinking\*



**“...the quality of our life and that of what we produce, make, or build depends precisely on the quality of our thoughts.”**

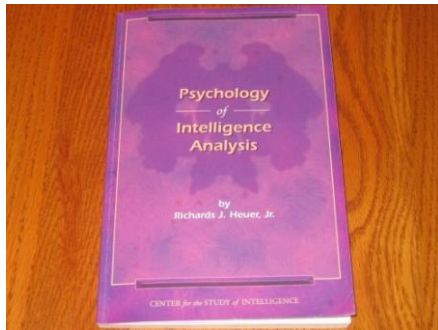
**Elements of thought:**



\* Foundations of Critical Thinking [www.criticalthinking.org](http://www.criticalthinking.org)



# Example: Heuer's Central Ideas



- “Tools and techniques that gear the analyst’s mind to apply higher levels of critical thinking can substantially improve analysis... structuring information, challenging assumptions, and exploring alternative interpretations.”

CS 7450

7

## Intelligence Process

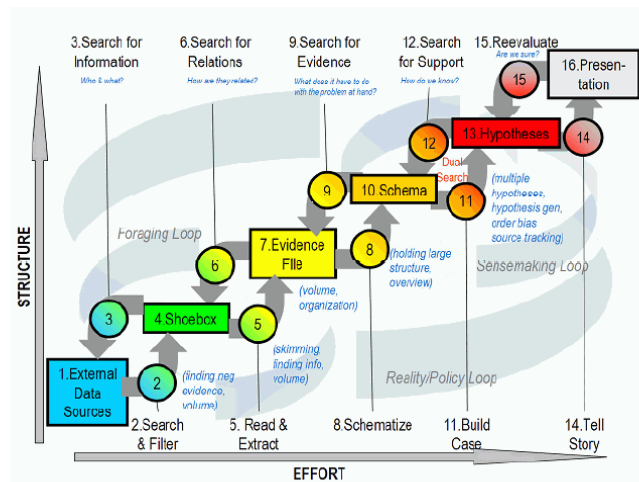


Figure 2.1. Notional model of sensemaking loop for intelligence analysis derived from CTA.

Pirolli & Card  
Int'l Conf Intelligence Analysis '05

Spring 2011

CS 7450

8

## Pain Points



- Cost structure of scanning and selecting items for further attention
- Analysts' span of attention for evidence and hypotheses

## HW 8 – Investigative Analysis



- Did you encounter those pain points?
- Discuss
  - How did you work on the problem?
  - What were the main challenges?

# Jigsaw



## Visualization for Investigative Analysis across Document Collections

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



**"Putting the pieces together"**

Spring 2011

CS 7450

11

# The Jigsaw Team



## Current:

Carsten Görg  
Zhicheng Liu  
Youn-ah Kang  
Jaeyeon Kihm

and many alumni

Spring 2011

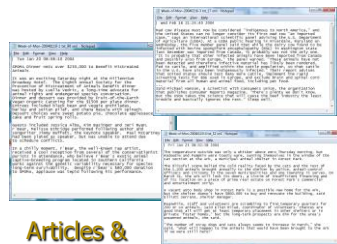
CS 7450

12

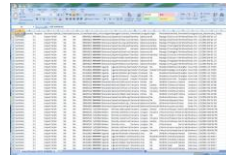
# Problem Addressed



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



Articles & reports



Spreadsheets



Blogs



XML documents

Spring 2011

CS-7450

13

# Our Focus

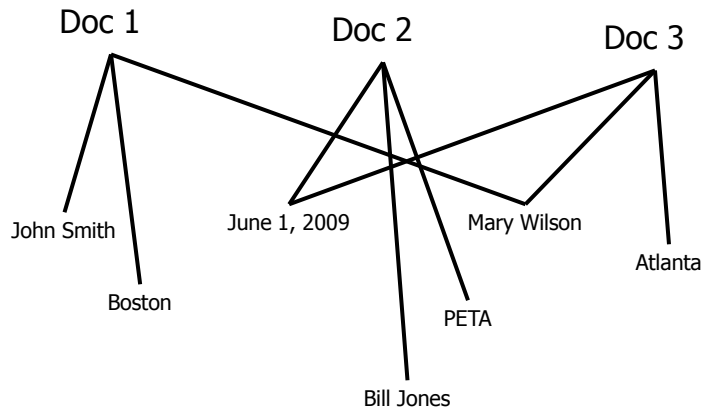


- Entities within the documents
  - Person, place, organization, phone number, date, license plate, etc.
- Thesis: A story/narrative/plot/threat within the documents will involve a set of entities in coordination

Spring 2011

CS 7450

14



Spring 2011

CS 7450

15

## Entity Identification



- Must identify and extract entities from plain text documents
  - **Crucial for our work**
- Not our main research focus – We use tools from others

Spring 2011

CS 7450

16



# Sample Document



Report: 20040510-4\_16  
May 14 2004

VANCOUVER, British Columbia - A Canadian immigration panel is considering whether accused environmental saboteur Tre Arrow can apply for refugee status in Canada.

Arrow, 30, who is wanted for fire bombing logging and cement trucks in Oregon, asked the Canadian authorities to remain in Canada as a political refugee at a hearing in Vancouver on Tuesday.

A key issue will be whether Arrow is affiliated with a terrorist group, which would immediately disqualify him from receiving refugee status in Canada, authorities said.

The Immigration and Refugee Board is scheduled to decide by May 31 whether Arrow is affiliated with the Earth Liberation Front, a group the FBI considers a terrorist organization responsible for scores of attacks on property over the past dozen years.

Spring 2011

CS 7450

17

# Entities Identified



**Source:**

**Date:** May 14, 2004

VANCOUVER, British Columbia - A Canadian immigration panel is considering whether accused environmental saboteur Tre Arrow can apply for refugee status in Canada.

Arrow, 30, who is wanted for fire bombing logging and cement trucks in Oregon, asked the Canadian authorities to remain in Canada as a political refugee at a hearing in Vancouver on Tuesday.

A key issue will be whether Arrow is affiliated with a terrorist group, which would immediately disqualify him from receiving refugee status in Canada, authorities said.

The Immigration and Refugee Board is scheduled to decide by May 31 whether Arrow is affiliated with the Earth Liberation Front, a group the FBI considers a terrorist organization responsible for scores of attacks on property over the past dozen years.

Spring 2011

CS 7450

18

## Sample Document 2



Title: Proving Columbus was Wrong  
Abstract: In this work, we show the world is really flat. To do this, we build a bunch of ships. Then we...  
PI: Amerigo Vespucci  
Co-PI: Vasco de Gama, Ponce de Leon  
Organization: Northwest Central Univ.  
Amount: 123,456  
Program Mgr: Ephraim Glinert  
Division: IIS  
ProgramElementCode: 2860

Spring 2011

CS 7450

19

## Entities Already Identified



Title: Proving Columbus was Wrong  
**Abstract:** In this work, we show the world is really flat. To do this, we build a bunch of ships. Then we...

---

**PI:** Amerigo Vespucci  
**Co-PI:** Vasco de Gama, Ponce de Leon  
**Organization:** Northwest Central Univ.  
**Amount:** 123,456  
**Program Mgr:** Ephraim Glinert  
**Division:** IIS  
**ProgramElementCode:** 2860

**Entities**

Spring 2011

CS 7450

20

# Connections



- 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

Spring 2011

CS 7450

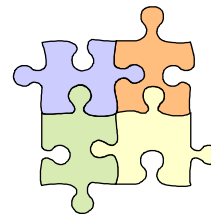
21

# Jigsaw

“Putting the pieces together”



- 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

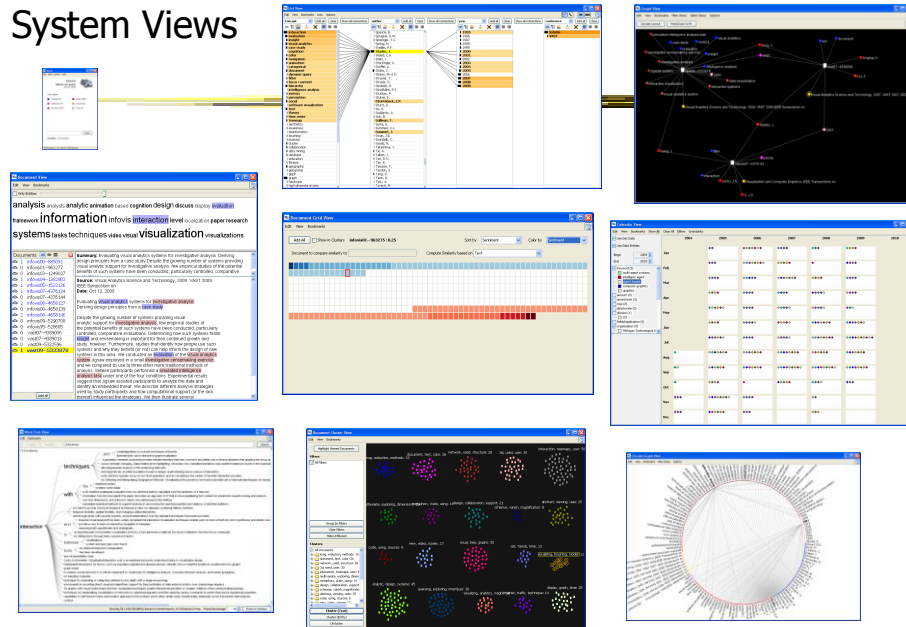


Spring 2011

CS 7450

22

# System Views



Spring 2011

CS 7450

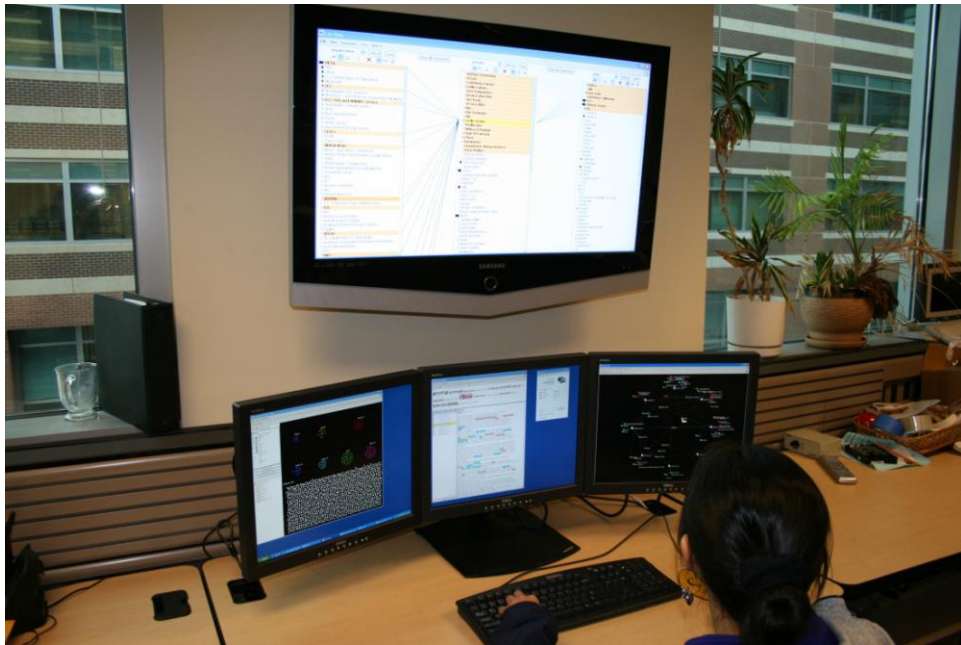
23



Spring 2011

CS 7450

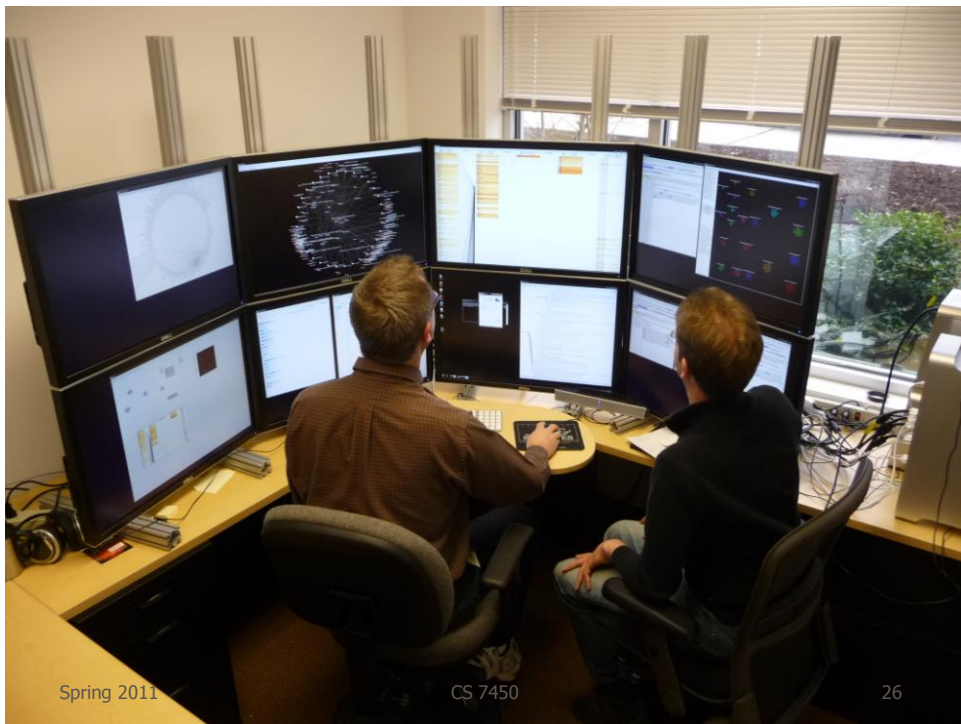
**Pixels Help**



Spring 2011

CS 7450

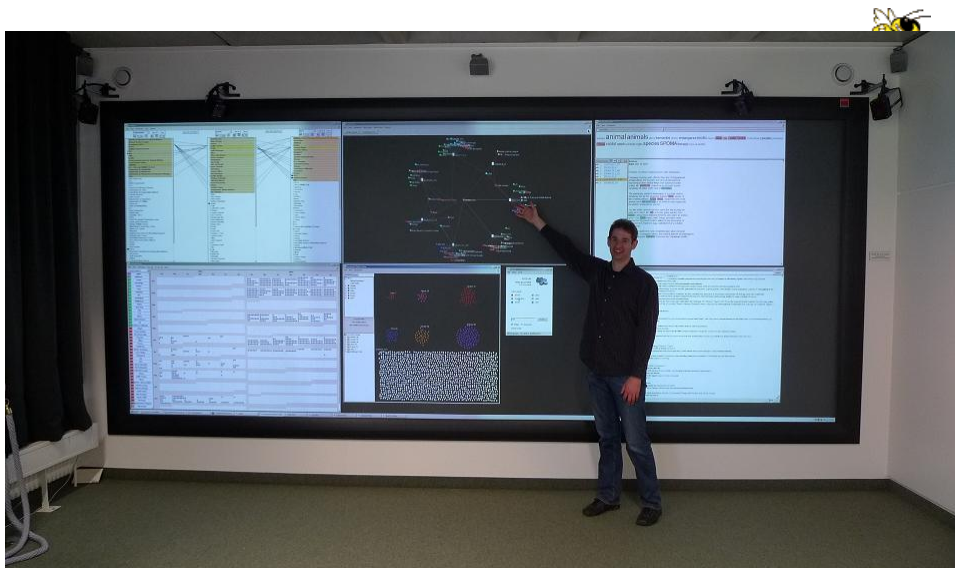
25



Spring 2011

CS 7450

26



Spring 2011

CS 7450

27

## Demo 1

- Data from HW 8
- Let's find the bad guys!

Spring 2011

CS 7450

28

## Demo 2



- Car reviews
  - Text: Consumer's comments
  - Entities: Various ratings (1-10), car features, other makes & models

Spring 2011

CS 7450

29

## Console



Entity types

The screenshot shows the Jigsaw application window. The title bar reads "Jigsaw" and the menu bar includes "File", "Views", "Entities", and "Tools". The main content area displays "JIGSAW infovis-vast" with a puzzle piece icon and "512 documents". Below this is a "Color Legend:" section with two columns of colored squares and labels: "author (1017)" (red), "concept (77)" (blue), "conference (2)" (magenta), "indexterm (1790)" (yellow), "journal (17)" (brown), "keyword (1202)" (pink), and "year (16)" (grey). An orange arrow points from the text "Entity types" to the "author (1017)" entry. At the bottom, there is a search bar with a "Search" button and two checkboxes: "Entities" (checked) and "Documents" (unchecked). The status bar at the very bottom says "Workspace: no active workspace".

Spring 2011

CS 7450

30



# Document View

Document View

analysis analysts analytic animation based cognition design discuss display evaluation

framework information infovis interaction level localization paper research

systems tasks techniques video visual visualization visualizations

Documents

- 1 infovis00-885091
- 0 infovis01-963277
- 0 infovis03-1249027
- 1 infovis04-1362902
- 1 infovis05-1532136
- 1 infovis07-4376134
- 0 infovis07-4376144
- 2 infovis08-4658127
- 0 infovis08-4658139
- 2 infovis08-4658146
- 0 infovis09-5290708
- 0 infovis95-526695
- 0 vast07-4389006
- 0 vast07-4389013
- 0 vast09-5332596
- 1 vast09-5333878

Summary: Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative

Source: Visual Analytics Science and Technology, 2009. VAST 2009. ICCC Symposium on

Date: Oct 12, 2009

Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study

Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative evaluations. Determining how such systems foster insight and sensemaking is important for their continued growth and study, however. Furthermore, studies that identify how people use such systems and why they benefit (or not) can help inform the design of new systems in this area. We conducted an evaluation of the visual analytics system Jigsaw employed in a small investigative-sensemaking exercise, and we compared its use to three other more traditional methods of analysis. Sixteen participants performed a simulated intelligence analysis task under one of the four conditions. Experimental results suggest that Jigsaw assisted participants to analyze the data and identify an embedded threat. We describe different analysis strategies used by study participants and how computational support (or the lack thereof) influenced the strategies. We then illustrate several

Important words in loaded docs

Automatic summary

Entities identified

Spring 2011

CS 7450

31

# List View

Lists of entities by type  
Connections highlighted

List View

Concept

- interaction
- evaluation
- insight
- visual analytics
- case study
- cognition
- color
- navigation
- animation
- categorization
- document
- dynamic query
- filter
- focus+context
- hierarchy
- intelligence analysis
- metrics
- perception
- social
- software visualization
- text
- theory
- time series
- treemap
- awareness
- biometrics
- brushing
- business
- cluster
- collaboration
- data mining
- database
- education
- ethics
- geographic
- graph
- hardware
- high-dimensional data

author

- I Spence, B.
- I Sprague, D.W.
- I Sreenager, T.C.
- I Spring, N.
- I Stadler, P.F.
- I Stasko, J.
- I Stead, C.A.
- I Stein, C.
- I Stodolinger, K.
- I Stoffel, A.
- I Stroh, C.
- I Storey, M.-A.D.
- I Sutherland, T.
- I Strayer, D.
- I Strobel, H.
- I Strothmann, P.J.
- I Studley, P.
- I Sulner, F.
- I Sturtebeck, E.P.
- I Shurtz, D.
- I Su, H.
- I Sudianto, A.
- I Suh, B.
- I Sullivan, T.
- I Suma, E.
- I Summers, K.L.
- I Suresh, J.
- I Swan, J.E.
- I Swindell, C.
- I Srinivas, N.
- I Takekuma, Y.
- I Tai, A.
- I Talbot, J.
- I Tan, D.S.
- I Tan, R.
- I Tanase, T.
- I Tandon, S.
- I Tang, D.
- I Tamm, E.
- I Tatu, A.
- I Tavaris, M.

year

- 1995
- 1996
- 1997
- 1998
- 1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009

conference

- infovis
- vast

Spring 2011

CS 7450

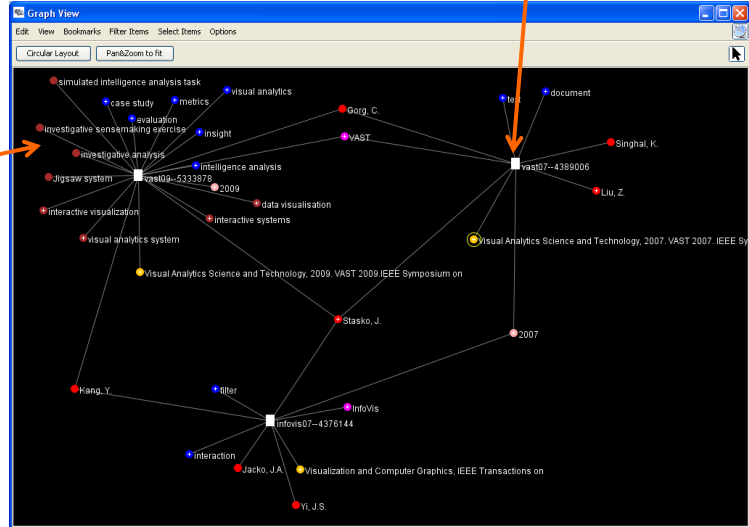
32



# Graph View

Document

Entities



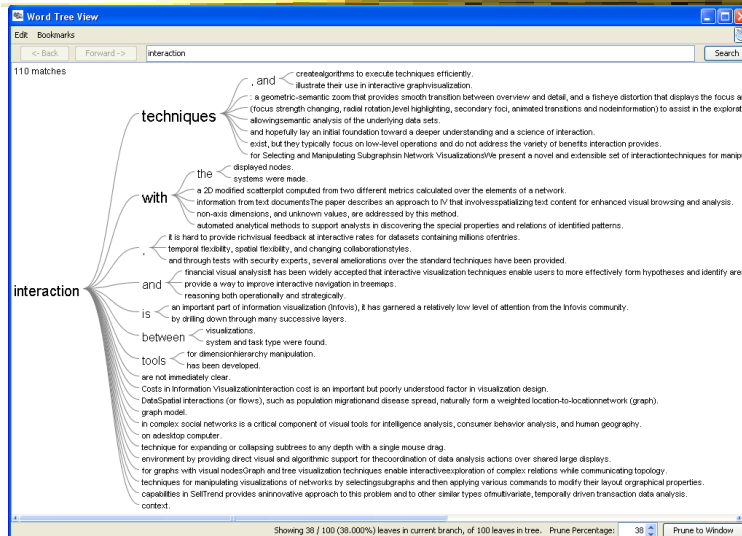
Spring 2011

CS 7450

33

# WordTree View

Context of a word in the collection



Spring 2011

CS 7450

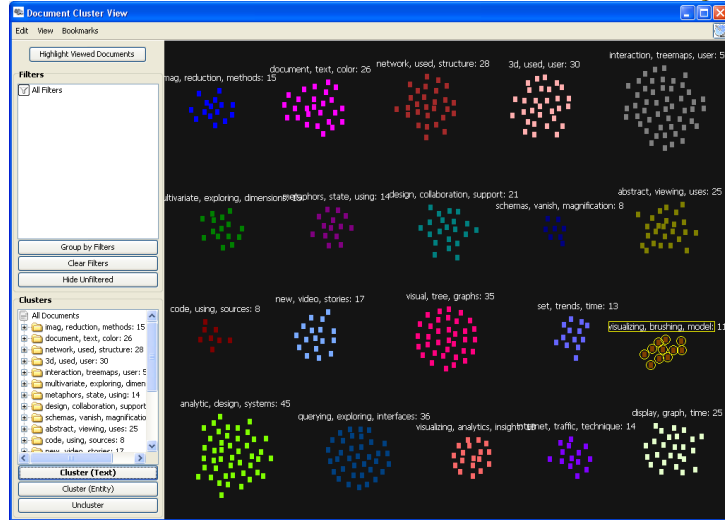
34

# Document Cluster View



Clustered by document text or by entities

Summarized by three words



Spring 2011

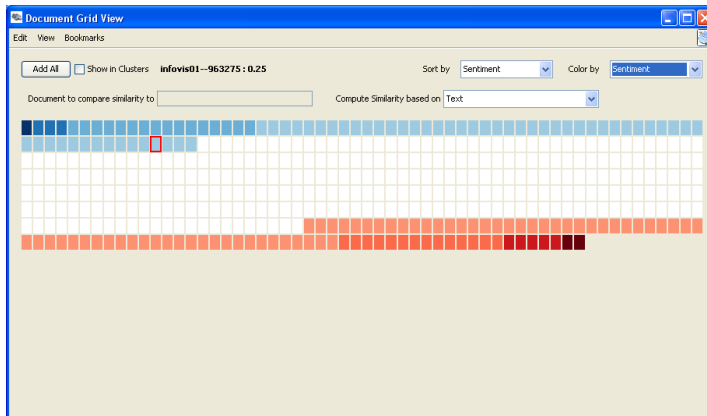
CS 7450

35

# Document Grid View



User controls order and color



Sentiment analysis shown here

Spring 2011

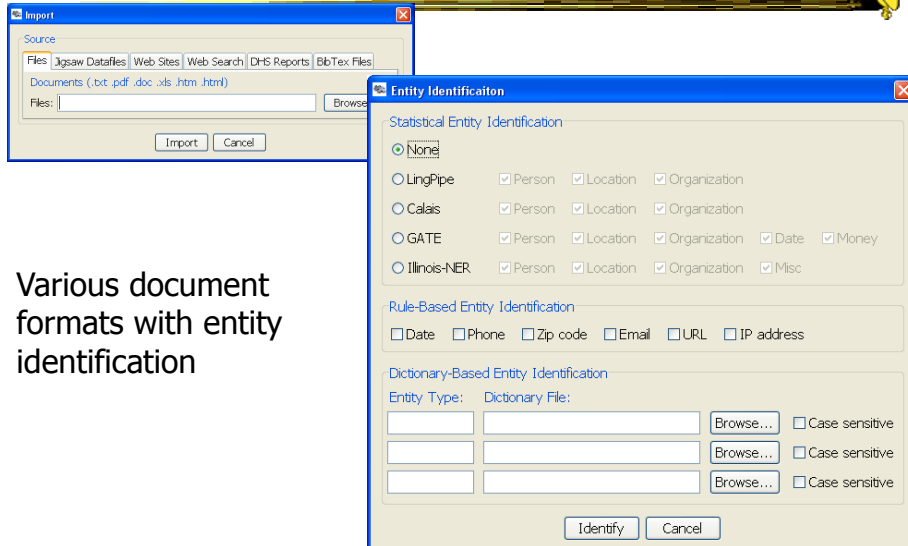
CS 7450

36





# Document Import



Various document formats with entity identification

Spring 2011

CS 7450

41

# Input Data Formats



- Text, pdf, Word, html, Excel
- Jigsaw data file format
  - Our own xml
- DB?
  - Go to Excel
  - Go to text, transform to Jigsaw data file

Spring 2011

CS 7450

42

```

<award>
<awardnumber>0640291</awardnumber>
<title>SGER: Distributed Spatial Partitioning Algorithms for Scalable Processing of Mobile
<nsfororganization>IIS </nsfororganization>
<programs>DATA MANAGEMENT SYSTEMS</programs>
<startdate>September 1, 2006</startdate>
<lastamendmentdate>September 12, 2007</lastamendmentdate>
<principalinvestigator>Liu, Ling</principalinvestigator>
<state>GA</state>
<organization>GA Tech Research Corporation - GA Institute of Technology </organizatio
<awardinstrument>Standard Grant </awardinstrument>
<programmanager>Le Gruenwald </programmanager>
<expirationdate>February 29, 2008</expirationdate>
<awardedamounttodate>65502</awardedamounttodate>
<co_pinames></co_pinames>
<piemailaddress>lingliu@cc.gatech.edu
<organizationstreetaddress>Office of Sponsored Programs </organizationstreetaddress>
<organizationcity>Atlanta </organizationcity>
<organizationstate>GA</organizationstate>
<organizationzip>30332</organizationzip>
<organizationphone>4048944819</organizationphone>
<nsfdirectorate>CSE </nsfdirectorate>
<programelementcodes>7485</programelementcodes>
<programreferencecodes>HPCC|9218|7484</programreferencecodes>
<fieldofapplications>0104000 Information Systems </fieldofapplicati
<awardnumber>0640291</awardnumber>
<abstract>IIS-0640219 Ling Liu <lt;lingliu@cc.gatech.edu>gt; Georgia Institute of Instit
</award>

```

## Scraped XML

Spring 2011

CS 7450

43

```

<document>
<docID>0808863</docID>
<docDate>July 1, 2008</docDate>
<docSource></docSource>
<docText>FODAVA-Lead: Dimension Reduction and Data Reduction: Foundations for Visualization

FODAVA-Lead: Dimension Reduction and Data Reduction: Foundations for Visualization The FODAVA (Foundations of
Data Analysis and Visualization) Lead research team at the Georgia Institute of Technology provides unified
expertise in the critical areas for providing leadership of the FODAVA effort, including machine learning and
computational statistics, information visualization, massive-dataset algorithms and data structures, and
optimization theory. The team is focused on the fundamental theory and approaches to make breakthroughs in data
representations and transformations. The work is directed along the two main axes of scale reduction, data reductio
<directorate>CSE</directorate>
<award-instr>Continuing grant</award-instr>
<programreferencecode>HPCC</programreferencecode>
<programreferencecode>9218</programreferencecode>
<keyword>visualization</keyword>
<keyword>algorithms</keyword>
<fieldofapplication>0000912 Computer Science</fieldofapplication>
<state>GA</state>
<organization>GA Tech Research Corporation - GA Institute of Technology</organization>
<keyword>data analysis</keyword>
<keyword>information visualization</keyword>
<keyword>machine learning</keyword>
<amount>1200000</amount>
<pi>Park, Haesun</pi>
<copi>John Staasko</copi>
<copi>Alexander Gray</copi>
<copi>Renato D. C. Monteiro</copi>
<copi>Vladimir Koltchinskii</copi>
<progmgr>Lawrence Rosenblum</progmgr>
<division>CCF</division>
<keyword>visual analytics</keyword>
<programelementcode>I114</programelementcode>
<programelementcode>H194</programelementcode>
</document>

```

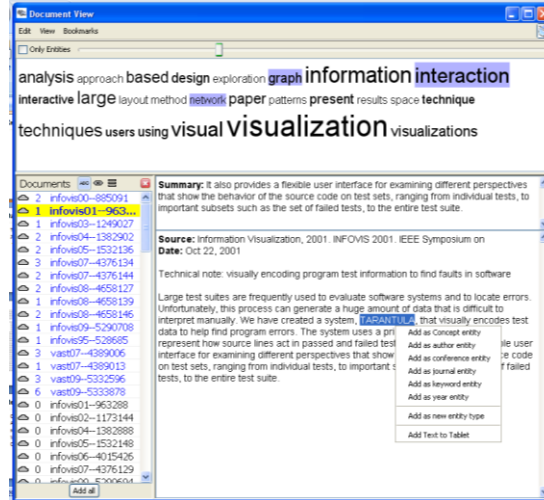
## Jigsaw Datafile Format

Spring 2011

CS 7450

44

# EI Correction

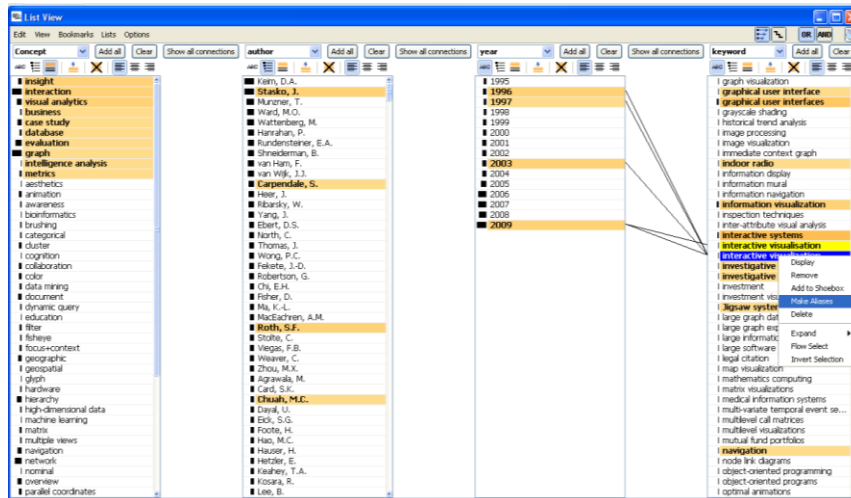


Spring 2011

CS 7450

45

# Entity Aliasing

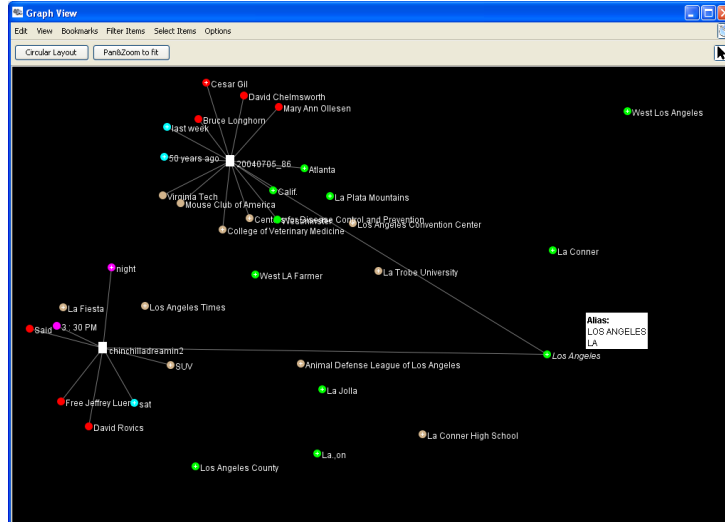


Spring 2011

CS 7450

46

# Alias Representation

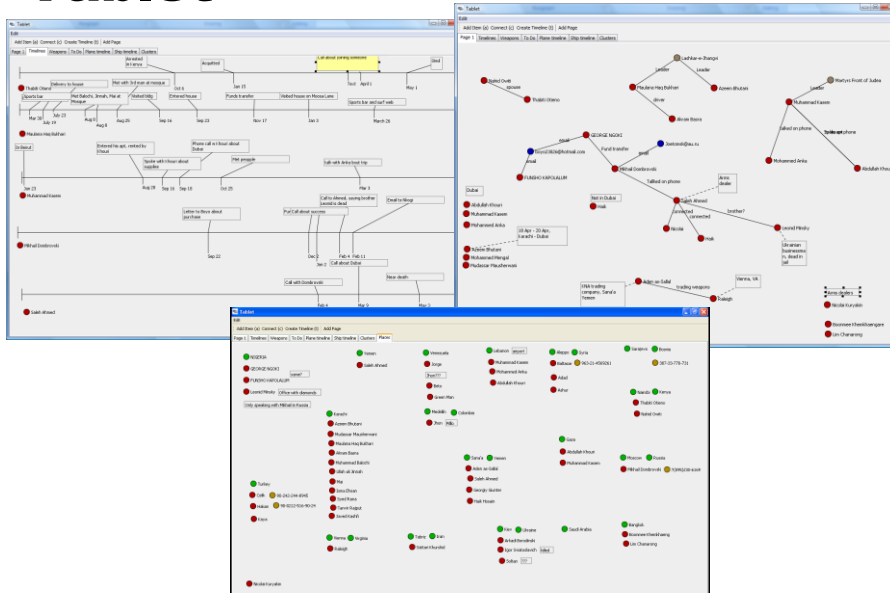


Spring 2011

CS 7450

47

# Tablet



Spring 2011

CS 7450

48



# Application Domains



- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES '10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, tripadvisor.com
  - Görg et al, HCIR '10
- Business Intelligence
  - Patents, press releases, corporate agreements, ...
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis '10

Spring 2011

CS 7450

49

# Potential Jigsaw Future Work



- Collaborative capabilities
- Improved evidence marshalling
- Present/browse investigation history
- Scalability upward
- Web document ingest
- Implement network algorithms
- DB import
- Wikipedia & Intellipedia
- Geospatial view
- Better timeline capabilities
- Reliability/uncertainty
- Other types of data
- Active crawling/RSS ingest
- Try it on display wall
- Deployment to real clients

Spring 2011

CS 7450

50

# Room to Improve



- What Jigsaw doesn't do so well now
  - The end-part of the Pirolli-Card model
    - Helping the analyst take notes, organize evidence, generate hypotheses, etc.  
(The Tablet is a first step)
  - Sometimes called "evidence marshalling"
  - Others have focused more on that aspect...

Spring 2011

CS 7450

51

# i2's Analyst Notebook



The screenshot shows the i2 Analyst Notebook website. At the top, there is a navigation bar with links for 'Contact Us', 'Download Center', 'Info Request', and a search box. Below this is a secondary navigation bar with 'Home', 'Company', 'Products', 'Solutions', 'Services', 'Partners', and 'Support'. The main content area is titled 'Products' and features a section for 'i2 Analyst's Notebook Powering Analysis'. This section includes a paragraph describing the software's ability to handle large volumes of raw data and generate actionable intelligence. Below the text are four small images illustrating different features: visualizing large volumes of data, revealing hidden connections, creating analytical charts, and communicating complex cases. A sidebar on the right lists various products and solutions, including 'Analytical Capabilities', 'Online Data Analysis', 'Information Sharing', 'Analytical Notebook SDK', 'Solutions', 'iBase', 'iBridge', 'ChartReader', 'PatternTracer', 'TextChart', 'ChartExplorer', 'Analyst's Workstation', 'iVR', and 'iXA'. The bottom of the page has a 'Done' button.

Spring 2011

CS 7450

52

# Analyst's Notebook



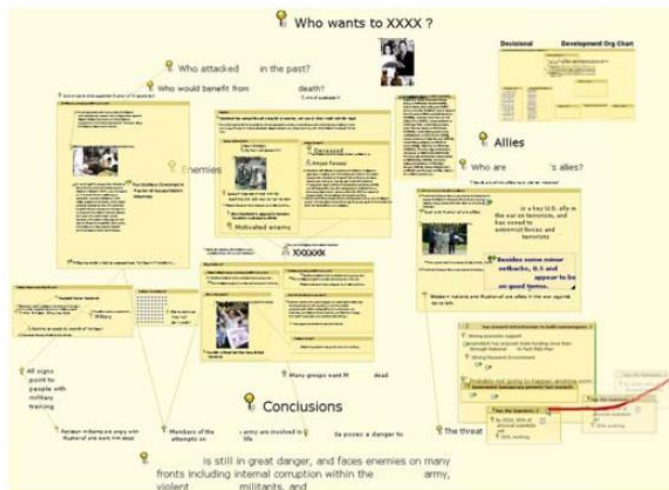
- Leading commercial tool in this space (law enforcement and intelligence agencies)
- Large zooming workspace where analyst creates networks of entities and notes
- Often used to produce presentation or story of analysis done

Spring 2011

CS 7450

53

# Oculus' Sandbox



Video

Wright et al  
CHI '06

Spring 2011

CS 7450

54

# Sandbox



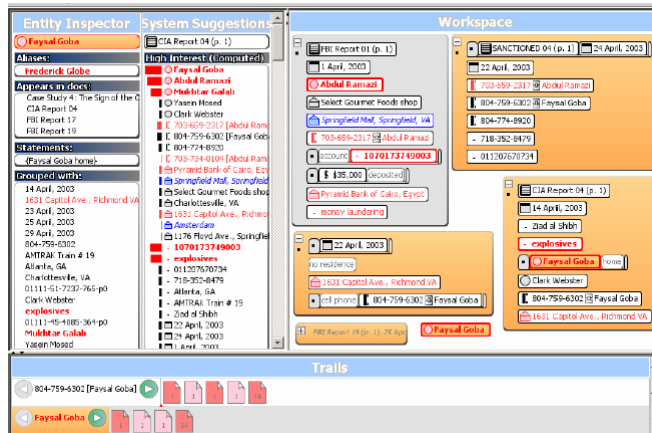
- Flexible space for inserting text and graphics
- Objects can be dragged-and-dropped from their other analysis tools
- Flexible level of detail
- Flexible gestures for making space, inserting, etc.
- Assertions with evidence gates
- Reasoning templates

Spring 2011

CS 7450

55

# PARC's Entity Workspace



Video

Bier, Card & Bodnar  
VAST '08

Spring 2011

CS 7450

56

# Entity Workspace



- Tools for rapid ingest of entities from documents
- Can snap together entities into groups
- Can indicate level of interest in objects
- Four main view panels, with zooming UI

Spring 2011

CS 7450

57

# Related Area of Interest



- Sensemaking
- A general term that has been used in a number of different contexts
  - E.g., How large corporations make decisions
- To me, ultimately about people working with data and information to understand it better

Spring 2011

CS 7450

58

# Sensemaking



Nice definition:

“A motivated , continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively.”

– Klein, Moon and Hoffman  
*IEEE Intelligent Systems '06*

# Alternate Definition



“The process of creating situation awareness in situations of uncertainty”

– D. Leedom, '01 SM Symp. Report

Situation awareness:

“It’s knowing what’s going on so you know what to do”

– B. McGuinness, quoting an Air Force pilot

## This Topic



- I work on it a lot now
- Interested in getting more work in this area started

Spring 2011

CS 7450

61

## HW 5



- Commercial tools critiques
- Things we noticed

Spring 2011

CS 7450

62

# Upcoming



- Zooming & panning
  - Reading
  
- Time Series Data
  - Reading
  - Aigner et al