

Visual Analytics for Investigative Analysis and Exploration of Document Collections

John Stasko

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

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments
A U.S. Department of Homeland Security Center of Excellence

**Georgia
Tech**



**College of
Computing**

School of Interactive Computing

Data Explosion

- Society is more complex
 - There simply is more “stuff”
- Computers, internet and web give people access to an incredible amount of data
 - news, sports, financial, purchases, etc...



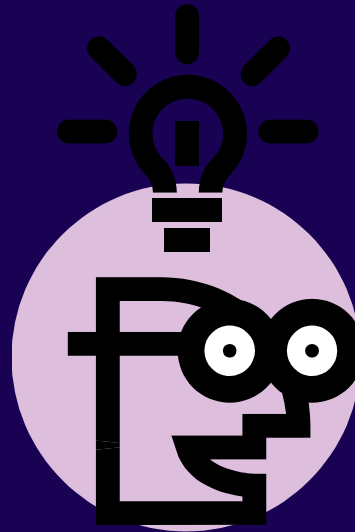
Data Overload

- Confound: How to make use of the data
 - How do we make sense of the data?
 - How do we harness this data in decision-making processes?
 - How do we avoid being overwhelmed?



Premise of my Work

- Visualization of data helps people understand it better



Visualization

- Often thought of as process of creating a graphic or an image
- Really is a cognitive process
 - Form a mental image of something
 - Internalize an understanding
- “The purpose of visualization is insight, not pictures”
 - Insight: discovery, decision making, explanation, analysis, exploration, learning



Main Idea

- Visuals help us think
 - Provide a frame of reference, a temporary storage area
- Cognition → Perception
- Pattern matching
- External cognition aid
 - Role of external world in thinking and reason

Larkin & Simon '87

Card, Mackinlay, Shneiderman '98



When to Apply?

- Many other techniques for data analysis
 - Data mining, DB queries, machine learning...
- Visualization most useful in **exploratory data analysis**
 - Don't know what you're looking for
 - Don't have a priori questions
 - Want to know what questions to ask

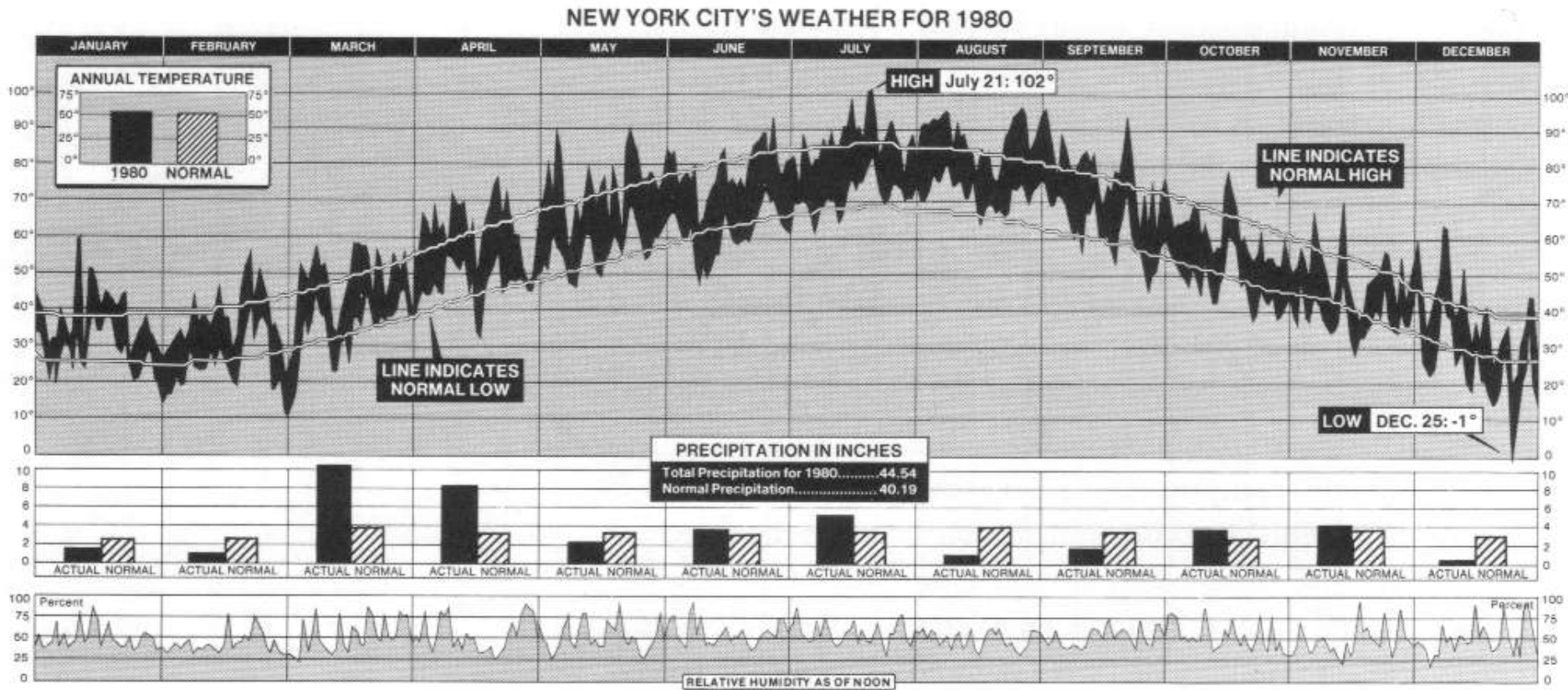


Some quick (static) examples...



NYC Weather

2220 numbers

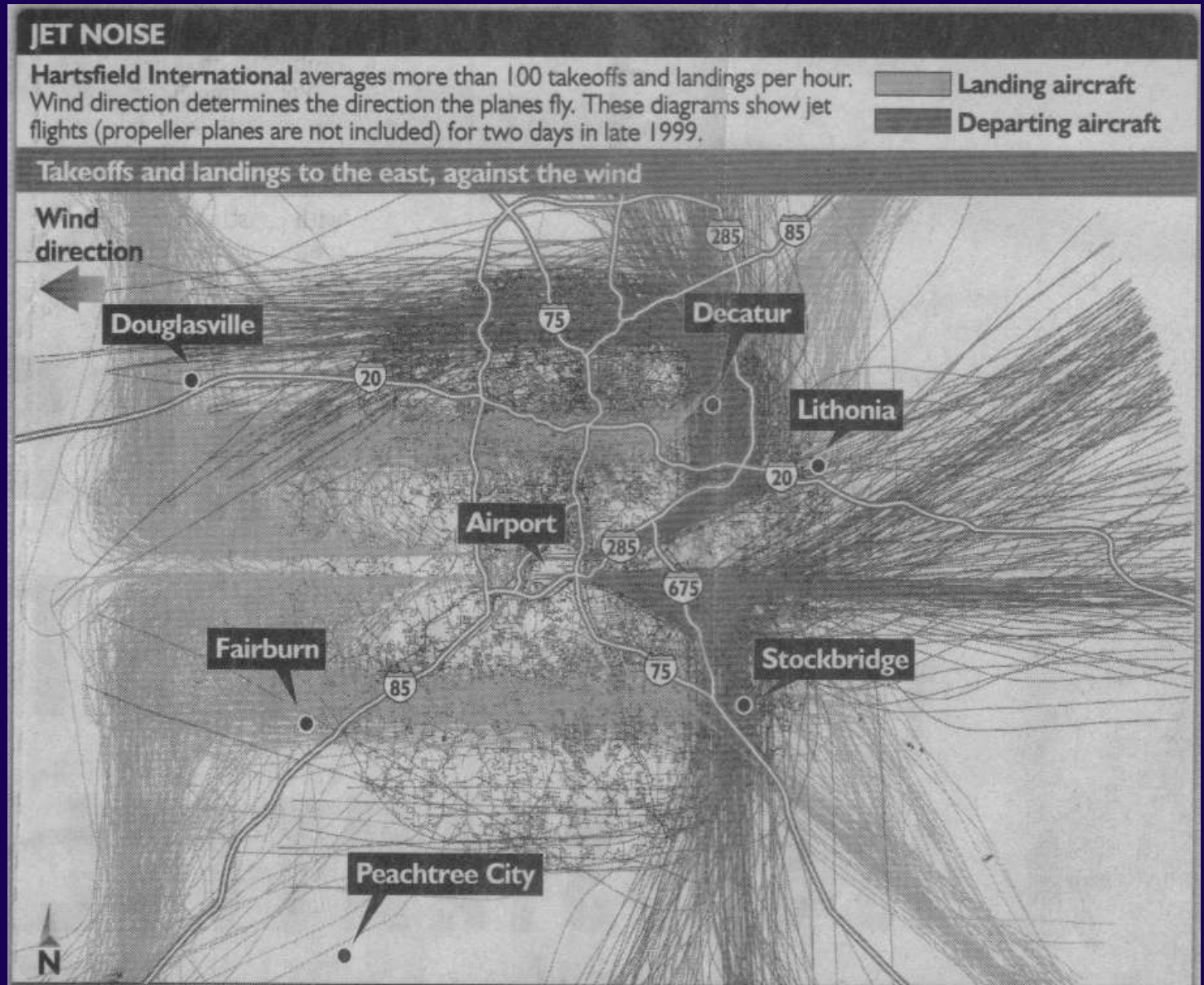


New York Times, January 11, 1981, p. 32.

E. Tufte, *Visual Display of Quant Info*



Atlanta Flight Traffic



Atlanta Journal
April 30, 2000



Reinforce my point with two examples



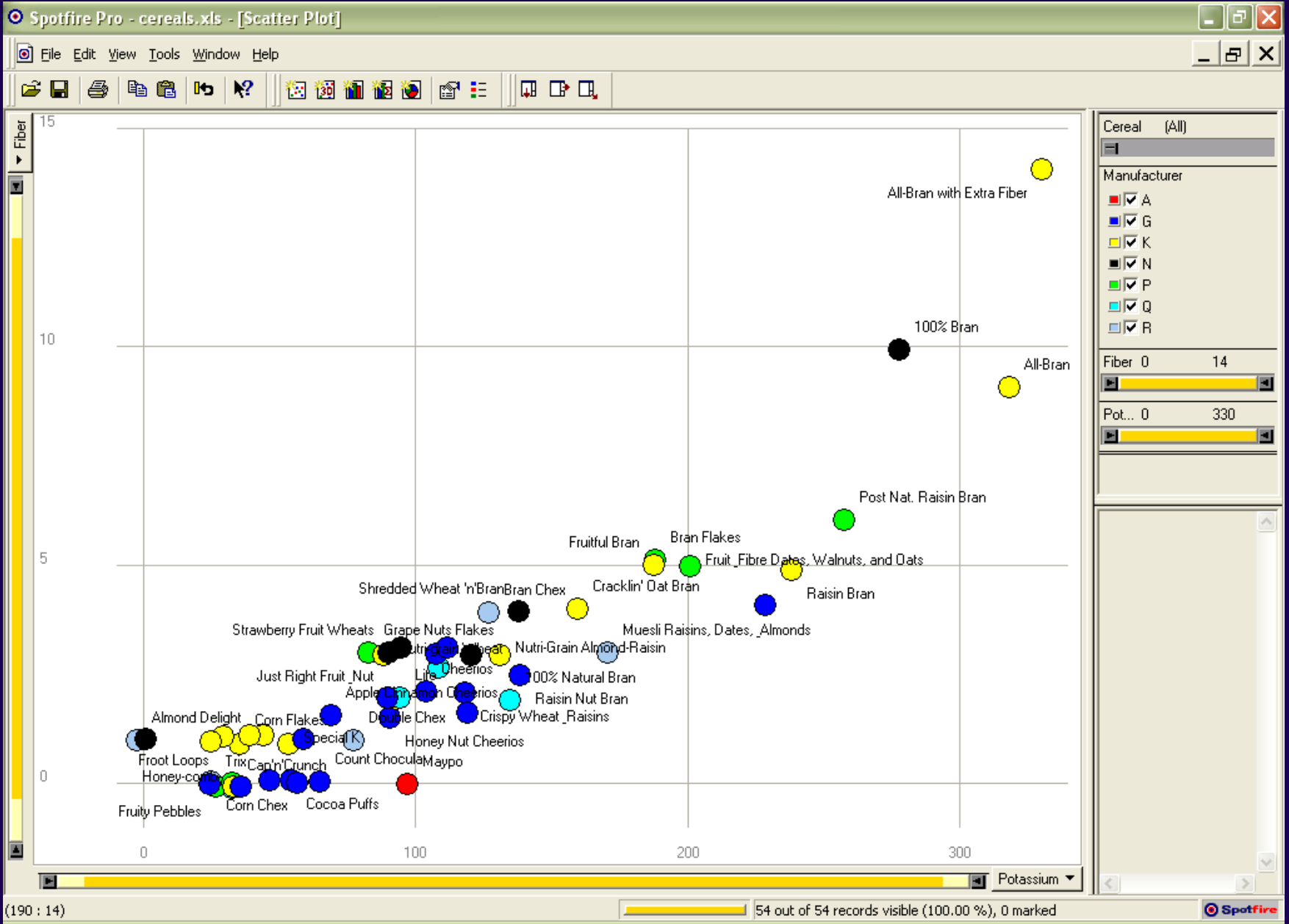
Questions:

- Which cereal has the most/least potassium?
- Is there a relationship between potassium and fiber?
- If so, are there any outliers?
- Which manufacturer makes the healthiest cereals?

	A	B	C	D
1	Cereal	Manufacturer	Fiber	Potassium
2	100% Bran	N	10	280
3	100% Natural Bran	Q	2	135
4	All-Bran	K	9	320
5	All-Bran with Extra Fiber	K	14	330
6	Almond Delight	R	1	0
7	Apple Cinnamon Cheerios	G	1.5	70
8	Bran Chex	R	4	125
9	Bran Flakes	P	5	190
10	Cap'n Crunch	Q	0	35
11	Cheerios	G	2	105
12	Cocoa Puffs	G	0	55
13	Corn Chex	R	0	25
14	Corn Flakes	K	1	35
15	Count Chocula	G	0	65
16	Cracklin' Oat Bran	K	4	160
17	Cream of Wheat (Quick)	N	1	0
18	Crispy Wheat & Raisins	G	2	120
19	Double Chex	R	1	80
20	Froot Loops	K	1	30
21	Frosted Flakes	K	1	25
22	Fruit & Fibre Dates, Wal	P	5	200
23	Fruitful Bran	K	5	190
24	Fruity Pebbles	P	0	25
25	Golden Grahams	G	0	45
26	Grape Nuts Flakes	P	3	85
27	Honey Nut Cheerios	G	1.5	90

28	Honey-comb	P	0	35
29	Just Right Fruit & Nut	K	2	95
30	Life	Q	2	95
31	Lucky Charms	G	0	55
32	Maypo	A	0	95
33	Muesli Raisins, Dates, &	R	3	170
34	Multi-Grain Cheerios	G	2	90
35	Nutri-Grain Almond-Rais	K	3	130
36	Nutri-grain Wheat	K	3	90
37	Oatmeal Raisin Crisp	G	1.5	120
38	Post Nat. Raisin Bran	P	6	260
39	Product 19	K	1	45
40	Quaker Oatmeal	Q	2.7	110
41	Raisin Bran	K	5	240
42	Raisin Nut Bran	G	2.5	140
43	Rice Krispies	K	0	35
44	Shredded Wheat	N	3	95
45	Shredded Wheat 'n'Bran	N	4	140
46	Shredded Wheat spoon	N	3	120
47	Smacks	K	1	40
48	Special K	K	1	55
49	Strawberry Fruit Wheats	N	3	90
50	Total Corn Flakes	G	0	35
51	Total Raisin Bran	G	4	230
52	Total Whole Grain	G	3	110
53	Trix	G	0	25
54	Wheaties	G	3	110
55	Wheaties Honey Gold	G	1	60





Fiber

Potassium



Even Tougher?

- What if you could only see one cereal's data at a time? (e.g. some websites)
- What if I read the data to you?



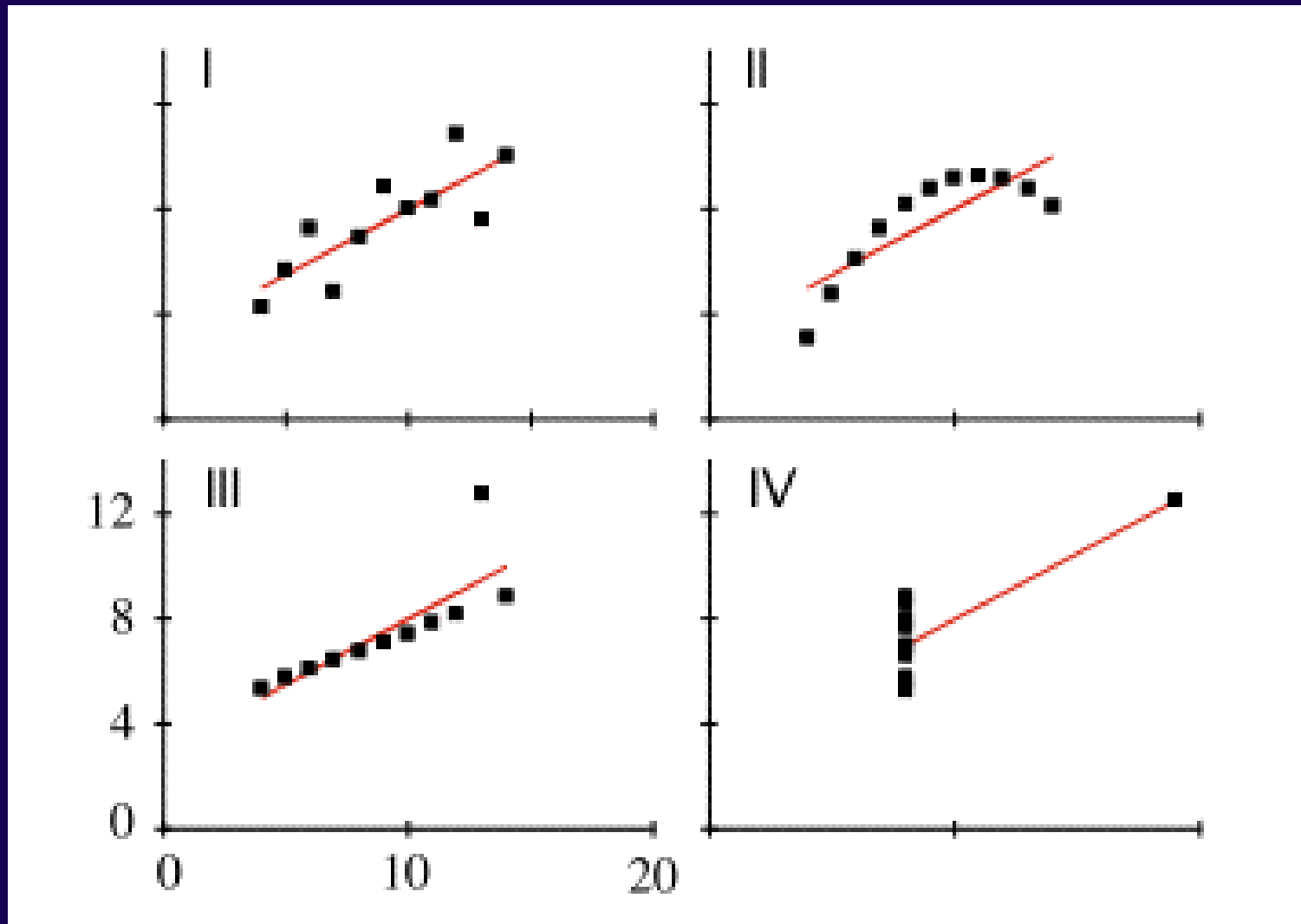
Four Data Sets

- Mean of the x values = 9.0
- Mean of the y values = 7.5
- Equation of the least-squared regression line is: $y = 3 + 0.5x$
- Sums of squared errors (about the mean) = 110.0
- Regression sums of squared errors (variance accounted for by x) = 27.5
- Residual sums of squared errors (about the regression line) = 13.75
- Correlation coefficient = 0.82
- Coefficient of determination = 0.67

<http://astro.swarthmore.edu/astro121/anscombe.html>



The Data Sets



Two Related Disciplines

- Information Visualization
- Visual Analytics



Information Visualization

- Using interactive computer visualizations to represent and communicate abstract data
 - Statistics, databases, software, ...
- Area emerged approximately 1990



Information Visualization

- Recent research trends
 - InfoVis for the Masses
 - Challenges of evaluation
 - Interaction is crucial



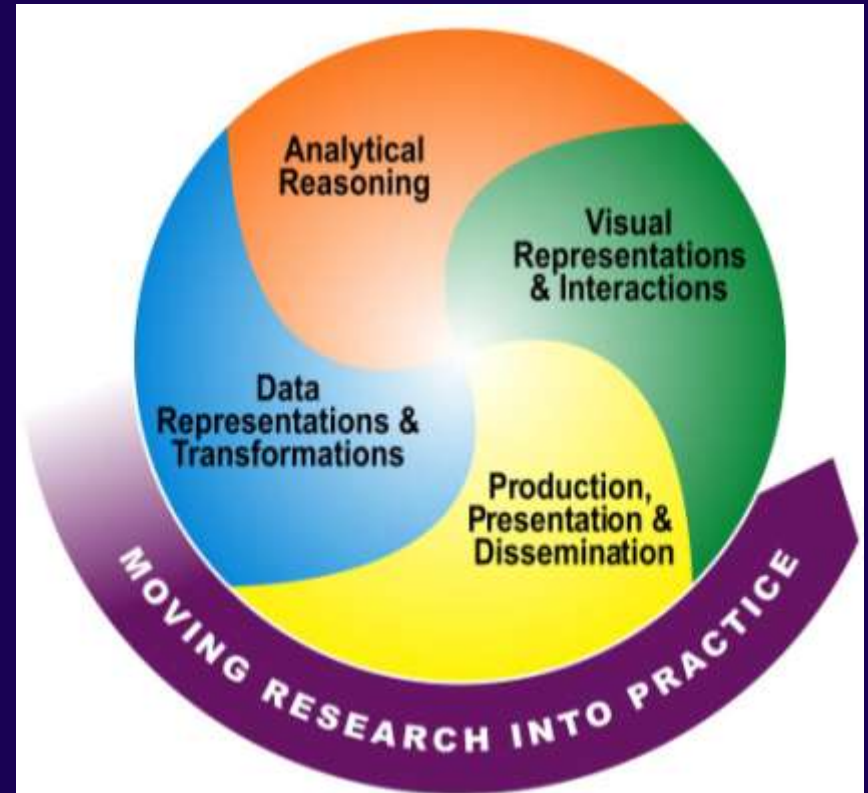
Visual Analytics

- Informal: Using visual representations to help make decisions
- Formal: The science of analytical reasoning facilitated by interactive visual interfaces
- InfoVis++
- Area emerged approximately 2005



Overview of the R&D Agenda

- Science of Analytical Reasoning
- Science of Visual Representations and Interactions
- Data Representations and Transformations
- Production, Presentation, and Dissemination
- Moving Research Into Practice
- Positioning for an Enduring Success



Slide courtesy of Jim Thomas



Visual Analytics: Beyond InfoVis

- Statistics, data representation and statistical graphics
- Geospatial and Temporal Sciences
- Applied Mathematics
- Knowledge representation, management and discovery
- Ontology, semantics, NLP, extraction, synthesis, ...
- Cognitive and Perceptual Sciences
- Communications: Capture, Illustrate and present a message
- Decision sciences

Slide courtesy of Jim Thomas



Sensemaking

“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



Jigsaw

- Visualization for Investigative Analysis across Document Collections
 - Targeting analysis
 - Law enforcement
 - Intelligence community

“Putting the pieces together”



The Jigsaw Team

Current:

Carsten Görg, Youn-ah Kang, Zhicheng Liu, Jaegul Choo, Alex Humesky, Hrishikesh Mantri, Vasili Pantazopoulos

Alumni:

Gennadiy Stepanov, Sarah Williams, Meekal Bajaj, Mohit Jain, Neel Parekh, Roger Pincombe, Kanupriyah Singhal, Xin Sun



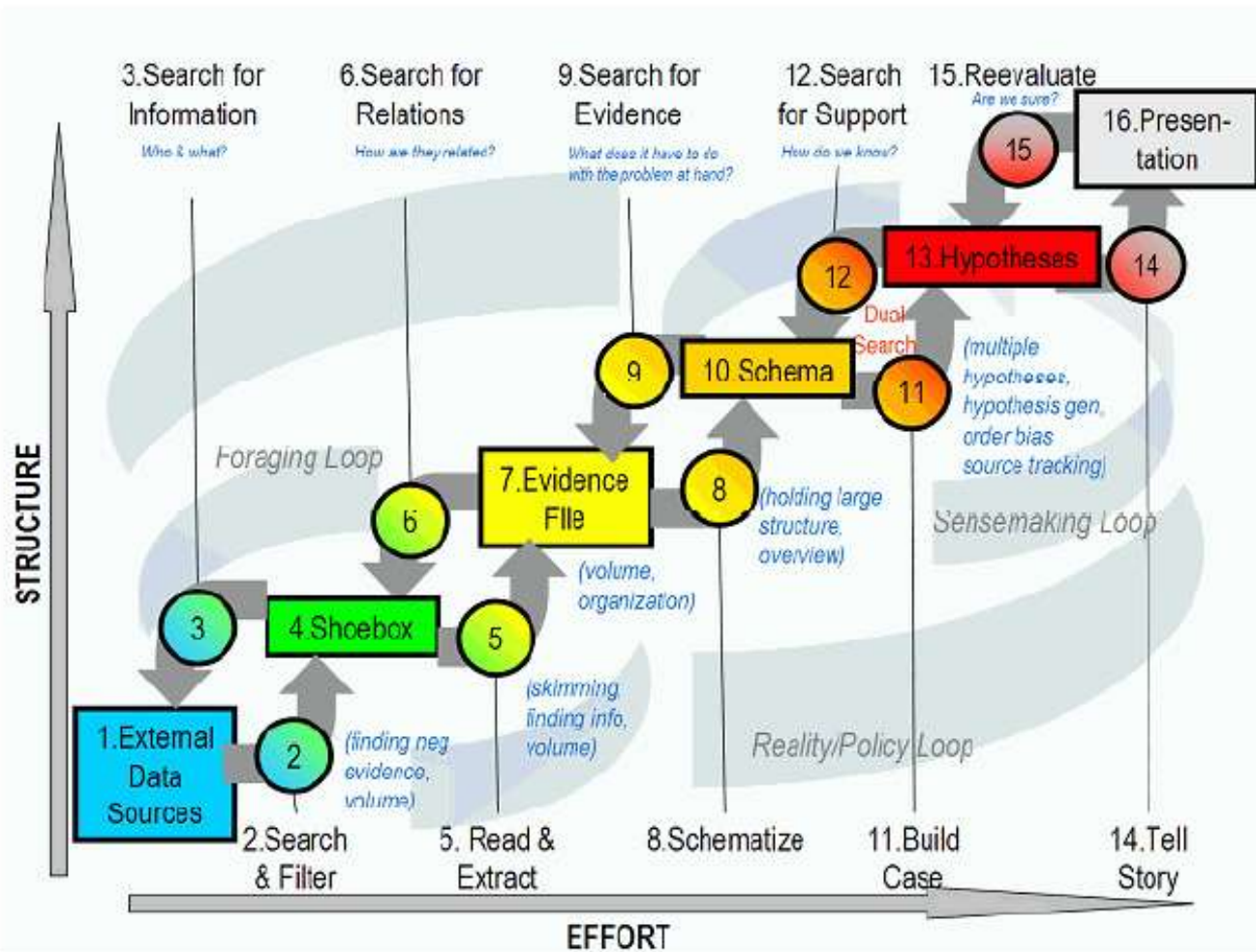


Figure 2.1. Notional model of sensemaking loop for intelligence analysis derived from CTA.

Pirolli & Card, ICIA '05



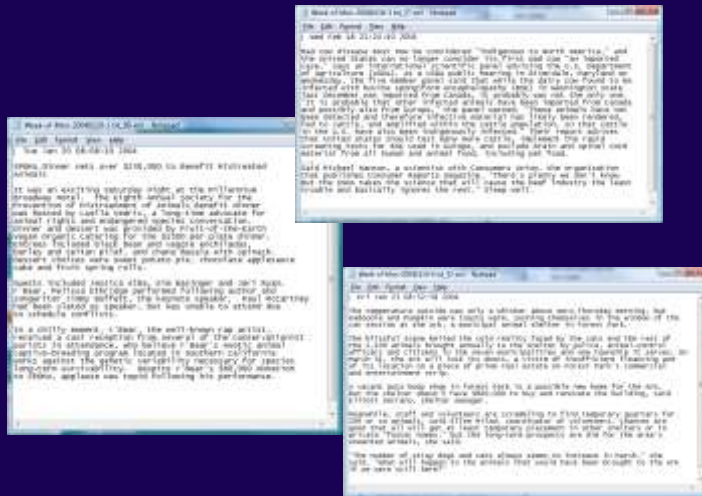
Pain Points

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



Problem Addressed

- Help investigative analysts discover plans, plots and threats embedded across the individual documents in large document collections



Documents/
case reports



DBs



Blogs



Example Documents

2) Report Date: 22 February, 2003. Surveillance report on Cesar Arze, whose residence is 77 Avenue Francis, Santo Domingo, Dominican Republic. Arze, who moved from Havana, Cuba to Santo Domingo in 1992, works as a medical technician in Santo Domingo. Arze is under surveillance because of information that he is associated with Cuban intelligence services. Arze was photographed in company with a man identified as Hector Lopez in Bogota, Columbia on 23 January, 2003. Lopez, a known representative of FARC, has conducted narcotics distribution activities throughout South and Central America and the Caribbean.

1) Report Date: 12 July, 2004. A routine customs inspection was performed on a package that was sent by a person named Pieter Dopple, 22 Hoveniersstraat, Antwerp, Belgium. This package was addressed to A. Hijazi, 1212 Lyons Ave, Newark, NJ. The customs form stated that this package contained two decorative clocks having a commercial value of \$250.00. The package instead contained 111 polished diamonds, whose value is estimated to be \$47,000. Discussions with the FBI resulted in a decision not to pursue a customs violation charge against Hijazi. The reason is that Hijazi is currently under FBI surveillance. This package was resealed and delivered to Hijazi by USPS.

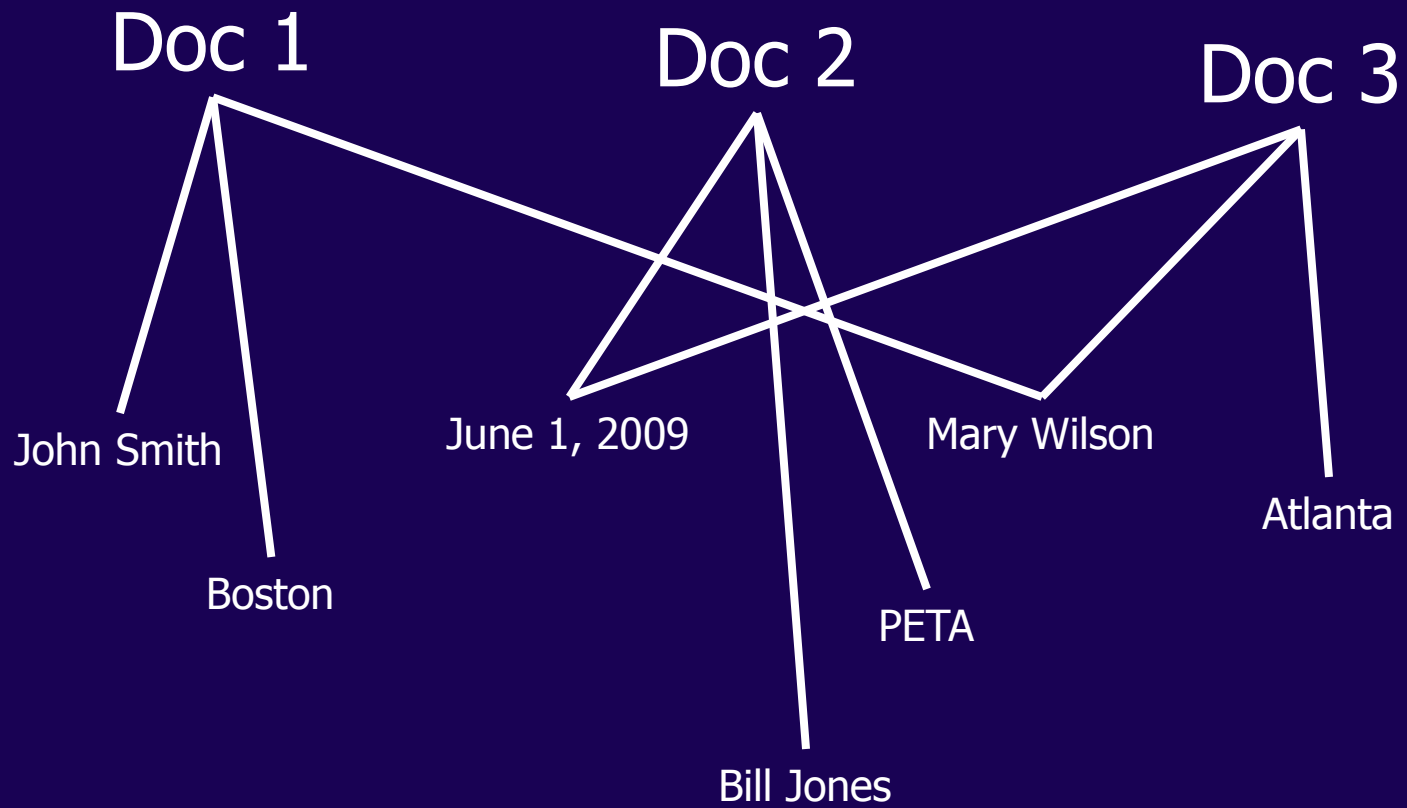
1) Report Date: 11 November, 2003. On 1 November, 2003 Mousa Salah, of no fixed address in Herndon, VA., was arrested at Dulles International Airport as he attempted to board a flight to London Heathrow. At the time of his arrest he was using a Jordanian passport in the name Shadi abu Hoshar. Salah has been wanted in connection with Hamas fund-raising and other activities in Northern Virginia. It is believed that Salah entered the USA illegally from Windsor, Canada into Detroit, MI in 2002. Salah was in possession of a Virginia driver's license # T21-23-8820 registered in the name Mousa Salah. Virginia DOT records show that this same license number was issued on 10 August, 2001 to a woman named April Stevens of Roanoke, VA. Ms Stevens is not the subject of further investigations; she teaches kindergarten classes in Roanoke. The address shown on Salah's driver's license does not exist. Several numbers and initials were written on the back of an envelope in possession of Salah at the time of his arrest; they are: (i) J. T., Detroit, (ii) A. H., Newark, (iii) M. M., Laurel.



Our Focus

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





Entity Identification

- Must identify and extract entities from plain text documents
 - Crucial for our work
- Not our main research focus – Collaborate with or 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.



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.



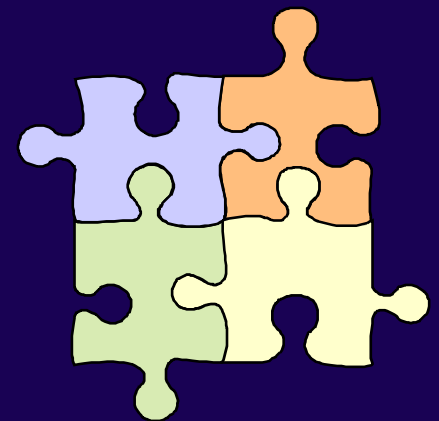
Connections

- 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



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



System Views

Document View

against agents **arrow** arson **authorities** being
canada canadian charges construction **elf** enforcement
 environmental **fbi** federal government **group** logging

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**

Report View

Color Legend

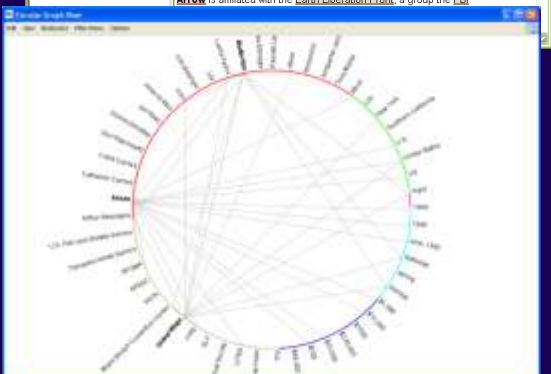
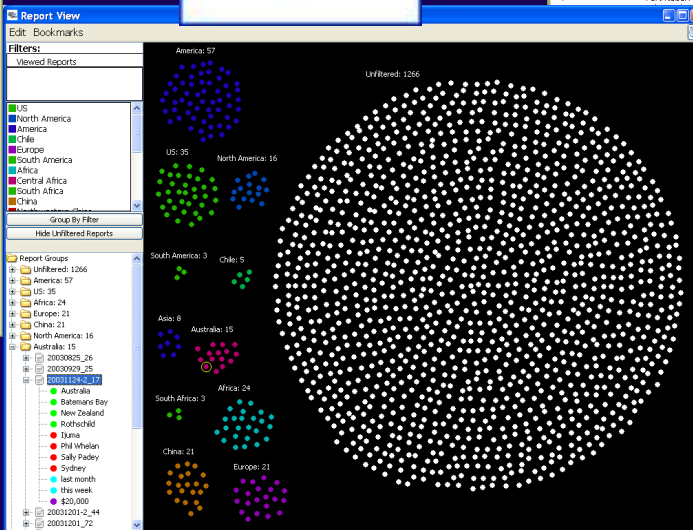
- place
- organization
- person
- date
- entity
- class

Search

List View

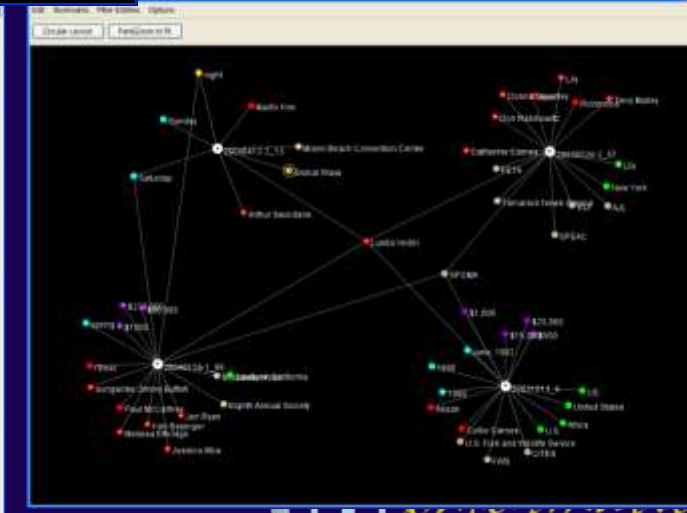
place person organization

- Abbotsford
- Alberta
- America
- Blaine
- British Columbia
- British Columbia
- Buckner
- Calgary
- Canada
- central alberta
- Chiapas
- Colo
- Collo
- Colorado
- Columbia
- Dawson City
- Eagle Creek
- Edmonton
- Fort Collins
- Fort Nelson
- Arnold Schwarzenegger
- Arrow
- Connor Cash
- Conner Cash, 22
- Cottrell, 23
- Draig Rosebraugh
- David Barabashi
- Duane Christopher Bradley
- Gary R. Perstein
- George Bush
- Jack
- Jeff Gaudy
- Jeff Koenigs
- Jim Widmer
- Keith
- Keith Jaspersen
- Michael Scarpitti
- News Headlines
- Pumpkin
- Robert Blacksmith
- Ron Arnold
- Scott McInnis
- Tre Arrow
- Wallace Stogier
- Wesley Smith
- Wesley Smith
- WILL-LIFE EXPERT CAUGHT SMUGGLING REPTILES
- William Jensen Cottrell
- Wolf



U.S. Bookmarks

Report selected	Sa	Mo	Tu	We	Th	Fr	Sa	Sa	Mo	Tu	We	Th	Fr	Sa
Jan														
Feb														
Mar														
Apr														
May														
Jun														
Jul														
Aug														
Sep														
Oct														
Nov														
Dec														



Web View

U.S. Bookmarks

Remove All Filters

U.S. Bookmarks

- night
- atlantic africa
- tasmania
- Gender National Park
- Canada
- England
- Congo River
- Mexico
- Robert Fabry
- John Carroll
- Jan Tatarowicz
- Montenegro
- Dale Anderson
- Anne Pacey
- University of Minnesota
- University of Cambridge
- Historical Museum of Natural
- today
- 6 million years ago
- 1 million years ago
- 1.8 million years ago

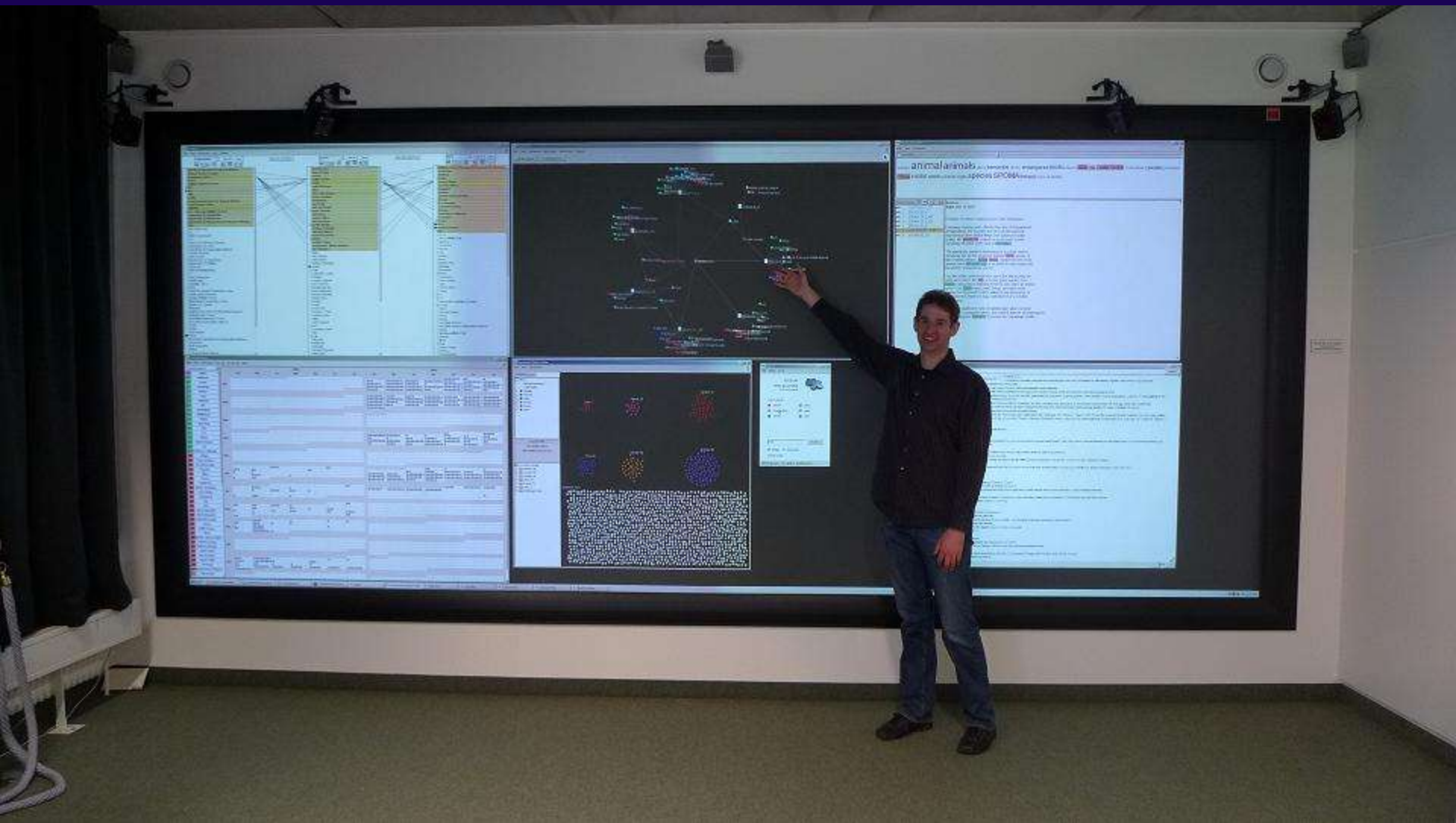


The Need for Pixels



Even More!





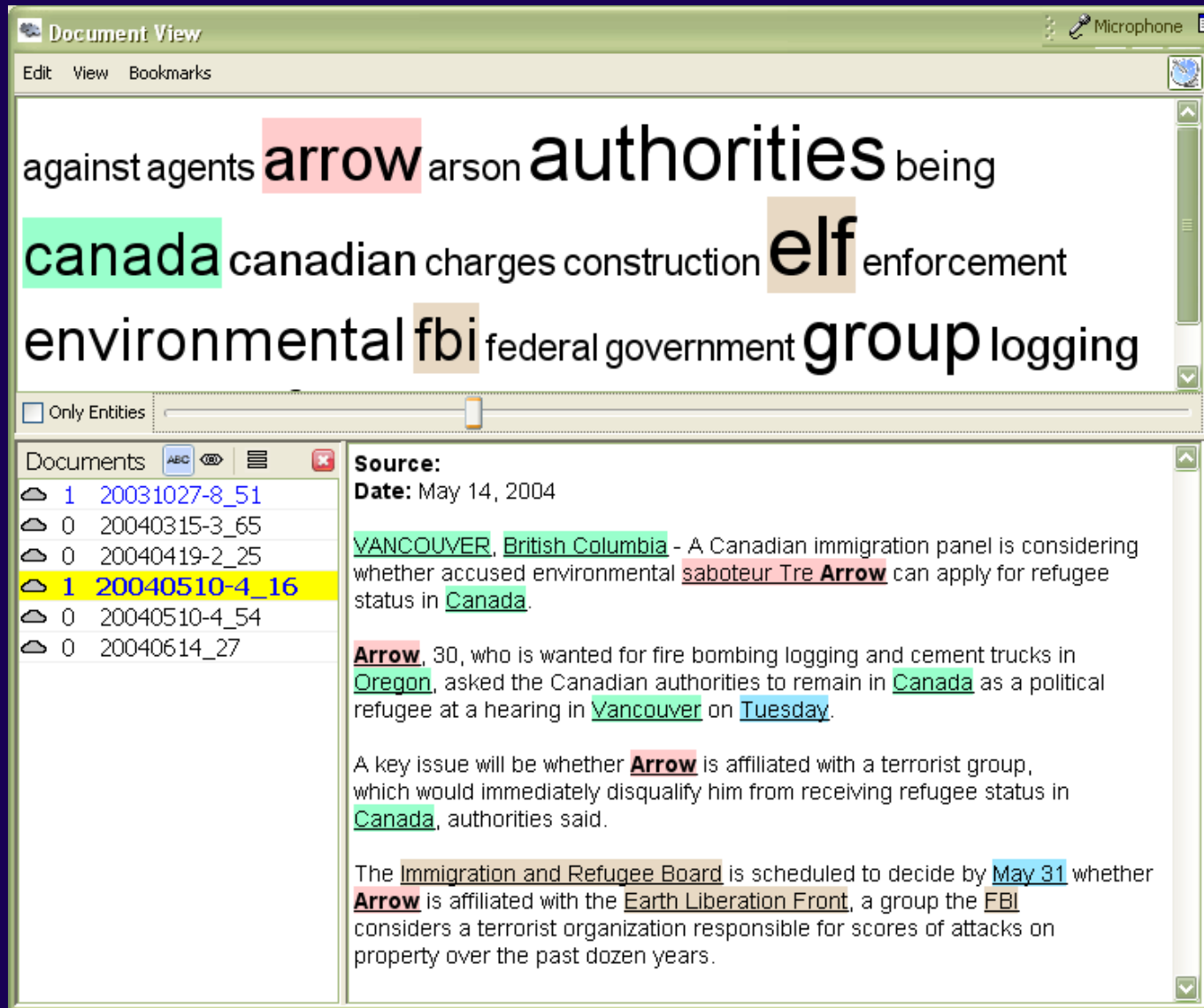
Demo



Console



Document View



Document View

Microphone

Edit View Bookmarks

against agents **arrow** arson **authorities** being
canada canadian charges construction **elf** enforcement
environmental **fbi** federal government **group** logging

Only Entities

Documents

- 1 20031027-8_51
- 0 20040315-3_65
- 0 20040419-2_25
- 1 20040510-4_16
- 0 20040510-4_54
- 0 20040614_27

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.



List View

The screenshot shows a 'List View' application window with three main panels, each displaying a list of items. The panels are labeled 'place', 'person', and 'organization'. Each panel has a dropdown menu, 'Add all', 'Clear', and 'Show all connections' buttons, and a toolbar with icons for sorting and filtering. Lines connect items across the lists, illustrating relationships. The 'Tre Arrow' item in the 'person' list is highlighted in yellow.

place

- Abbotsford
- Alberta
- America**
- Blaine
- British Columbia**
- British Columbia
- Buckner
- Calgary
- Canada**
- central Alberta
- Chiapas**
- Colo**
- Colo.**
- Colorado
- Columbia
- Dawson City
- Eagle Creek**
- Edmonton
- Fort Collins
- Fort Nelson
- Great Slave Lake
- Hollywood
- Iraq**
- Long Island**
- LOS ANGELES**
- Los Angeles**
- Mexico**
- Montana
- N.M**
- N.Y**
- northern British Columbia
- Northwest Territories
- Oregon**
- Portland
- Puget Sound

person

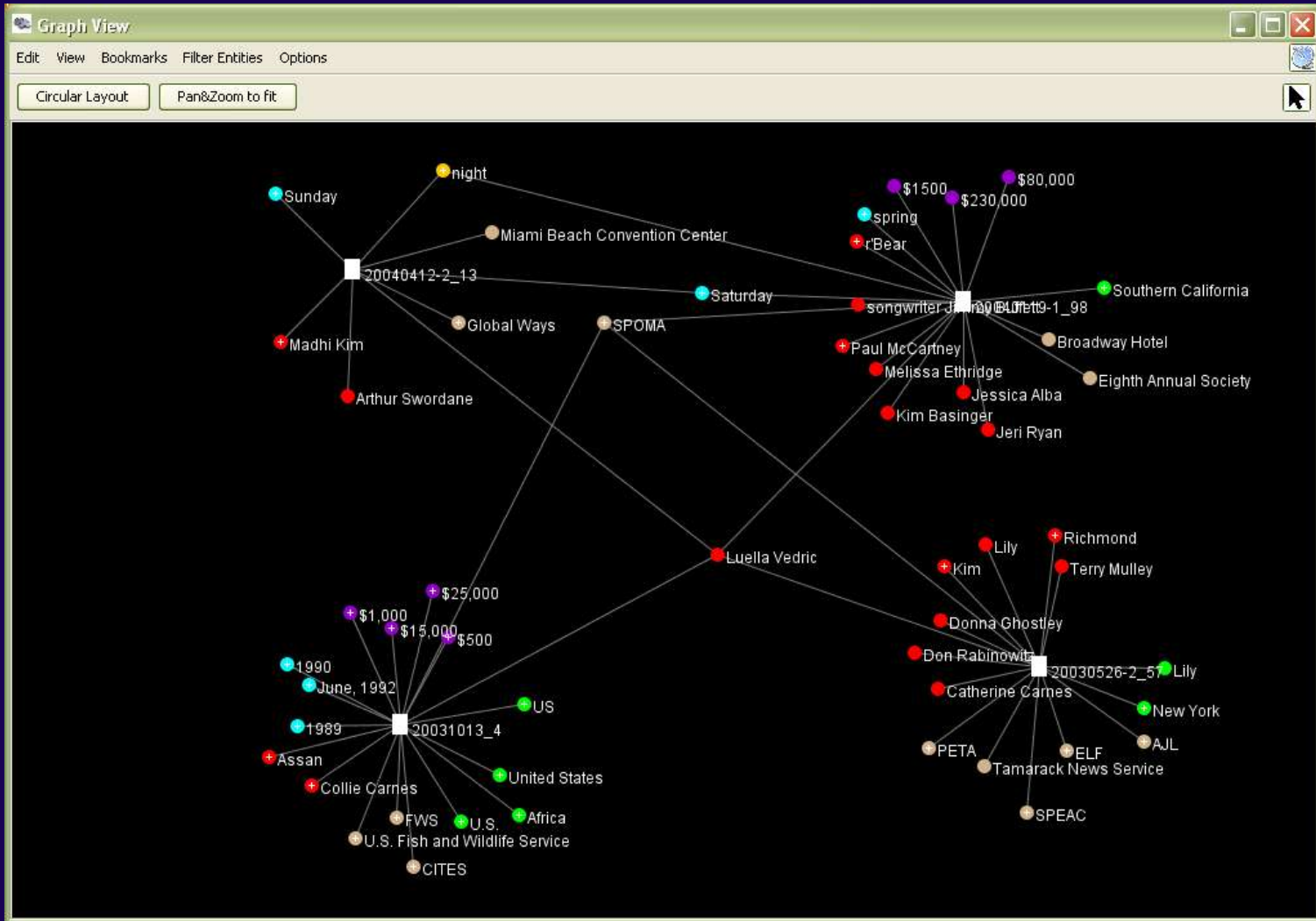
- Arnold Schwarzenegger**
- Arrow**
- Connor Cash
- Connor Cash, 22
- Cottrell, 23
- Craig Rosebraugh
- David Barbarash
- Duane Christopher Bradley
- Gary R. Perlstein
- George Bush
- Jack
- Jeff Guidry
- Jeff Koenings
- Jim Widmier
- Keith
- Keith Jespersion
- Law**
- Michael Scarpitti**
- News Headlines
- Pumpkin
- Robert Blecksmith**
- Ron Arnold**
- Rosebraugh, 31
- saboteur Tre Arrow
- Scott McInnis**
- Tre Arrow**
- Wallace Stegner
- Wesley Smith
- WILDLIFE EXPERT CAUGHT SMUGGLING REPTILES
- William Jensen Cottrell**
- Wolf

organization

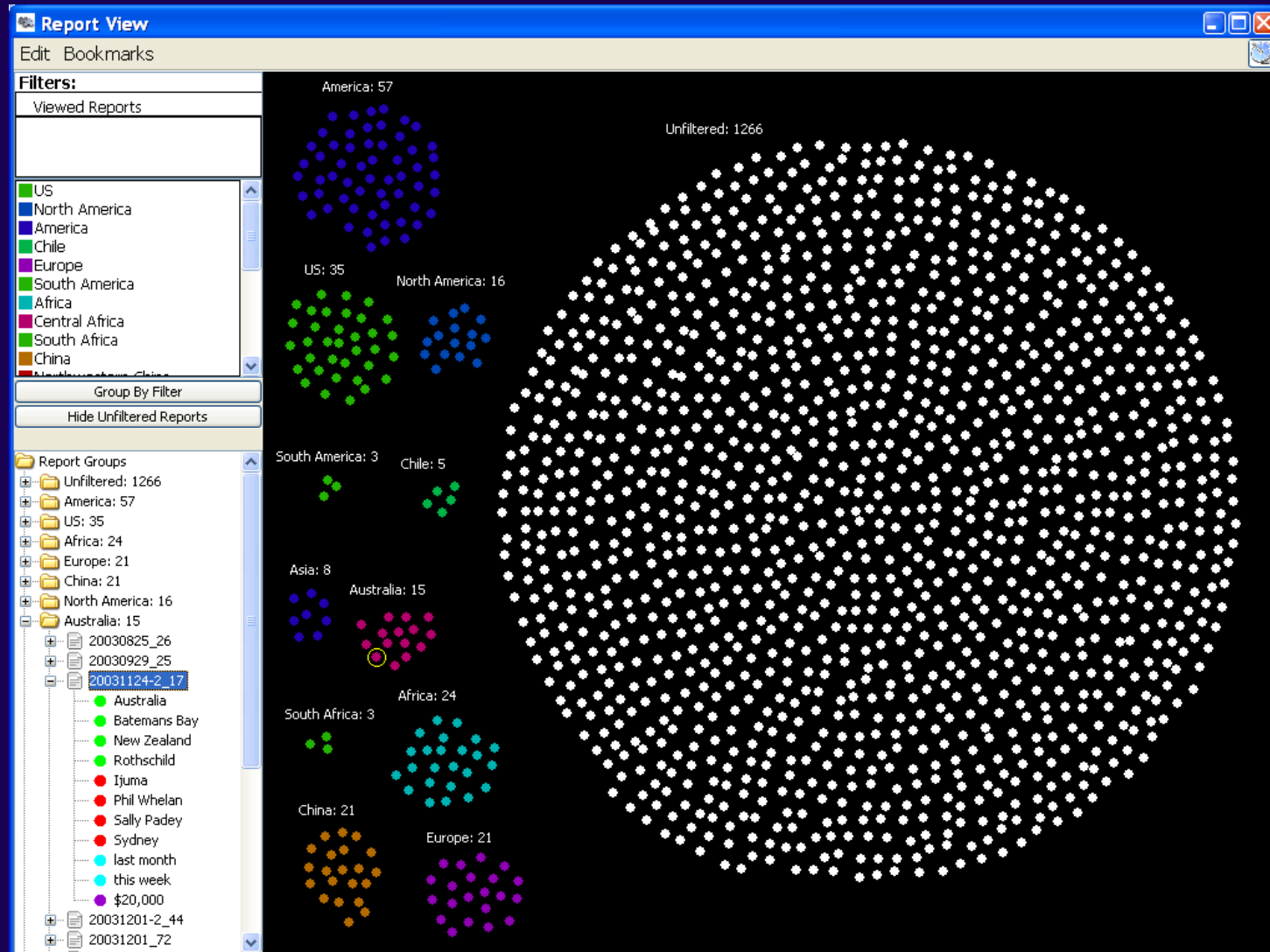
- ALF
- Animal Liberation Front**
- California Institute of Technology**
- CWD
- Department of Fish and Wildlife
- Earth First
- Earth Liberation Front**
- ELF**
- ELF Web
- FBI**
- FDA
- Fourth World War**
- Immigration and Refugee Board**
- Joint Terrorism Task Force**
- Portland State University**
- San Bernardino Central Detention Cen.**
- Sarvey Wildlife Center
- Schoppert Logging Company**
- Supreme Court**
- SUV**
- U.S. Department of Agriculture
- U.S. District Court
- US Forrest Service**
- Washington Fish and Wildlife Commission
- WHO KILLED JANE



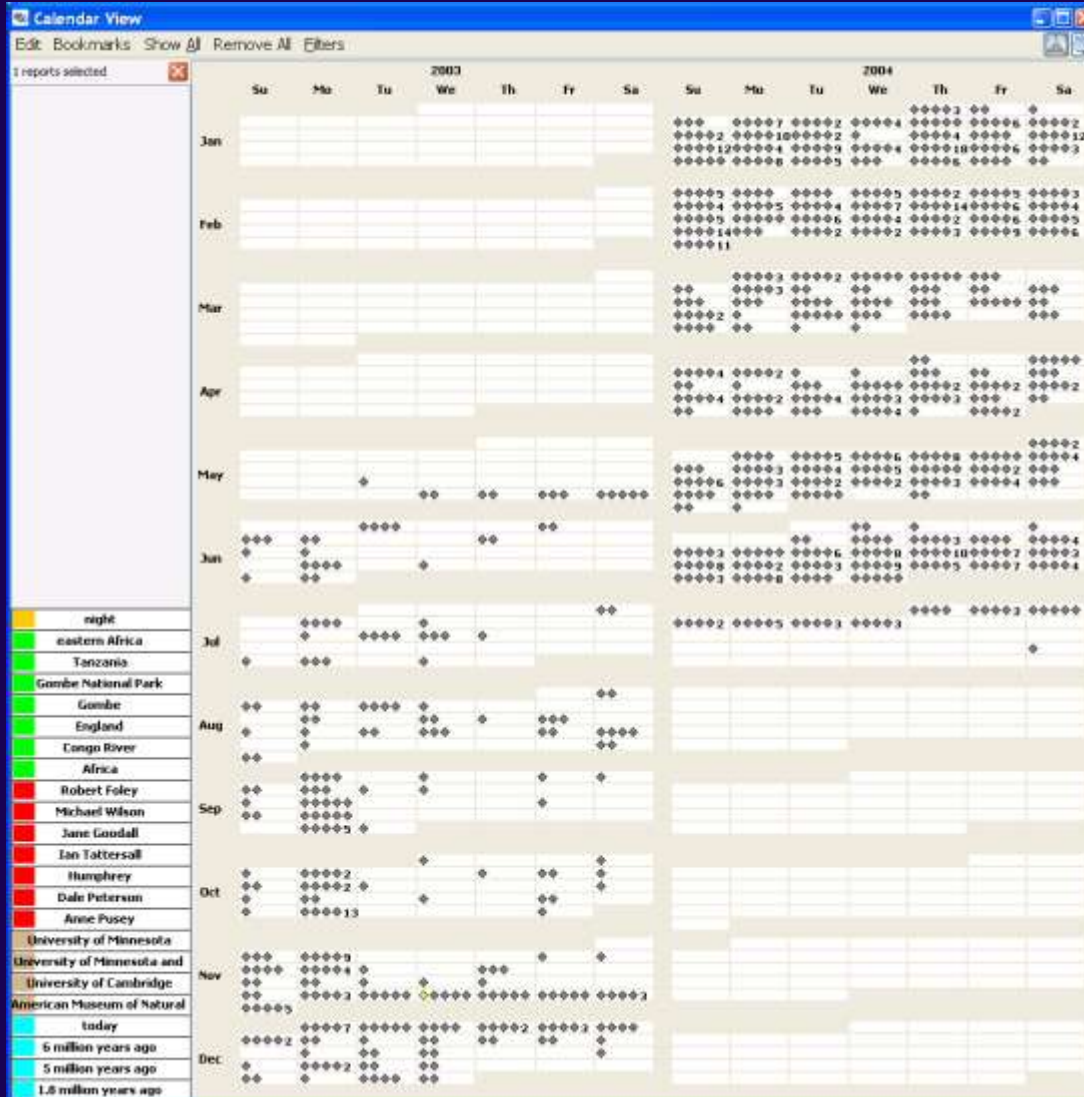
Graph View



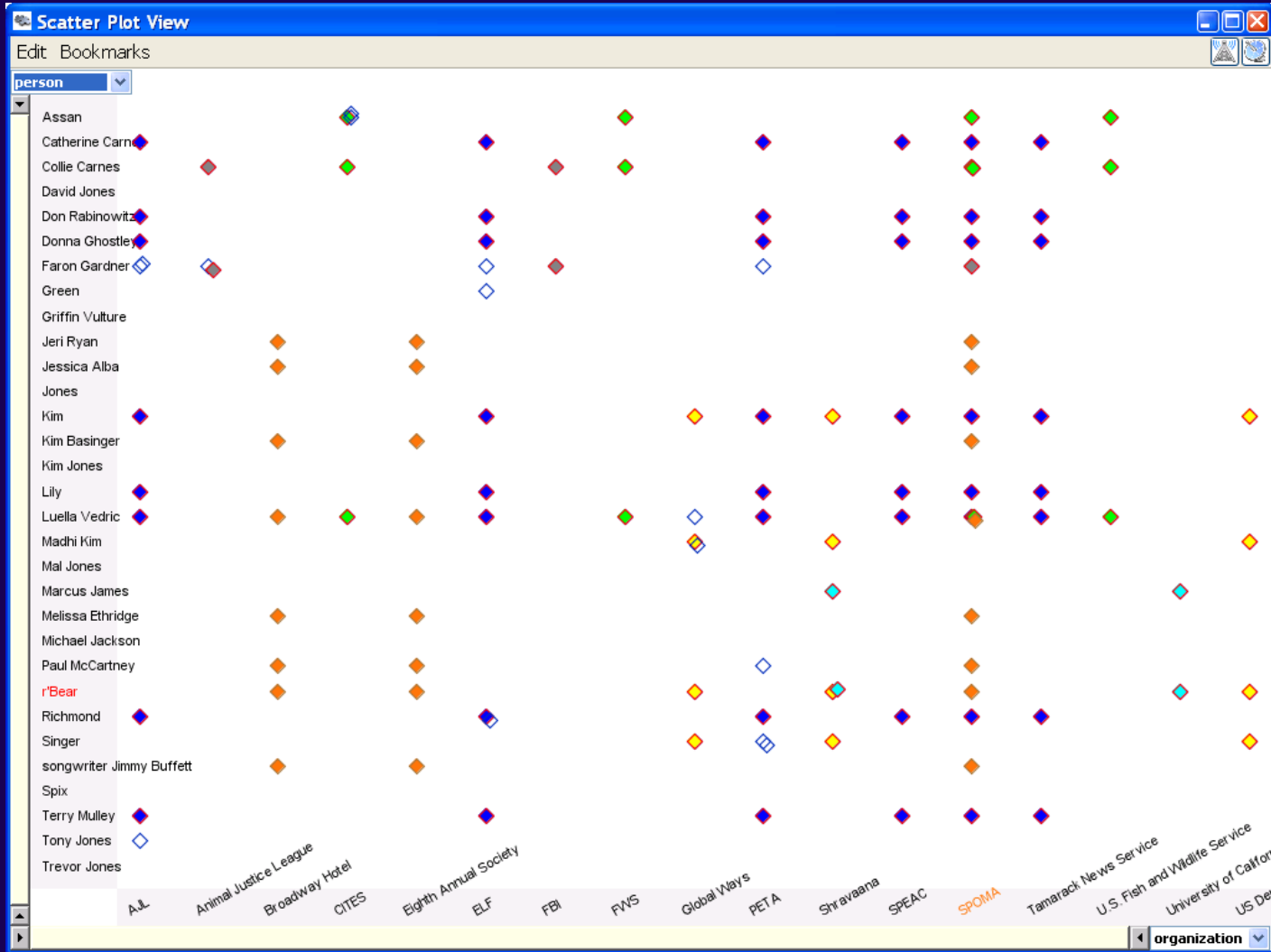
Report Cluster View



Calendar View



Scatterplot View



Timeline View



Related Work

- GeoTime, TRIST, & SANDBOX: Oculus
- Analyst's Notebook: i2
- KANI, Prospekt: PNNL
- WebTAS: ISS Inc.
- CorpusView & Entity Workspace: PARC



What It Does & Does Not Provide

- Is not a “theme” view
- Is a visual index onto documents and entities
- Must read the reports!



Contributions

- Individual view representations and interactions – focus on entity connections
- Advocacy/illustration of using multiple views and large screen space
- Relatively simple UI interactions
- Synthesis into a working system providing investigative analysis capabilities



New Developments



Document Management

Import

Source

Files | Jigsaw Projects | Web Sites | Web Search | DHS Reports | BibTex Files

Documents (.txt .pdf .doc .xls .htm .html)

Files:

Entity Identification

None LingPipe Calais GATE Docs per batch (GATE)

Define mapping for Excel file "2009-040811-0001_Working Copy.xls", Sheet "4048223559"

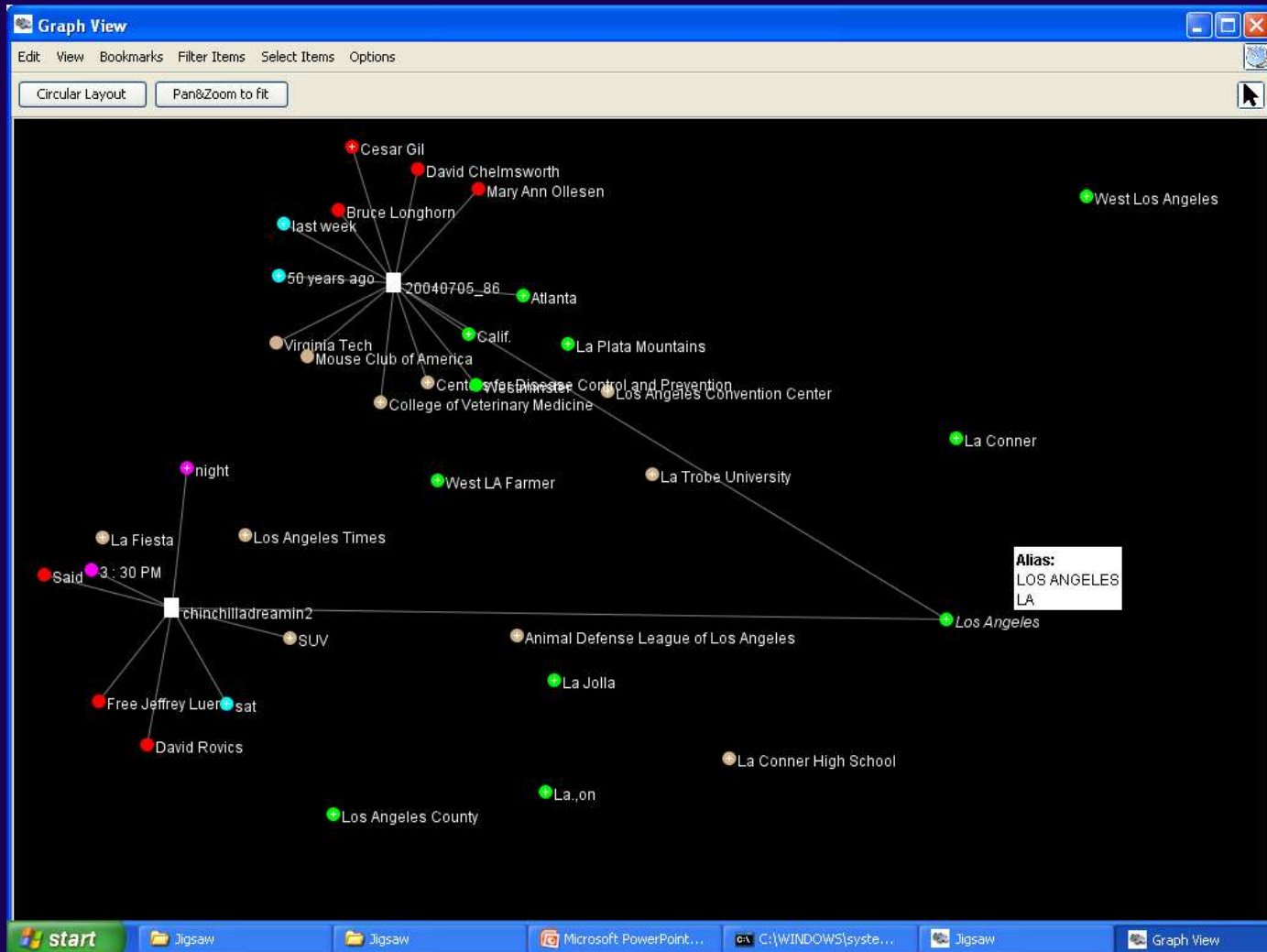
Data starts on row:

	A	B	C	D	E	F	G	
	Ignore	Ignore	Ignore	Ignore	Ignore	Ignore	Ignore	Ignore
1	CALLING_NBR	CALLED_NBR	DIALED_DIGITS	M_R_#	START_DATE	END_DATE	DURATION	REPOLL
2	4.046606159E9	4.048223559E9	4.048223559E9	Inbound	01-Aug-2008	01-Aug-2008	1057.0	624.0
3	8.645672861E9	1.14048223559E11	4.048223559E9	Routed_Call	01-Aug-2008	01-Aug-2008	30.0	624.0
4	4.048223559E9	6.785311024E9	6.785311024E9	Outbound	01-Aug-2008	01-Aug-2008	74.0	624.0
5	6.786773549E9	4.048223559E9	4.048223559E9	Inbound	01-Aug-2008	01-Aug-2008	345.0	624.0
6	7.703040864E9	4.048223559E9	4.048223559E9	Undetermined	01-Aug-2008	01-Aug-2008	2.0	624.0
7	6.785799544E9	1.14048223559E11	4.048223559E9	Routed_Call	01-Aug-2008	01-Aug-2008	29.0	624.0
8	7.703040864E9	1.14048223559E11	4.048223559E9	Routed_Call	01-Aug-2008	01-Aug-2008	31.0	624.0
9	6.785799544E9	1.14048223559E11	4.048223559E9	Routed_Call	01-Aug-2008	01-Aug-2008	29.0	624.0
10	6.785799544E9	1.14048223559E11	4.048223559E9	Routed_Call	01-Aug-2008	01-Aug-2008	29.0	624.0
11	4.048223559E9	8.645672861E9		Outbound	01-Aug-2008	01-Aug-2008	0.0	295.0
12	6.786773549E9	4.048223559E9		Inbound	01-Aug-2008	01-Aug-2008	0.0	293.0
13	8.645672861E9	4.048223559E9		Inbound	01-Aug-2008	01-Aug-2008	0.0	293.0
14	4.048223559E9	6.786773549E9		Outbound	01-Aug-2008	01-Aug-2008	0.0	293.0
15	6.785799544E9	1.14048223559E11	4.048223559E9	Routed_Call	01-Aug-2008	01-Aug-2008	5.0	624.0

More flexible document import & entity identification



Entity Aliasing



Recommend Related Entities

Entities recommended for: Los Angeles Times, Cesar Gil

date	money	organization	person	place	time
today last year Wednesday last week Thursday Friday 2003 last month Last year this month last fall		PETA Humane Society FBI USDA U.S. Department of Ag... ELF Creutzfeldt Jakob Fund for Animals Chiron FDA SPOMA	Dennis Kucinich Kucinich Faron Gardner Michael Markarian Collie Carnes Robert L. Ehrlich Jr. Frans de Waal Sarah Brosnan	United States U.S. Washington America Texas California Europe Africa Los Angeles San Francisco Ohio	last night late night

Recommended entity: **Faron Gardner**

Path: **Los Angeles Times** - {20030714-2_25} - **Animal Justice League** - {20030602-1_66 / 20030818_23} - **Faron Gardner**

Path: **Cesar Gil** - {20030609_4} - **Faron Gardner**

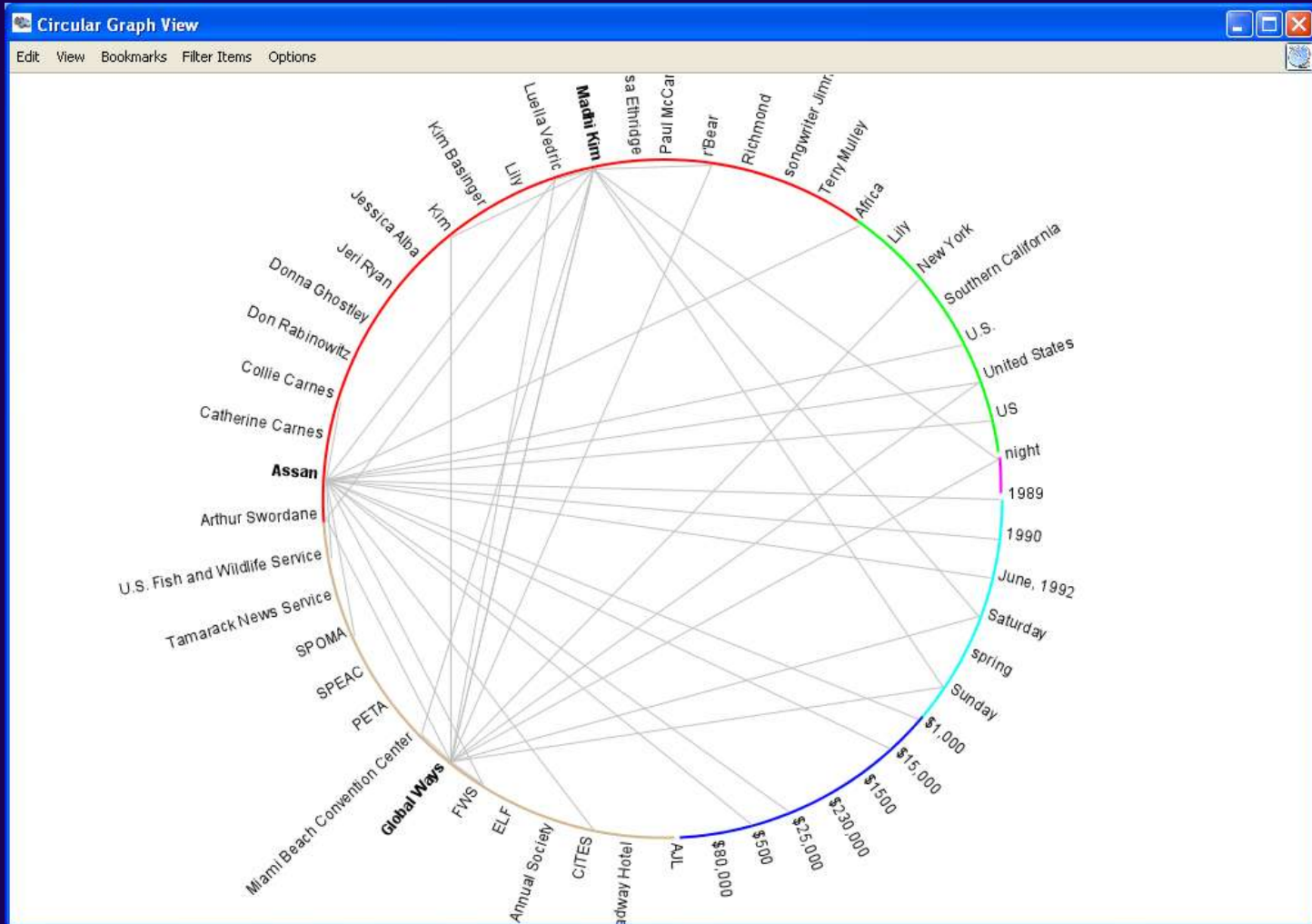


Amplifying Automated Analysis

- Auto-summarization
- Analyst can ask “What’s similar?” or “What’s different?”
- Can the system suggest hypotheses about the data?
 - Difficult problem, in general



Entity Connection View



WordTree View

Word Tree View

Edit Bookmarks

<< Back Forward >> dog Search

572 matches

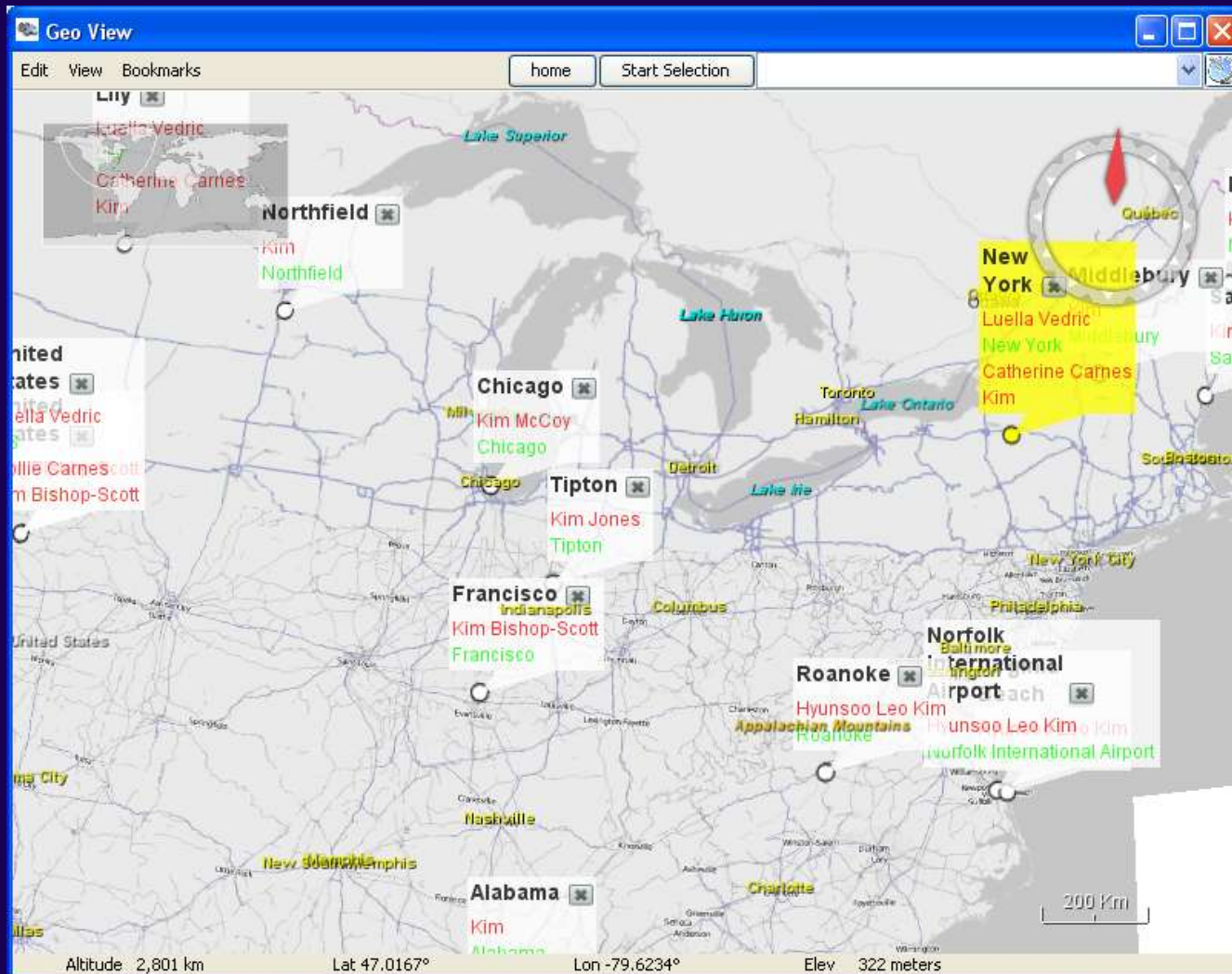
dog

- which Ford had taken in only a week before because its owner could no longer cope with it.
- a Lab mix that, Gretta says, "needs a lot of attention.
- but did she murder the chickens?
- " says Tom Burns.
- recently diagnosed with Addison's disQ.
- running at large, he bit a 10-year-old boy minutes after he attacked the girl," Caldwell said.
- life could be considered paradise on Earth,with an indulgent baby-boomer generation creating such institutions asdoggie day care, doggie spas and doggie vacations.
- 's behavior of barking and running, as well as the Zambitos' practice of keeping Cecil in the kitchen, away from visitors, should have been evidence enough the dog was capak
- and cat fur" and to protect consumers from buying falsely labeled products.
- 1 cat available at the pound and there are many cats and kittens available at PETSMART on Route 2, Warwick.
- food , a far cry from the scraps he had barely subsisted on.
- at \$9.
- is declared dangerous," Mr.
- a happy dog.
- owners topurchase licenses for their pets, as proof that the animal has beenvaccinated, but only about one fifth of all dog owners actually comply.
- are strongly urged to contact the City Council and express their opposition to the ordinance.
- will be reunited with you if they ever become lost.
- or cat always wears a collar with identification tags even if it has been microchipped or tattooed.
- to a couch potato," Smith said.
- run free.
- had a tracking microchip implanted under its skin.
- in the household, should be protected from fleas that may cause painful diseases such as flea allergy dermatitis.
- that had been hit by the truck was their own dog Mellow, who had been across the street playing with other dogs.
- a young, white, female mongrel - was found scared and whimpering in a trash bin at the Landis Gardner factory on East Sixth Street by a company security guard.
- has already bitten, he said.
- Owner Association, to speak for them.
- furimports and exports, he said.
- on a leash and encouraged Collier to allow the dog to sniff his hand, as the dog had known him from prior visits, according to court papers.
- when they came to pick her up.
- owners said they would be more loyal to a hotel chain that accepted dogs, even if they weren't traveling with their dog.
- bites annually in Colorado, 1,400 will be prosecuted and 467 will be required to pay the fee.
- control laws which come into force on Monday.
- fighting ring that was uncovered by the Edgecombe County Sheriff's Department on Monday.
- from the Muncie Animal Shelter and the Delaware County commissioners, who could only suggest that he take the animal to the Randolph County Animal Shelter.
- into a store or restaurant, that is often not the case with service dogs that assist those with other disabilities, especially when the disability is not visible, experts say.
- pound is in operation, he said.
- who benefited from Andy?
- with a bit of the Bard in his bark.
- would go where the children went.
- was capable of attacking.

start Jigsaw Jigsaw Microsoft Power... C:\WINDOWS\... Jigsaw Graph View Word Tree View



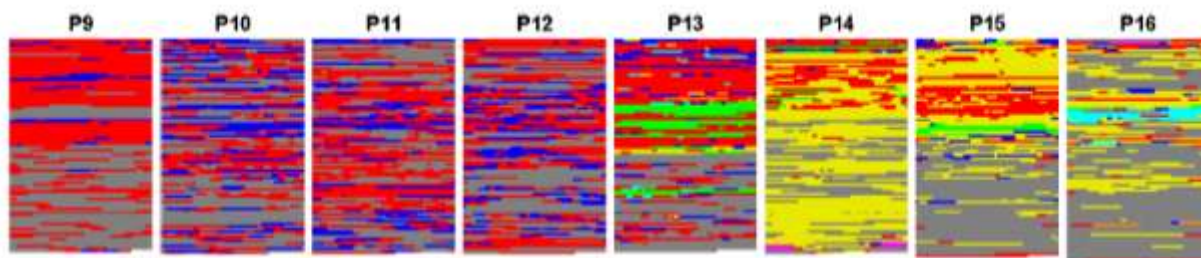
GeoSpatial View



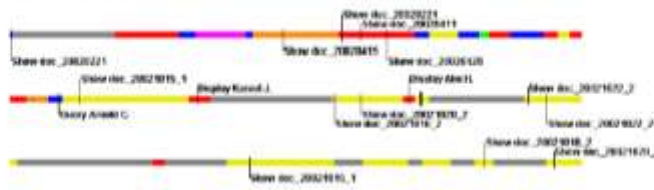
Evaluation

Overview, filter & detail
 Build from detail
 Hit the keyword
 Find a clue, follow the trail

	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Grading Task Sheet	-1.75	17	4	13	13	10.5	10.5	-3.5	5.5	-8.25	4	7.5	14.5	13.5	7	17
Grading Debriefing	2	2.5	1	5.5	3	4	1.5	3	3.5	2.5	1.5	6	6	2.5	5.5	5
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
Avg. Score/Setting	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
Number 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	78:48
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
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT



- Main View
- Document View
- List View
- Graph View
- Calendar View
- Document Cluster View
- Timeline View
- Task Sheet



To appear:
 Kang, Görg & Stasko
 VAST '09



Key Focus

- Transitioning to real (field) use
- Bundle the system into a package for download and use
- Available upon request

- Work in progress...



WAJAC Use

"Recently, the Washington Joint Analytical Center (WAJAC) assisted a local police department in analyzing 345 incident reports pertaining to graffiti. Jigsaw was an invaluable tool when it came to examining reports. Without the use of Jigsaw, examining reports would have been tedious and time consuming, and connections between entities may not have been seen.

The local police department was very impressed with Jigsaw and stated that the program will become very useful in tracking graffiti activity, and possibly other criminal activity.

WAJAC uses Jigsaw on a continuous basis to examine information in our intelligence database. Jigsaw is also used to analyze incidents relating to the security of the Washington State Ferry System. It provides a searchable database and quick analytical tool for tracking potential threats to the largest ferry transportation system in the U.S."

CHAD R MELTON
Criminal Intelligence Analyst
WAJAC/FBI-FIG



Cold Case

- Worked on an active police cold case
- About 1000 documents
 - Scanned and OCR'ed
- Some pluses and minuses



New Domains

- Bioinformatics & genomics
 - Many, many articles about a gene --- How do they connect and relate?
 - Entity-Entity relationships now matter more
- Air traffic statutes: US – Europe
- Legal e-discovery
- ...



Issues/Questions/Reflections

- Where do I start?
- All those views!
- Entity Identification, entity identification, entity identification, entity identification...
- Straightforward UIs matter!



Future Work

- “Thinking” view (evidence marshalling)
- Investigative history browser
- Amplified computational analysis
- Entity connection types
- Representing reliability & uncertainty



Reflections on Document Analysis

- Definitely can generalize out from investigative analysis
- Visual analytics can really help
- Issues...



Document Analysis Issues

- 1. Understand the task & use
 - Intelligence analysis
 - Identify the threat
 - Understand a situation
 - Others?



Document Analysis Issues

- 2. Need more text & linguistics analysis
 - Leverage computational linguistics
 - E.g., sentiment analysis
 - Collaborate with colleagues who are experts in this area



Document Analysis Issues

- 3. Handle dynamic collections of documents
 - Static collection vs. document stream
 - Adapt to document feeds & crawling/trolling



Document Analysis Issues

- 4. Help analysts manage the investigative state better
 - Helping analysts resume and return to earlier points
 - What else do they really want to do?



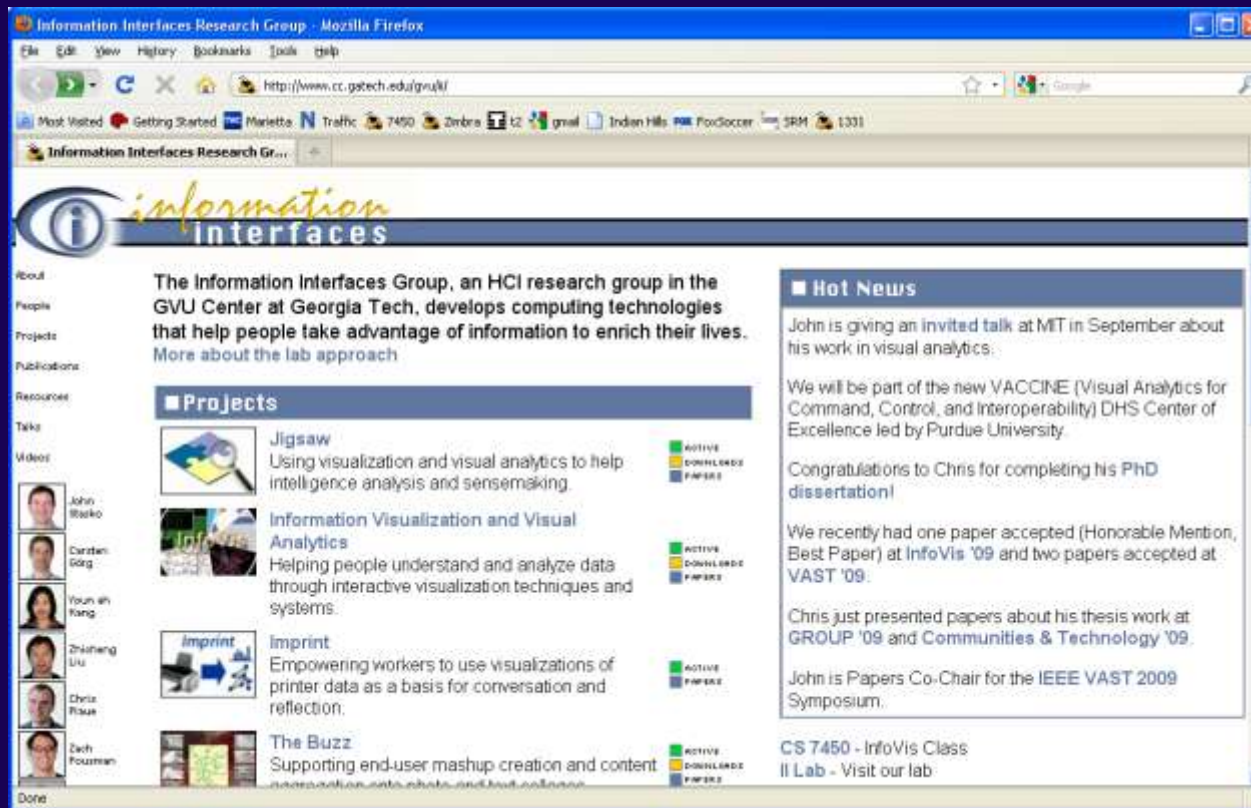
Document Analysis Issues

- 5. Include many types of documents
 - Blogs, spreadsheets, press releases, Twitter, Wikipedia, patents, academic papers, ...
 - Can each's structure/form be exploited?
 - Important to understand what an analyst would want to do with each



To Learn More

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



The screenshot shows a Mozilla Firefox browser window displaying the website for the Information Interfaces Research Group. The browser's address bar shows the URL <http://www.cc.gatech.edu/ii/>. The website features a navigation menu on the left with links for About, People, Projects, Publications, Resources, Talks, and Videos. The main content area includes a header with the group's logo and name, a central text block describing the group's mission, and a 'Projects' section listing several ongoing projects: Jigsaw, Information Visualization and Visual Analytics, Imprint, and The Buzz. Each project entry includes a small icon, a brief description, and status indicators for 'ACTIVE', 'DOWNLOADS', and 'PAPERS'. On the right side, there is a 'Hot News' section with several news items, including an invited talk at MIT, participation in the VACCINE center, a PhD dissertation completion, and paper acceptances at InfoVis '09 and VAST '09. At the bottom right, there is a link to 'CS 7450 - InfoVis Class' and a 'Visit our lab' link. The browser's status bar at the bottom indicates 'Done'.



Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the new 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?

