

Visual Analytics 2



CS 7450 - Information Visualization
November 20, 2013
John Stasko

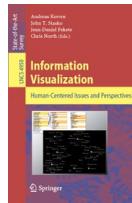
Agenda



- 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



- Thinking¹ - or reasoning - involves objectively connecting present beliefs with evidence in order to believe something else
- Critical Thinking¹ 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.
- Intelligence¹ 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*

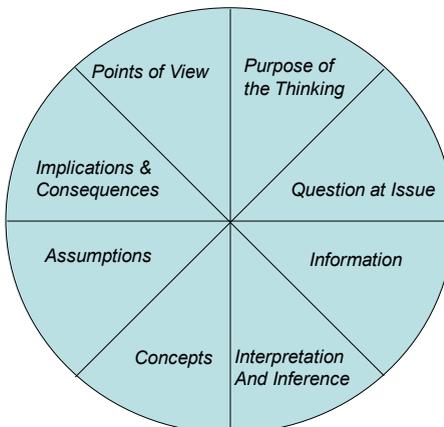
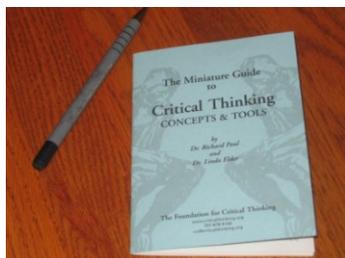


Critical Thinking*



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

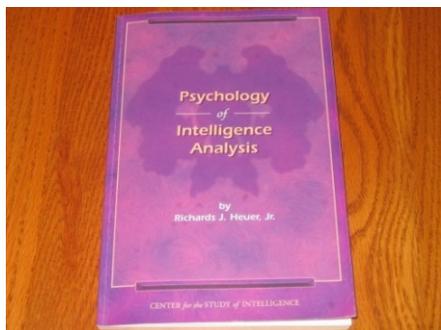
Elements of thought:



* Foundations of Critical Thinking www.criticalthinking.org



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

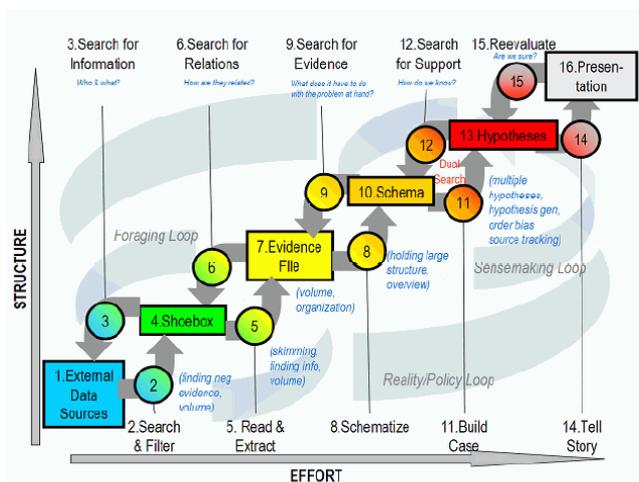


Figure 2.1. Notional model of sensemaking loop for intelligence analysis derived from CTA.

Pirolli & Card

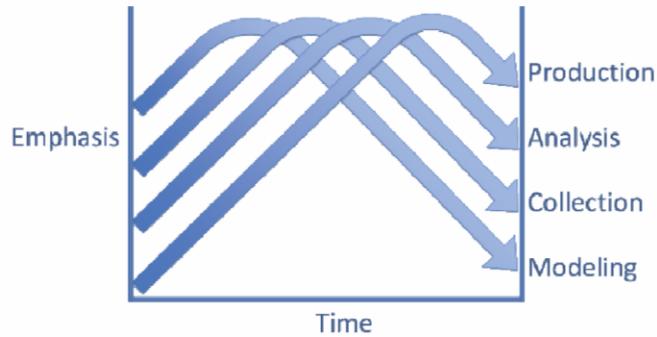
Intl Conf Intelligence Analysis '05

Fall 2013

CS 7450

8

Intelligence Process



Wheaton
In preparation

Fall 2013

CS 7450

9

Pain Points



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

Fall 2013

CS 7450

10



Visualization for Investigative Analysis across Document Collections

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

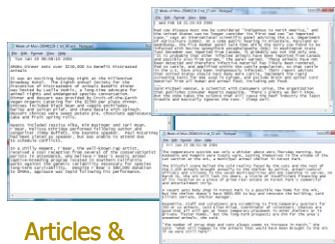
"Putting the pieces together"



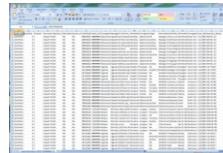
Problem Addressed

Analogy

Help "investigators" explore, analyze and understand large document collections



Articles & reports



Spreadsheets



XML documents



Blogs

The Jigsaw Team



Carsten Görg
Zhicheng Liu
Youn-ah Kang
Jaeyeon Kihm
Jaegul Choo
Chad Stolper
Anand Sainath

and many others

Fall 2013

CS 7450

13

Our Focus

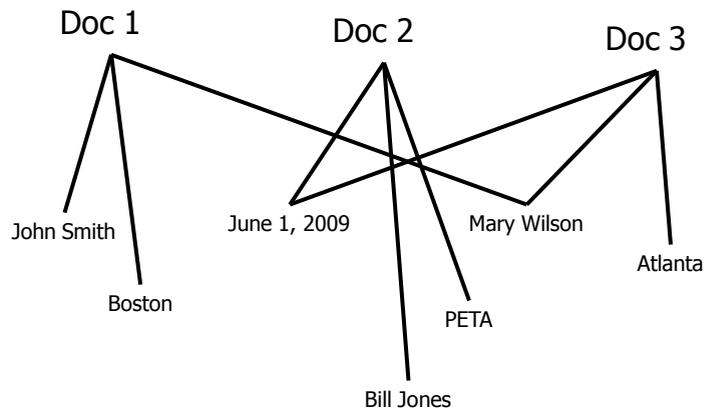


- Entities within the documents
 - Person, place, organization, phone number, date, license plate, etc.
- Thesis: A story/narrative/plot/threat within the documents will involve a set of entities in coordination

Fall 2013

CS 7450

14



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

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.

Fall 2013

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.

Fall 2013

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

Fall 2013

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

Fall 2013

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

Fall 2013

CS 7450

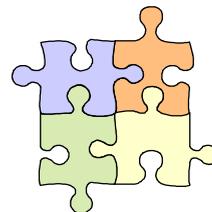
21

Jigsaw

“Putting the pieces together”



- Computational analysis of document text
 - Entity identification, document similarity, clustering, summarization, sentiment
- Multiple visualizations (views) of documents, analysis results, entities and their connections
- Views are highly interactive and coordinated

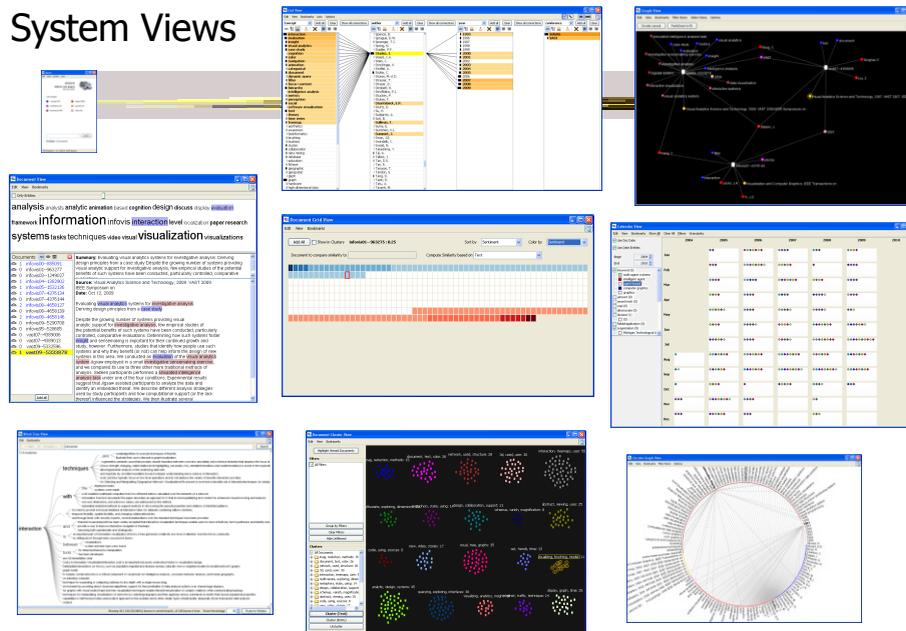


Fall 2013

CS 7450

22

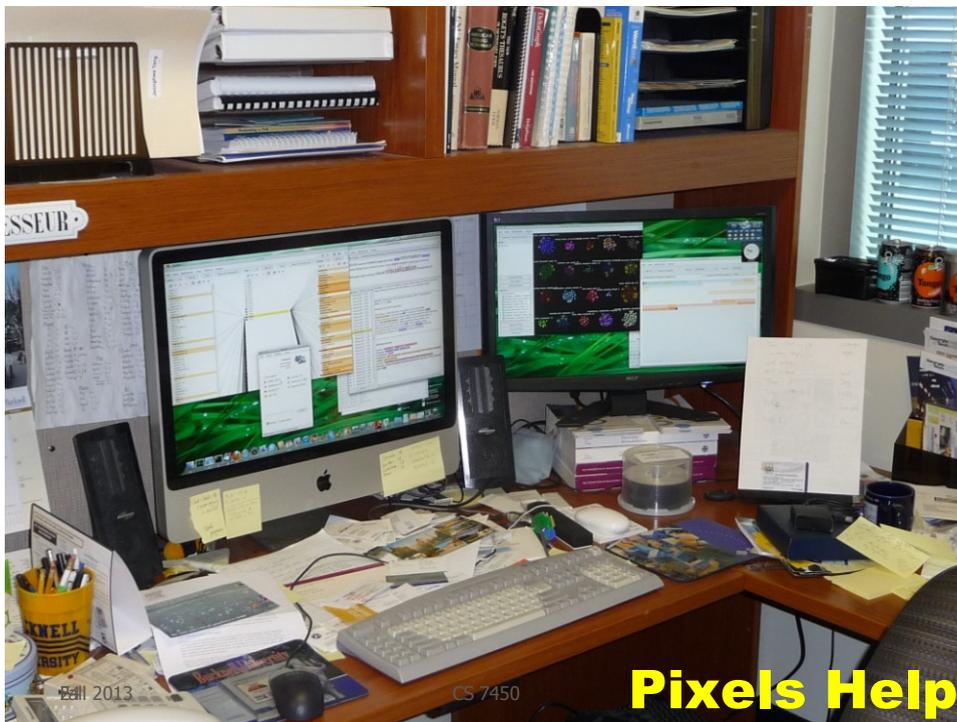
System Views



Fall 2013

CS 7450

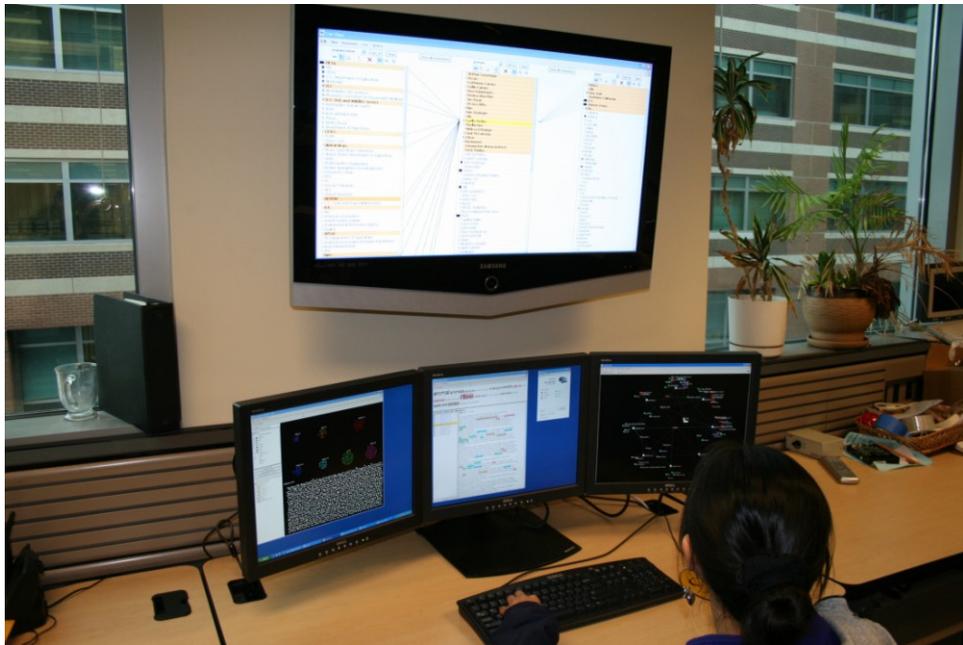
23



Fall 2013

CS 7450

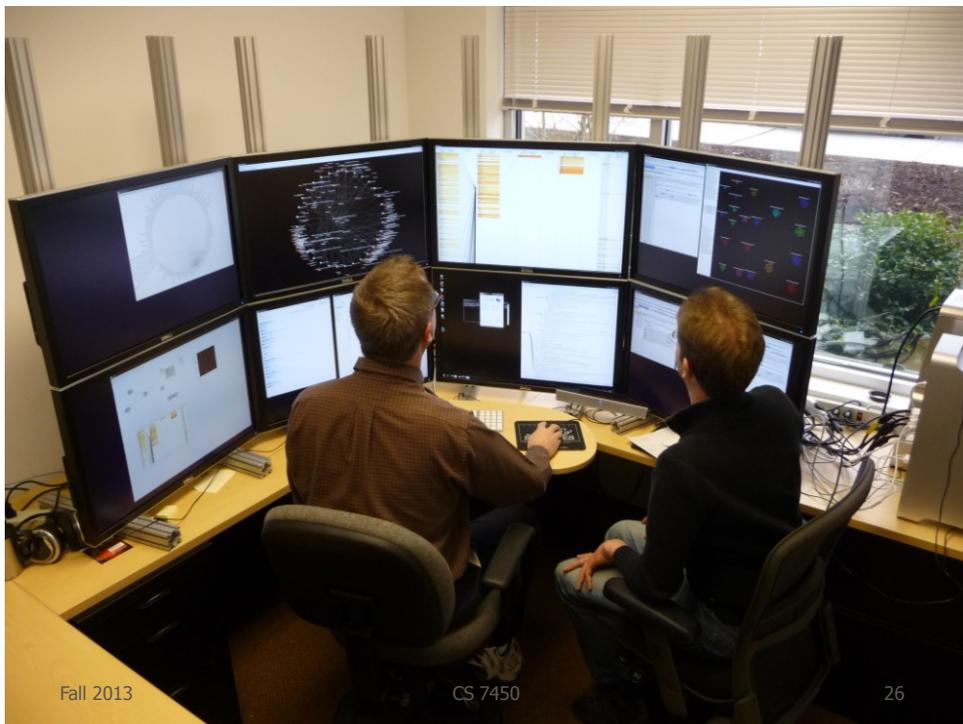
Pixels Help



Fall 2013

CS 7450

25



Fall 2013

CS 7450

26

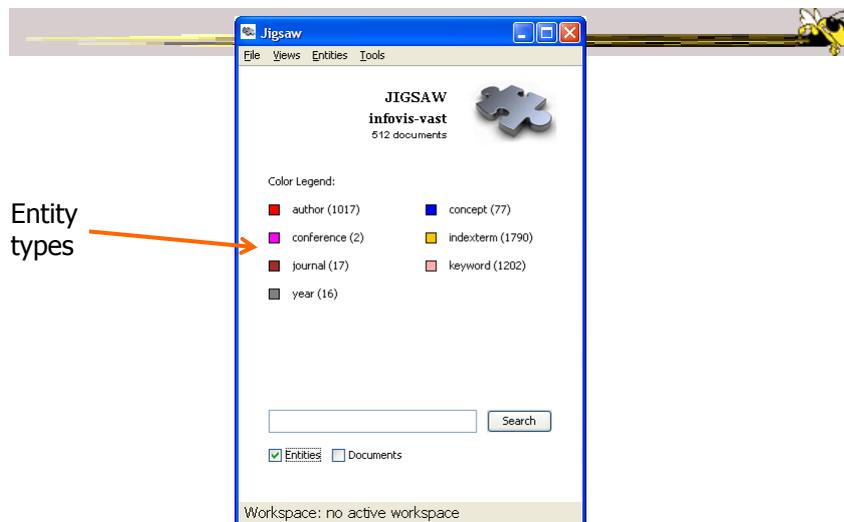


Fall 2013

CS 7450

27

Console



Fall 2013

CS 7450

28

Document View

The Document View window displays a list of documents on the left and a summary of the selected document on the right. The summary text includes:

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. IEEE Symposium on

Date: Oct 12, 2009

Entities identified: 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

Fall 2013

CS 7450

29

List View

Lists of entities by type
Connections highlighted

The List View window displays hierarchical lists of entities by type. The entities are organized into columns: Concept, author, year, and conference. The Concept column lists various terms such as interaction, evaluation, insight, visual analytics, case study, cognition, color, navigation, animation, categorical, document, dynamic query, filter, focus+context, hierarchy, intelligence analysis, metrics, perception, social, software visualization, text, theory, time series, treemap, aesthetics, awareness, bioinformatics, brushing, business, cluster, collaboration, data mining, database, education, filters, geographic, graph, hardware, and high-dimensional data. The author column lists names such as Spence, B., Sprague, D.W., Sreenager, T.C., Spring, N., Stadler, P.F., Stasko, J., Stead, C.A., Stein, C., Stodolinger, K., Stoffel, A., Stroh, C., Storey, M.-A.D., Strayer, T., Strayer, D., Strobel, H., Strothmann, P.J., Studley, P., Sturkenbeck, E.P., Sturtz, D., Su, H., Sudianto, A., Suh, B., Sullivan, T., Suma, E., Summers, K.L., Suresh, J., Swan, J.E., Swindell, C., Suresh, N., Takekuma, Y., Tal, A., Talbot, J., Tan, D.S., Tan, R., Tanase, T., Tandon, S., Tang, D., Tamm, E., Tatu, A., and Tavanti, M. The year column lists years from 1995 to 2009. The conference column lists conferences such as Infovis and VAST. Connections between entities are highlighted with lines.

Fall 2013

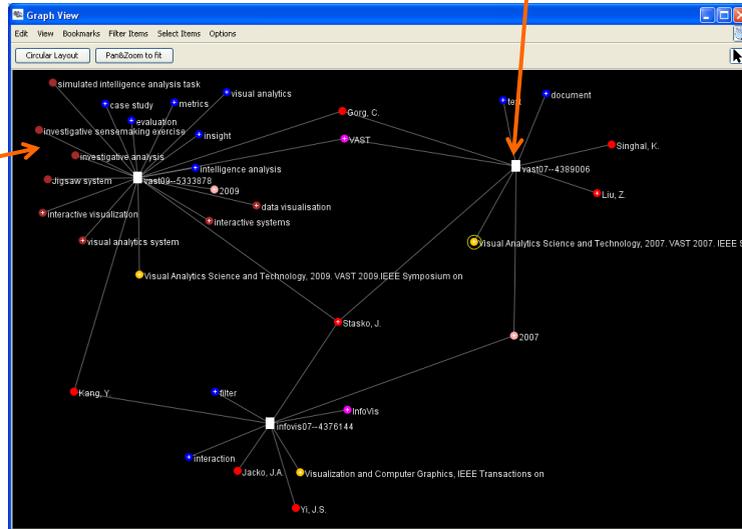
CS 7450

30

Graph View

Document

Entities



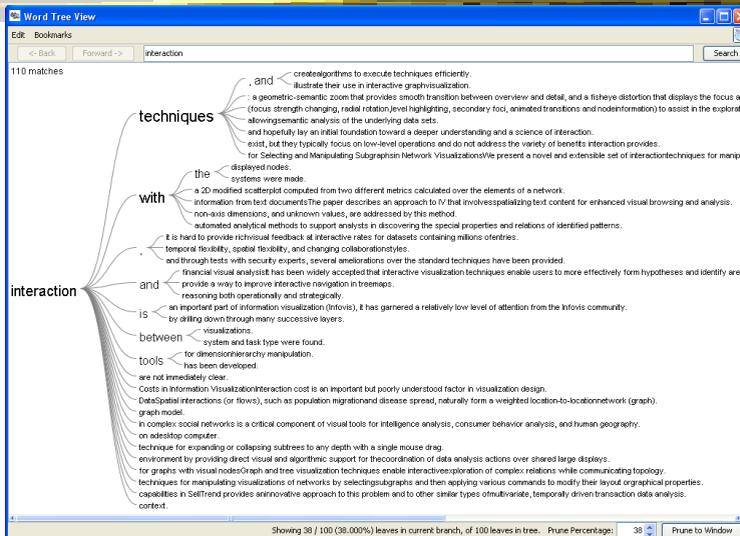
Fall 2013

CS 7450

31

WordTree View

Context of a word in the collection



Fall 2013

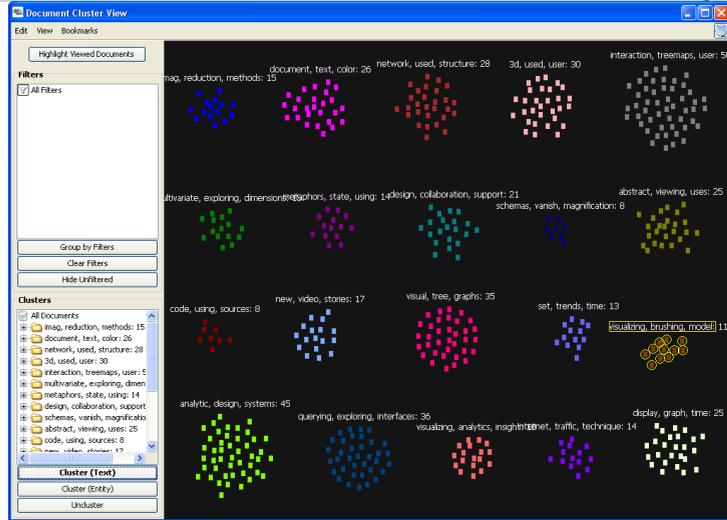
CS 7450

32

Document Cluster View

Clustered by document text or by entities

Summarized by three words



Fall 2013

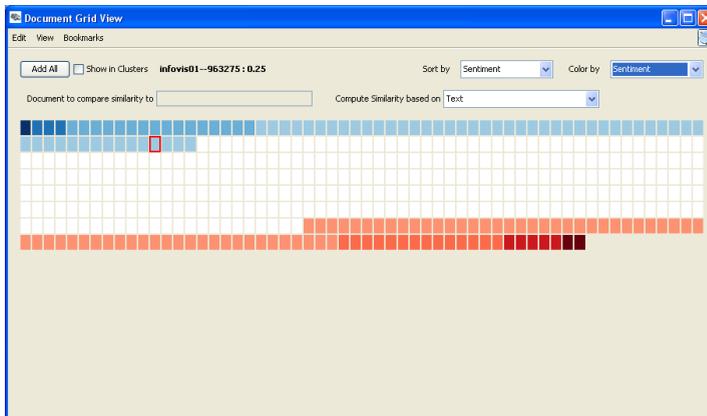
CS 7450

33

Document Grid View

User controls order and color

Sentiment analysis shown here



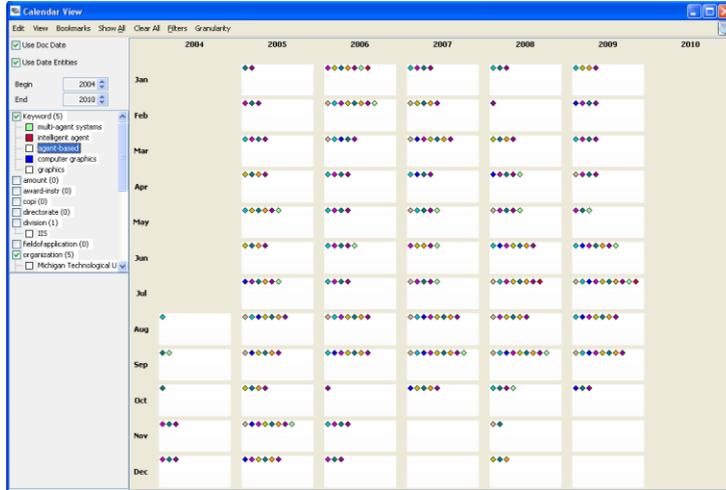
Fall 2013

CS 7450

34

Calendar View

Showing connections between entities and dates



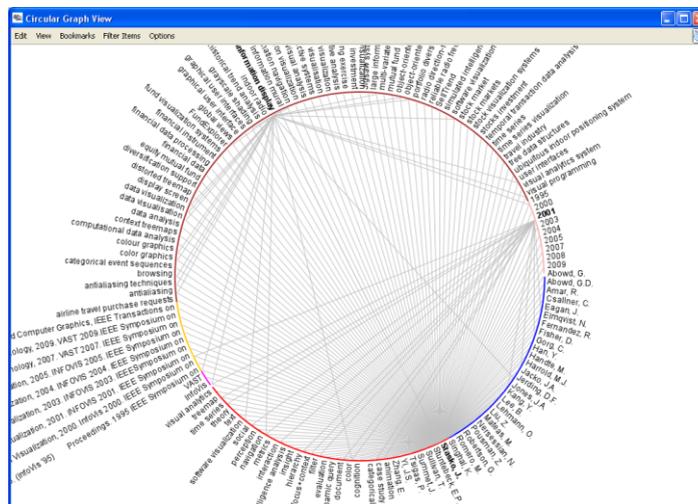
Fall 2013

CS 7450

35

Circular Graph View

Connections between entities



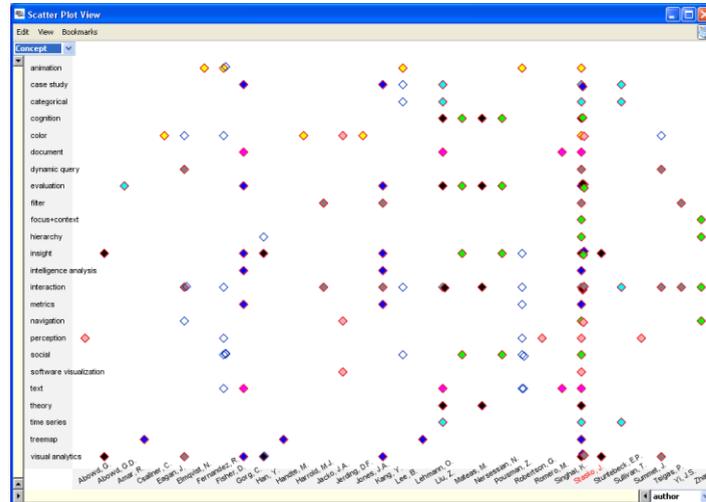
Fall 2013

CS 7450

36

Scatterplot View

Documents containing pairs of entities



Fall 2013

CS 7450

37

Demo 1

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

Fall 2013

CS 7450

38

Demo 2



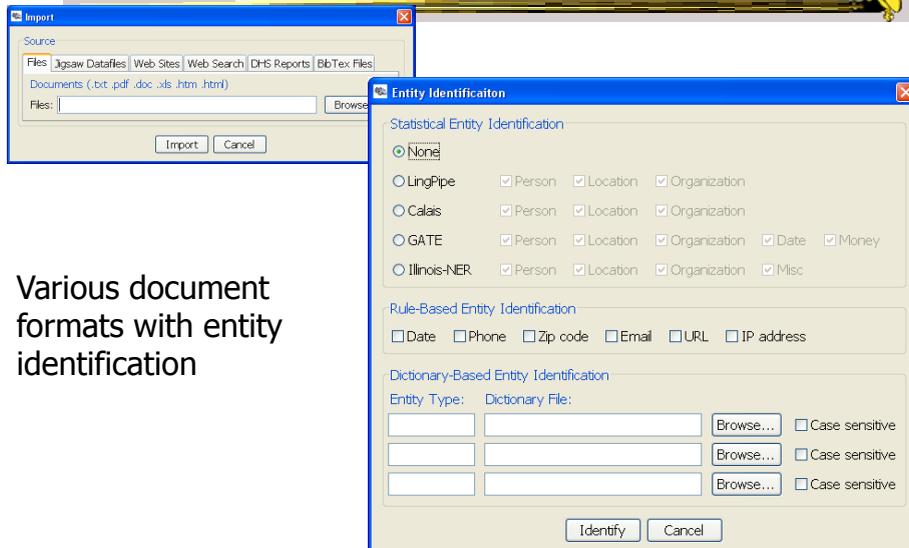
- InfoVis & VAST papers
 - Text: paper title and abstract
 - Entities: author, keyword, year, conference, “concept”

Computational Analyses



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

Document Import



Various document formats with entity identification

Fall 2013

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

Fall 2013

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

Fall 2013

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>

```

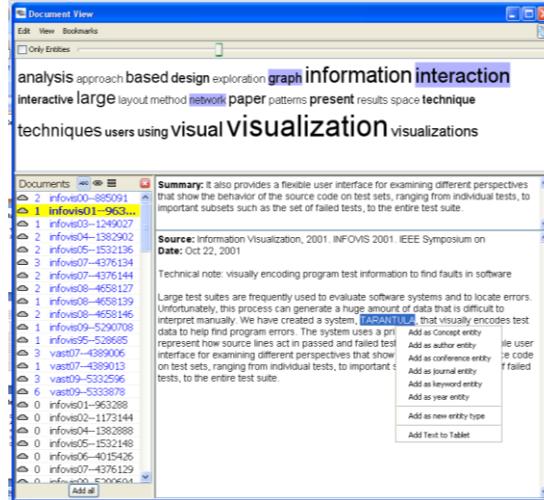
Jigsaw Datafile Format

Fall 2013

CS 7450

44

EI Correction

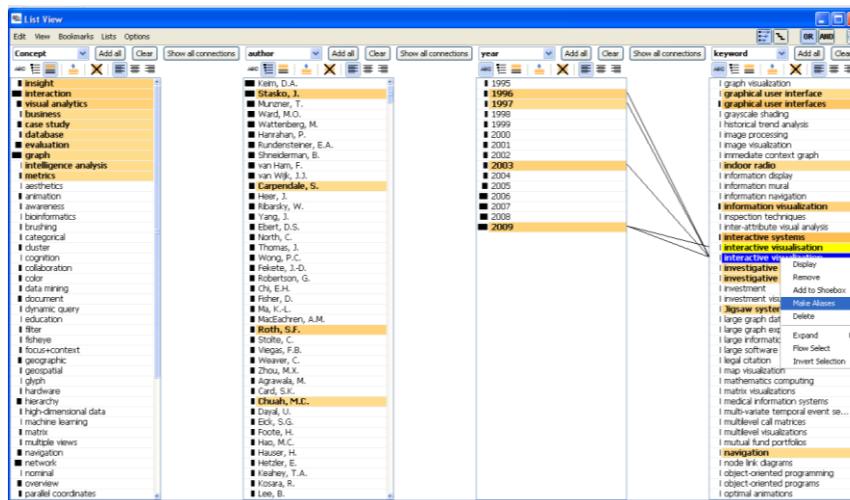


Fall 2013

CS 7450

45

Entity Aliasing

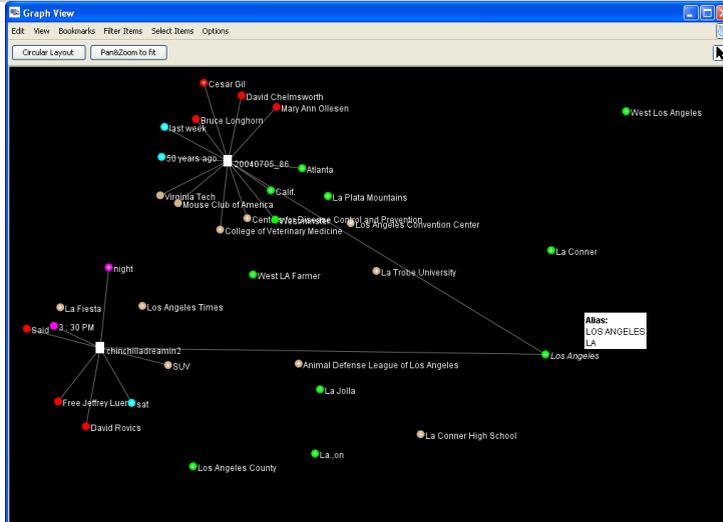


Fall 2013

CS 7450

46

Alias Representation

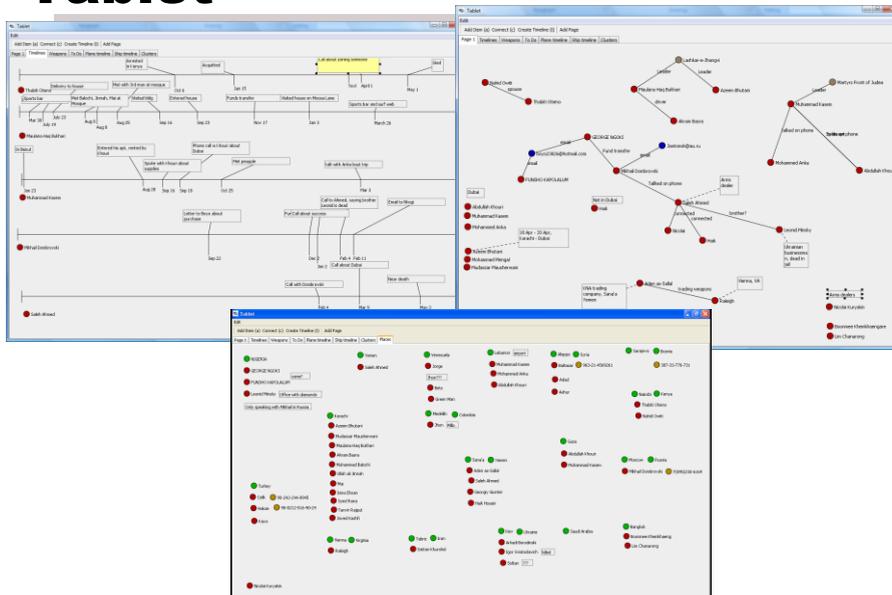


Fall 2013

CS 7450

47

Tablet



Fall 2013

CS 7450

48

More Complex Task Eval



- Consider investigative analysis tasks involving sensemaking, awareness, and understanding
- Research questions
 - How do people use systems?
 - What characteristics matter?
 - What should we measure/observe?
- Exploring methods for utility evaluation

Kang et al
VAST '08 & TVCG '11

Fall 2013

CS 7450

49

System Examined - Jigsaw



The screenshot displays the Jigsaw system interface, which is a complex data visualization tool. It features several panels: a top-left panel with a 'JIGSAW StudyData' sidebar containing a color legend for Person, Location, Organization, and Date; a central 'Document Cluster View' showing a network of nodes and edges; a bottom-left 'Person' list with a search bar and a 'Workspace: no active workspace' indicator; a bottom-right 'Network Graph' showing a detailed view of the relationships between nodes; and a central text pane displaying search results for 'source' with details like 'Date: Oct 18, 2002' and a paragraph of text. The interface is designed for exploring and analyzing large volumes of interconnected data.

Fall 2013

CS 7450

50

Study Design



- Task and dataset Your HW 7
 - 50 simulated intelligence case reports
Each a few sentences long
23 were relevant to plot
 - Identify the threat & describe it in 90 minutes

Source: doc017

Date: Oct 22, 2002

Abu H., who was released from custody after the September 11 incidents and whose fingerprints were found in the U-Haul truck rented by Arnold C. [see doc033] holds an Egyptian passport. He is now known to have spent six months in Afghanistan in the summer of 1999.

Fall 2013

CS 7450

51

Study Design - Settings



1: Paper



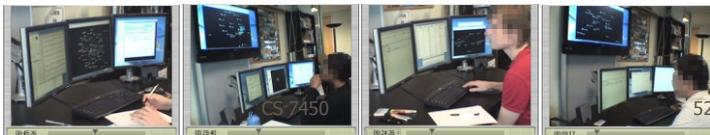
2: Desktop



3: Entity



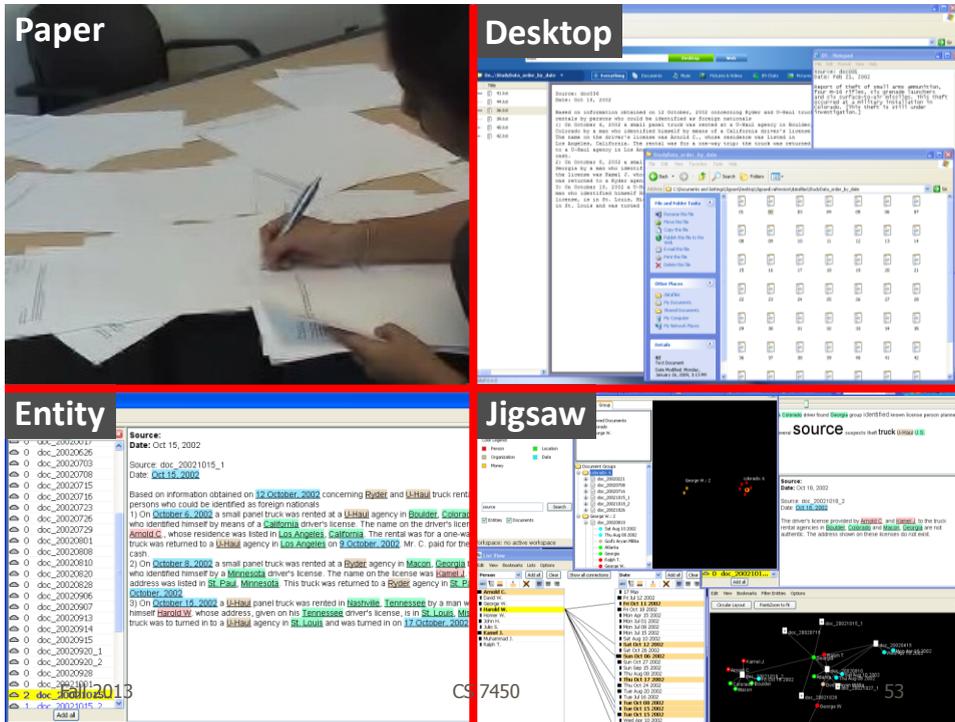
4: Jigsaw



Fall 2013

CS 7450

52



Performance Measures

- Task sheets (like VAST Contest)
 - Three components (relevant people, events, locations)
 - +1 for correct items, -1 for a misidentified items
- Summary narrative
 - Subjective grading from 1 (low) to 7 (high)
- Two external raters
- Normalized, each part equal, mapped to 100-point scale

Results



	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Final Score	22.87	65.00	24.26	87.08	62.08	67.13	42.13	29.41	52.23	15.00	29.26	81.19	95.05	58.07	75.20	90.00
Performance	Fair	Very good	Fair	Excellent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excellent	Excellent	Good	Very good	Excellent
Average Score	49.80				50.19				44.42				79.59			
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23
# of Queries					19	18	48	8	23	61	59	91	44	4	26	8
First Query					40:49	19:55	2:47	12:41	1:31	0:29	0:59	3:12	0:18	5:35	25:37	4:18
Amount of Notes	Many	None	Many	Some	Many	Some	Few	Some	Some	None	None	Few	Some	Few	Few	Few
First Note Taking	0:07		0:05	0:16	1:53	19:57	2:47	8:20	2:37		3:14		0:48	0:32	5:15	78:45
First Task Sheet	43:20	32:53	70:13	3:25	61:35	20:26	7:33	64:11	28:09	0:52	2:55	7:20	48:26	41:48	43:00	5:33

Fall 2013

CS 7450

55

Results



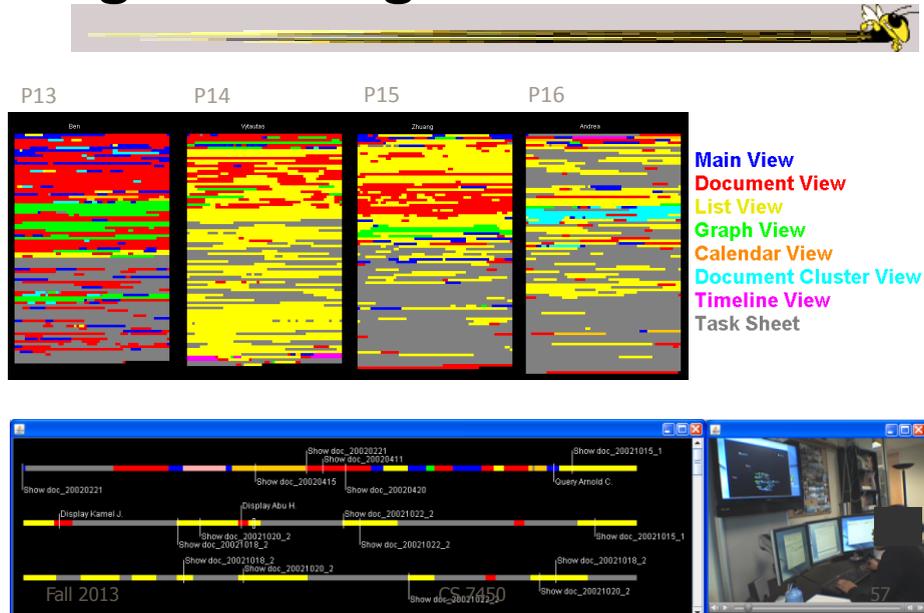
	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Final Score	22.87	65.00	24.26	87.08	62.08	67.13	42.13	29.41	52.23	15.00	29.26	81.19	95.05	58.07	75.20	90.00
Performance	Fair	Very good	Fair	Excellent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excellent	Excellent	Good	Very good	Excellent
Average Score	49.80				50.19				44.42				79.59			
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23
# of Queries					19	18	48	8	23	61	59	91	44	4	26	8
First Query					40:49	19:55	2:47	12:41	1:31	0:29	0:59	3:12	0:18	5:35	25:37	4:18
Amount of Notes	Many	None	Many	Some	Many	Some	Few	Some	Some	None	None	Few	Some	Few	Few	Few
First Note Taking	0:07		0:05	0:16	1:53	19:57	2:47	8:20	2:37		3:14		0:48	0:32	5:15	78:45
First Task Sheet	43:20	32:53	70:13	3:25	61:35	20:26	7:33	64:11	28:09	0:52	2:55	7:20	48:26	41:48	43:00	5:33

Fall 2013

CS 7450

56

Jigsaw Usage Patterns



Investigative Strategies

1. Overview, filter and detail (OFD)
2. Build from detail (BFD)
3. Hit the keyword (HTK)
4. Find a clue, follow the trail (FCFT)

P16: "I like this people-first approach. Once I identify key people, then things that are potentially important come up, too. I'm an impatient person and don't want to read all documents chronologically."

Results by Strategy



	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excellent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excellent	Excellent	Good	Very good	Excellent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Fall 2013

CS 7450

59

Results by Strategy



	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excellent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excellent	Excellent	Good	Very good	Excellent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Fall 2013

CS 7450

60

Results by Strategy



	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excellent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excellent	Excellent	Good	Very good	Excellent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Fall 2013

CS 7450

61

Results by Strategy



	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excellent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excellent	Excellent	Good	Very good	Excellent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Fall 2013

CS 7450

62

Results by Strategy



	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excellent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excellent	Excellent	Good	Very good	Excellent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Tool Design Implications

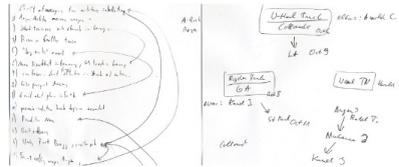


- Support finding starting points/clues
- Guide the analyst to follow the right trail
- Support different strategies of SM process
- Support smooth transition between SM stages
- Provide a workspace
- Allow flexibility in organizing
- Support to find next steps when dead-end
- Facilitate further exploration

Jigsaw's Influence



- Supporting different strategies
- Showing connections between entities
- Helping users find the right clue
- Helping users focus on essential information
- Reviewing hypotheses
- Increasing motivation



Fall 2013

CS 7450

65

Evaluation Recommendations



- Compare system usage to traditional methods
- Collect qualitative data, support with quantitative data
- Consider questions to be answered
- Possible metrics
 - Number of documents viewed
 - When note-taking initiated
 - The quantity of representations created
 - Amount of time and effort in organizing
 - Time spent in reading/processing relevant information

Fall 2013

CS 7450

66

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

Fall 2013

CS 7450

67

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

Fall 2013

CS 7450

68

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...

Fall 2013

CS 7450

69

i2's Analyst Notebook

The screenshot displays the i2 Analyst Notebook website. At the top, there is a navigation bar with links for 'Contact Us', 'Download Center', and 'Info Request', along with a search bar and a dropdown menu for 'USA, Canada & Caribbean'. 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 sub-section for 'i2 Analyst's Notebook Powering Analysis'. This section includes a brief description of the software's capabilities and a list of features. A sidebar on the right lists various products, including 'Analyst's Notebook', 'Analytical Capabilities', 'Online Data Analysis', 'Information Sharing', 'Analyst's Notebook SDK', 'Solutions', 'iBase', 'iBridge', 'ChartReader', 'PatternTracer', 'TextChart', 'ChartExplorer', 'Analyst's Workstation', 'iVR', and 'iXA'. The bottom of the page shows a 'Done' button.

Fall 2013

CS 7450

70

Analyst's Notebook



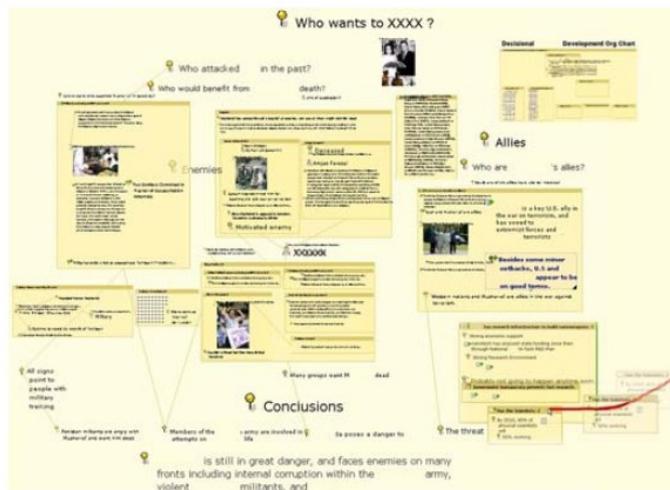
- Leading commercial tool in this space (law enforcement and intelligence agencies)
- Large zooming workspace where analyst creates networks of entities and notes
- Often used to produce presentation or story of analysis done

Fall 2013

CS 7450

71

Oculus' Sandbox



Video

Wright et al
CHI '06

Fall 2013

CS 7450

72

Sandbox

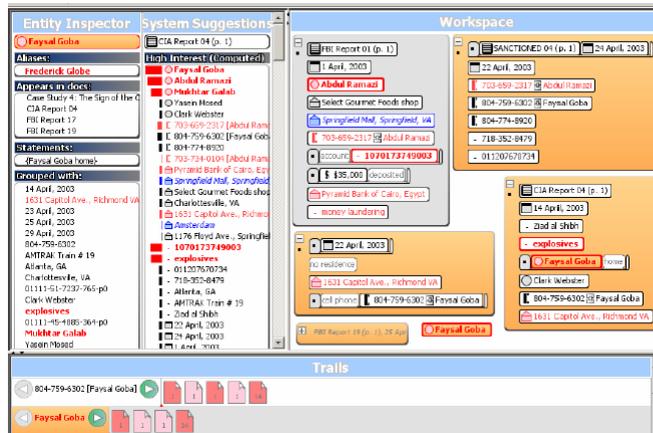
- Flexible space for inserting text and graphics
- Objects can be dragged-and-dropped from their other analysis tools
- Flexible level of detail
- Flexible gestures for making space, inserting, etc.
- Assertions with evidence gates
- Reasoning templates

Fall 2013

CS 7450

73

PARC's Entity Workspace



Video

Bier, Card & Bodnar
VAST '08

Fall 2013

CS 7450

74

Entity Workspace



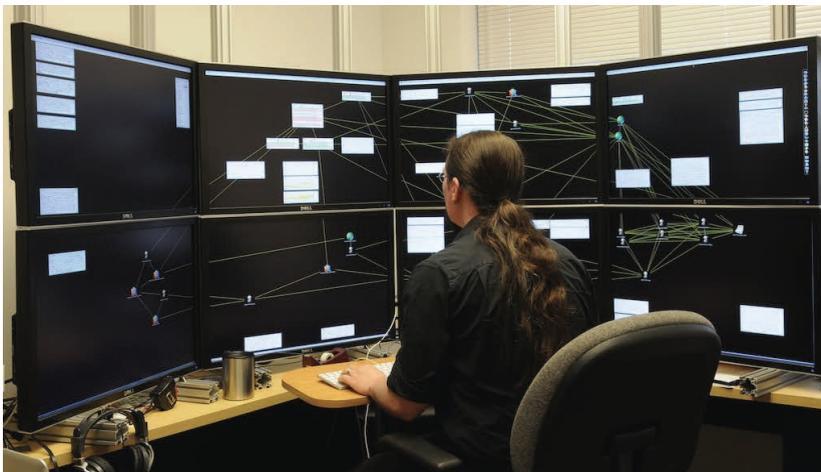
- Tools for rapid ingest of entities from documents
- Can snap together entities into groups
- Can indicate level of interest in objects
- Four main view panels, with zooming UI

Fall 2013

CS 7450

75

VT's Analyst's Workspace



Fall 2013

CS 7450

Video

76

Analyst's Workspace



- Uses spatial affordances from a large display area for benefit in sensemaking
- Analysts move around and arrange items (documents, entities, search results) to externalize the thinking process
 - Like working with pieces of paper on a conference table, but with computational capabilities

Andrews & North
VAST '12

Fall 2013

CS 7450

77

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

Fall 2013

CS 7450

78

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

HW 7



- Be an intelligence analyst
 - Use Jigsaw (available on web)
 - Documents on t-square
- Turn in: Your paragraph description of “threat” + process description (and any materials you want to submit)
- Due Tuesday 26th @ 4pm
 - 1 hardcopy, not email

Upcoming



- Time series data
 - Reading
Aigner et al '08

- Evaluation
 - Reading
Carpendale '08