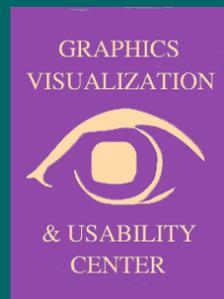


# Jigsaw: Supporting Investigative Analysis through Interactive Visualization

John Stasko\*, Carsten Görg,  
Zhicheng Liu, Kanupriyah Singhal

Information Interfaces Research Group  
School of Interactive Computing & GVI Center  
Georgia Institute of Technology  
Southeastern RVAC

\*Microsoft Research



# The Team

- Carsten Görg  
Zhicheng Liu  
Vasili Pantazopoulos  
Neel Parekh  
Kanupriyah Singhal  
Gennadiy Stepanov  
Sarah Williams



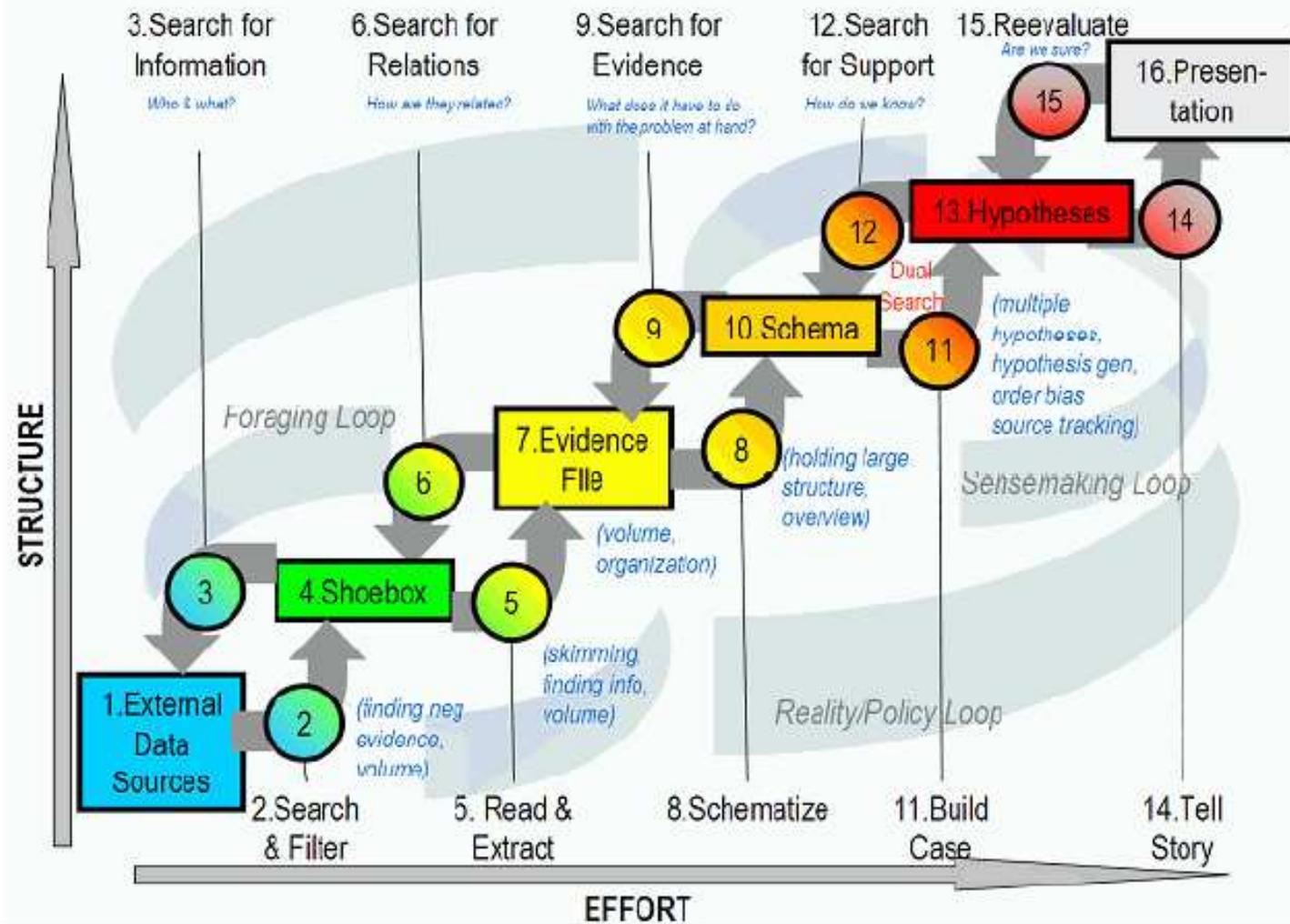


Figure 2.1. Notional model of sensemaking loop for intelligence analysis derived from CTA.

Pirolli & Card, ICIA '05



# Pain Points

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



# The Problem

- Assist investigative analysts in information foraging and sense-making given a multitude of documents/reports
- Help them discover plans and plots embedded throughout the documents, plots difficult to find by simply examining and reading the documents



# An Exemplary Report

**Report Date: 12 May, 2003 [From Fulton County PD].**

A package of beef and a letter from "Animal Justice League" claiming that meat had been poisoned in 20 Los Angeles supermarkets was left at the Los Angeles Times, 1st Street offices. The paper also received a phone call taking credit for the action and stating, "We will take direct action against animal abuse in whatever form is necessary to stop the cruelty." A similar threat was made against a supermarket in Upsala in November. No poisoned meat was found at either supermarket.

What if there are  
thousands of these?



# Our Focus

- Entities within the documents
  - Person, place, organization, phone number, date, license plate, etc.
- 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



# Jigsaw

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

“Putting the pieces together”



# Connections

- Entities relate/connect to each other to make a larger “story”
- Simple definition:
  - Two entities are connected if they appear in a report together



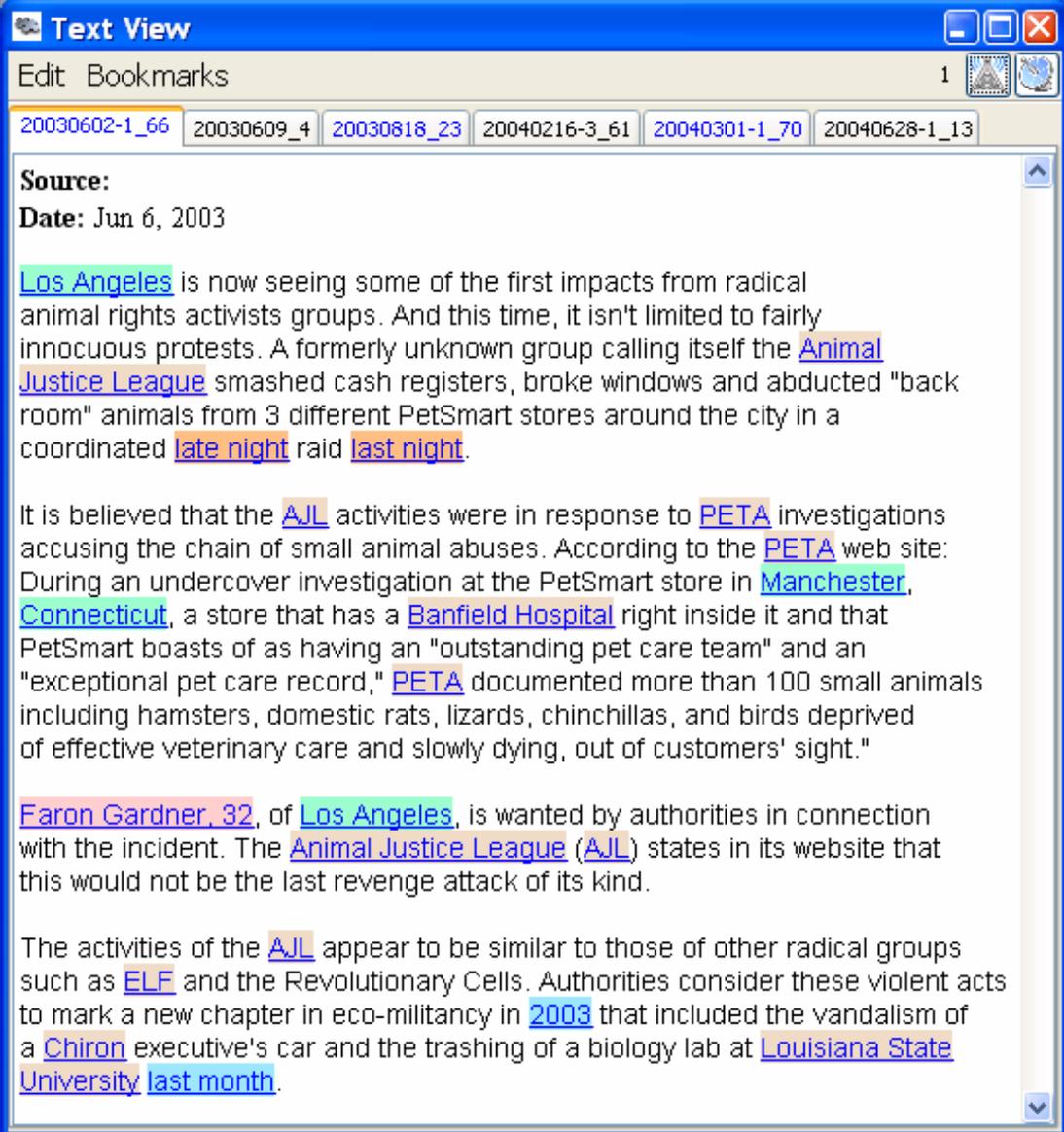
# The Need for Pixels



# Console



# Text View



**Text View**

Edit Bookmarks 1

20030602-1\_66 20030609\_4 20030818\_23 20040216-3\_61 20040301-1\_70 20040628-1\_13

**Source:**  
**Date:** Jun 6, 2003

[Los Angeles](#) is now seeing some of the first impacts from radical animal rights activists groups. And this time, it isn't limited to fairly innocuous protests. A formerly unknown group calling itself the [Animal Justice League](#) smashed cash registers, broke windows and abducted "back room" animals from 3 different PetSmart stores around the city in a coordinated [late night](#) raid [last night](#).

It is believed that the [AJL](#) activities were in response to [PETA](#) investigations accusing the chain of small animal abuses. According to the [PETA](#) web site: During an undercover investigation at the PetSmart store in [Manchester, Connecticut](#), a store that has a [Banfield Hospital](#) right inside it and that PetSmart boasts of as having an "outstanding pet care team" and an "exceptional pet care record," [PETA](#) documented more than 100 small animals including hamsters, domestic rats, lizards, chinchillas, and birds deprived of effective veterinary care and slowly dying, out of customers' sight."

[Faron Gardner, 32](#), of [Los Angeles](#), is wanted by authorities in connection with the incident. The [Animal Justice League \(AJL\)](#) states in its website that this would not be the last revenge attack of its kind.

The activities of the [AJL](#) appear to be similar to those of other radical groups such as [ELF](#) and the Revolutionary Cells. Authorities consider these violent acts to mark a new chapter in eco-militancy in [2003](#) that included the vandalism of a [Chiron](#) executive's car and the trashing of a biology lab at [Louisiana State University last month](#).

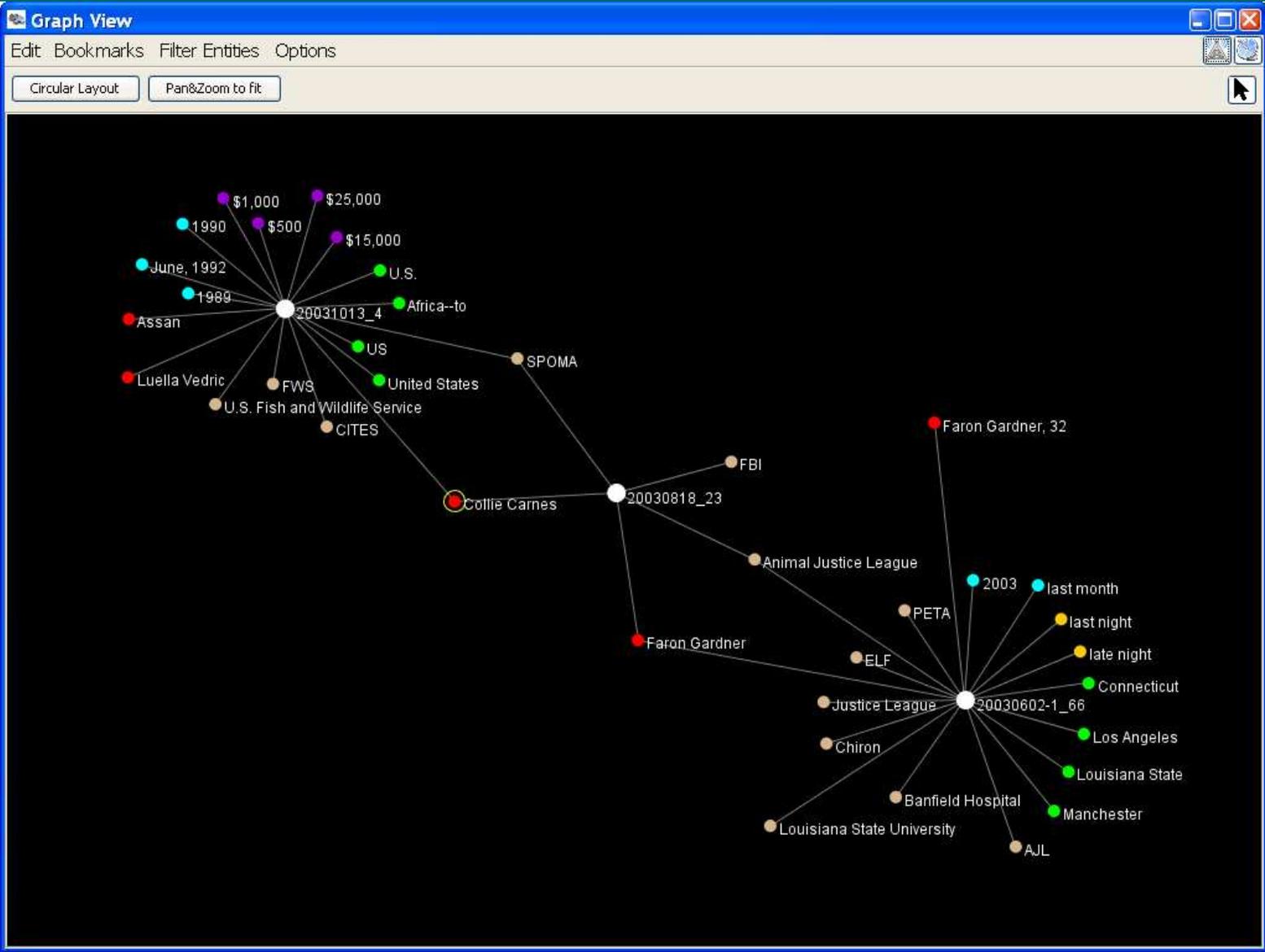


# List View

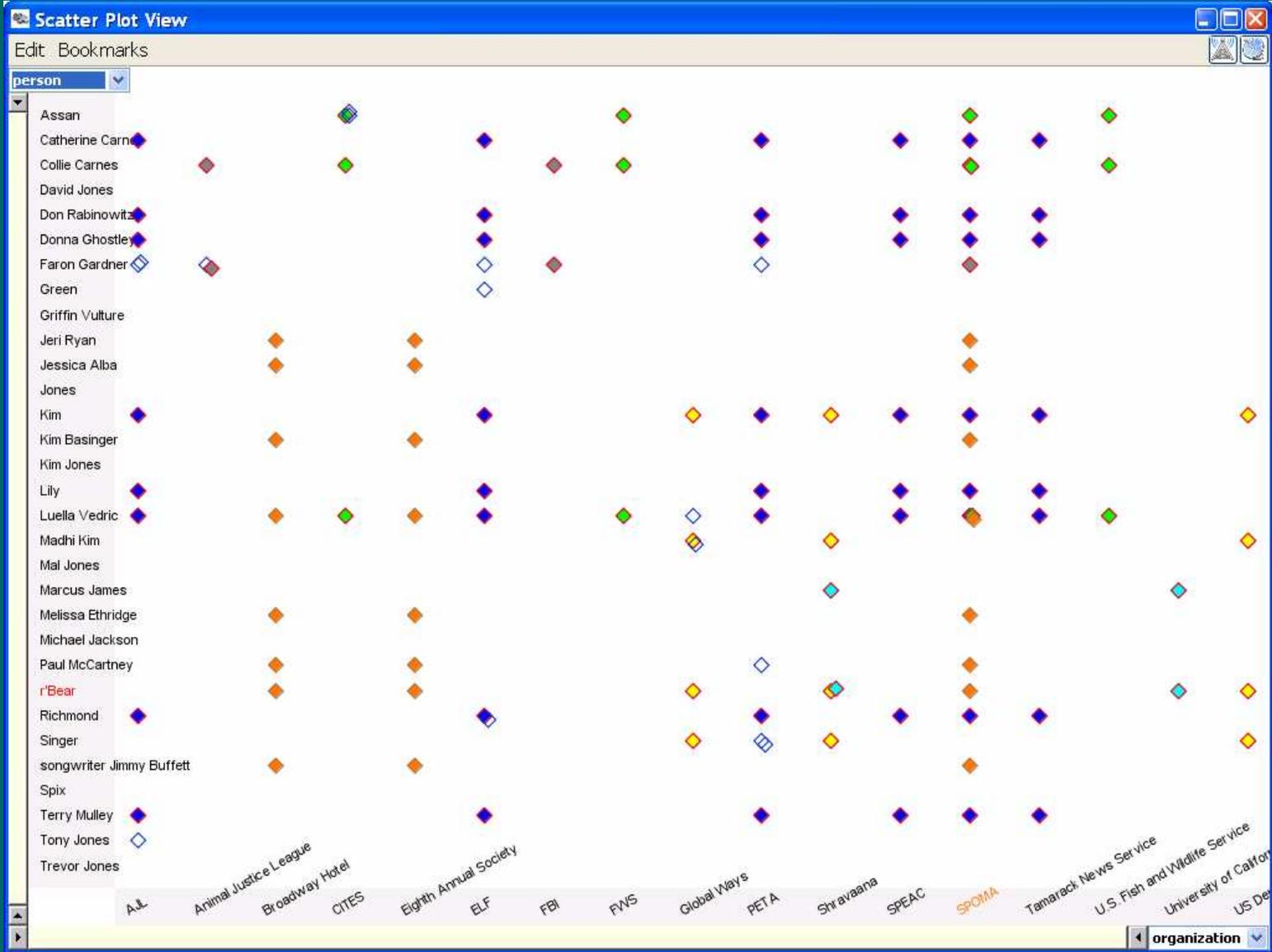
The screenshot shows a 'List View' application window with a blue title bar and standard window controls. The interface is divided into two main panes: 'person' on the left and 'organization' on the right. Both panes have a 'Value' dropdown menu, 'Move active up', 'Unselect', 'Add All', and 'Clear' buttons. A 'Show all connections' button is located between the panes. The 'person' list includes names like Arthur Swordane, Catherine Carnes, Collie Carnes, Don Rabinowitz, Donna Ghostley, Jeri Ryan, Jessica Alba, Kim, Kim Basinger, Lily, Luella Vedric, and Madhi Kim. The 'organization' list includes entities like AJL, Animal Justice League, Banfield Hospital, Broadway Hotel, Chiron, CITES, Delta Society, Eighth Annual Society, ELF, FBI, FWS, Global Ways, Justice League, Louisiana State University, Miami Beach Convention Center, Northfield Elementary School, Pet Partners, PETA, Shravaana, SPEAC, SPOMA, Tamarack News Service, U.S. Fish and Wildlife Service, University of California Medical Center, and US Department of Agriculture. Lines connect specific individuals to various organizations, illustrating a network of relationships.



# Graph View

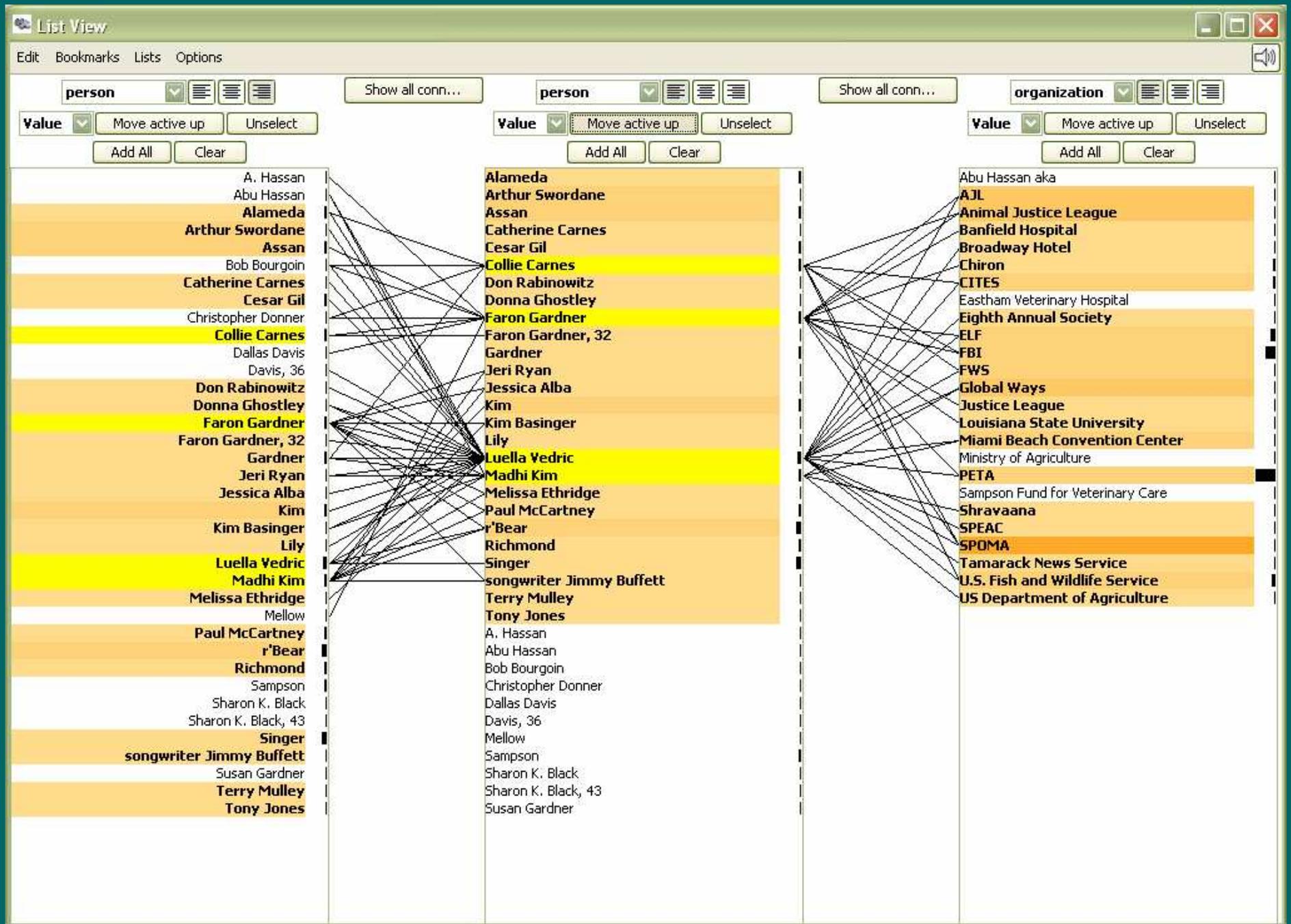


# Scatterplot View









## Use tips - 2







# What It Does & Does Not Provide

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



# Related Work

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

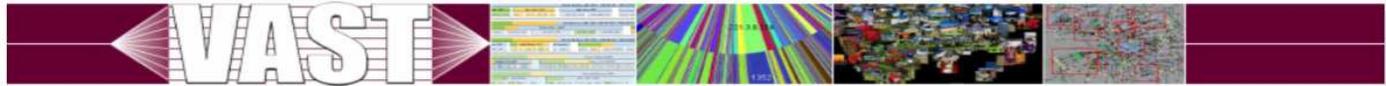


http://www.cs.umd.edu/hcil/VASTcontest07/ - Windows Internet Explorer

http://www.cs.umd.edu/hcil/VASTcontest07/ Live Search

Links Google Gmail Marietta Redmond MSN Maps KPLU 88.5 KWJZ 98.9

Gmail - Inbox (5) http://www.cs.umd.edu...



IEEE Symposium on Visual Analytics Science and Technology 2007 • October 30 to November 1, 2007

## VAST 2007 CONTEST

### Blue Iguanodon

**News:**  
 we posted the **RESULTS and winners!**  
 We agreed not to post the solution until the end of the semester because the contest problem is being used in class projects this fall  
 Two page summaries from the teams will be posted here by the time of the symposium.  
 See [History of changes and FAQs](#)

#### Overview of the event

The VAST Contest is a participation category of the [IEEE VAST 2007 Symposium](#) (part of [VIS 2007](#)). It continues in the footsteps of the [VAST 2006 contest](#) as its purpose is to promote the development of benchmark data sets and metrics for visual analytics, and to establish a forum to advance visual analytics evaluation methods.

The dataset consist of news stories and blog entries, along with background information and some multimedia materials (small maps and data tables). Participants are to investigate a major law enforcement/counter-terrorism scenario, form their hypotheses, and collect supporting evidence. There are several tasks that teams will need to tackle on the way. These include 1) processing the text and multimedia information to identify entities of interest (e.g. people, places and activities), 2) depicting this information visually using interactive visualizations and other tools to aid in the analysis of the information; 3) answering the contest questions based on the analysis; and 4) producing a video demonstration of their system showing how they arrived at those answers.

The initial data set processing may be accomplished either by using software written by the teams, by using commercial software, by partnering with another organization, or by using the preprocessed version provided by the contest organizing committee.

Contest entries will be judged on both the correctness of the analysis of the situation and the utility of the tools in conducting the analysis.

Participants will have several months to prepare their submissions. Selected entries will present their work at the conference. A small number of teams will be invited to participate in a special closed-door session. In that session they will work on a new but similar problem with professional analysts who will provide feedback and guidance to help refine the tools.

#### How does it compare to last year?

The general format and schedule of the event remains the same.  
 Most of the raw data will use the same format as last year, so you can use the 2006 materials to see how you are doing since the answers are available for 2006.

New this year: top entries will coauthor a journal paper. We also provide a pre-processed version of the dataset (e.g. with entities already extracted) so that more teams can participate. Your team will need to decide whether to enter the contest using the preprocessed data or the raw data.

#### Data set, tasks, and judging

- [INFORMATION about the data set, tasks and judging criteria](#) (Flier to advertise the contest)
- Contest **RESULTS and winners**
- [Register and DOWNLOAD the RAW DATA SET for 2007](#)
- [Register and DOWNLOAD the PREPROCESSED DATA SET for 2007](#) (**STOP:** read the INFORMATION page above first)

Done Internet | Protected Mode: On 100%



# Contributions

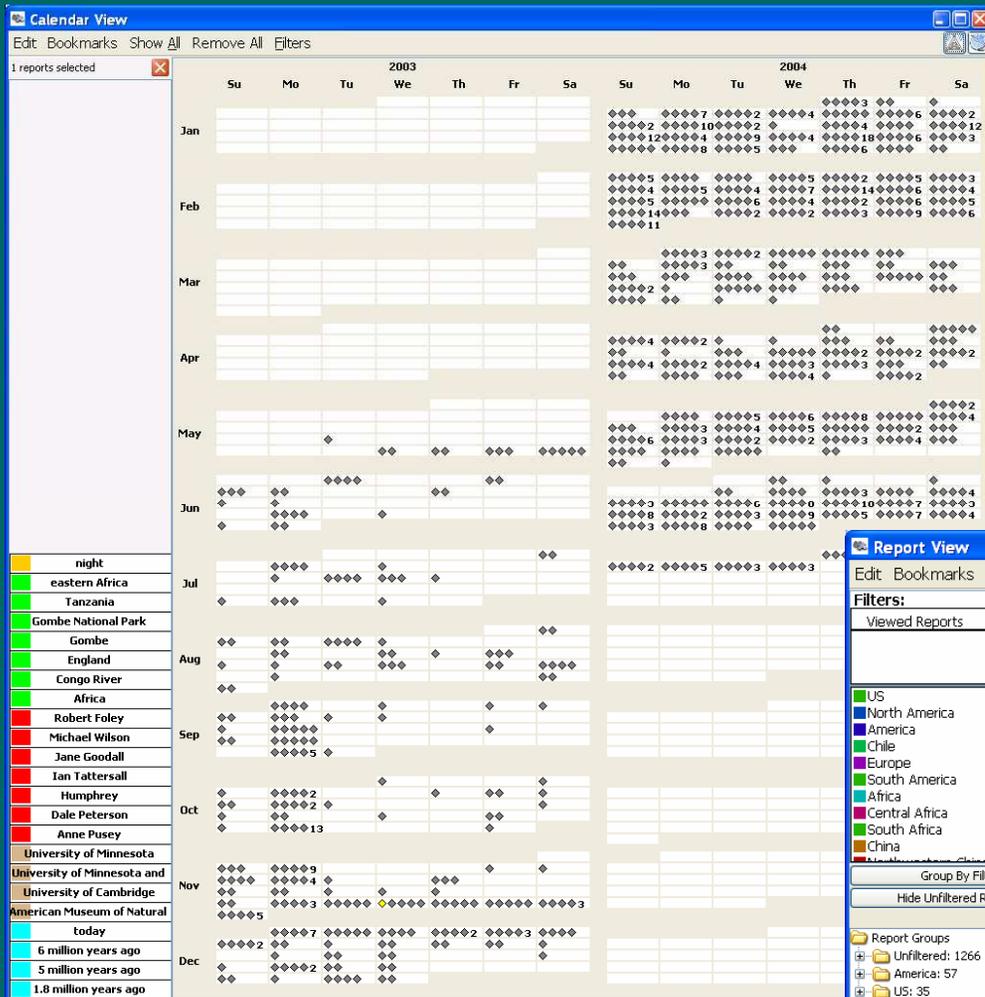
- Individual view representations and interactions – focus on entity connections
- Advocacy/illustration of using multiple views and large screen space
- Relatively simple UI interactions
- Synthesis into a working system providing investigative analysis capabilities



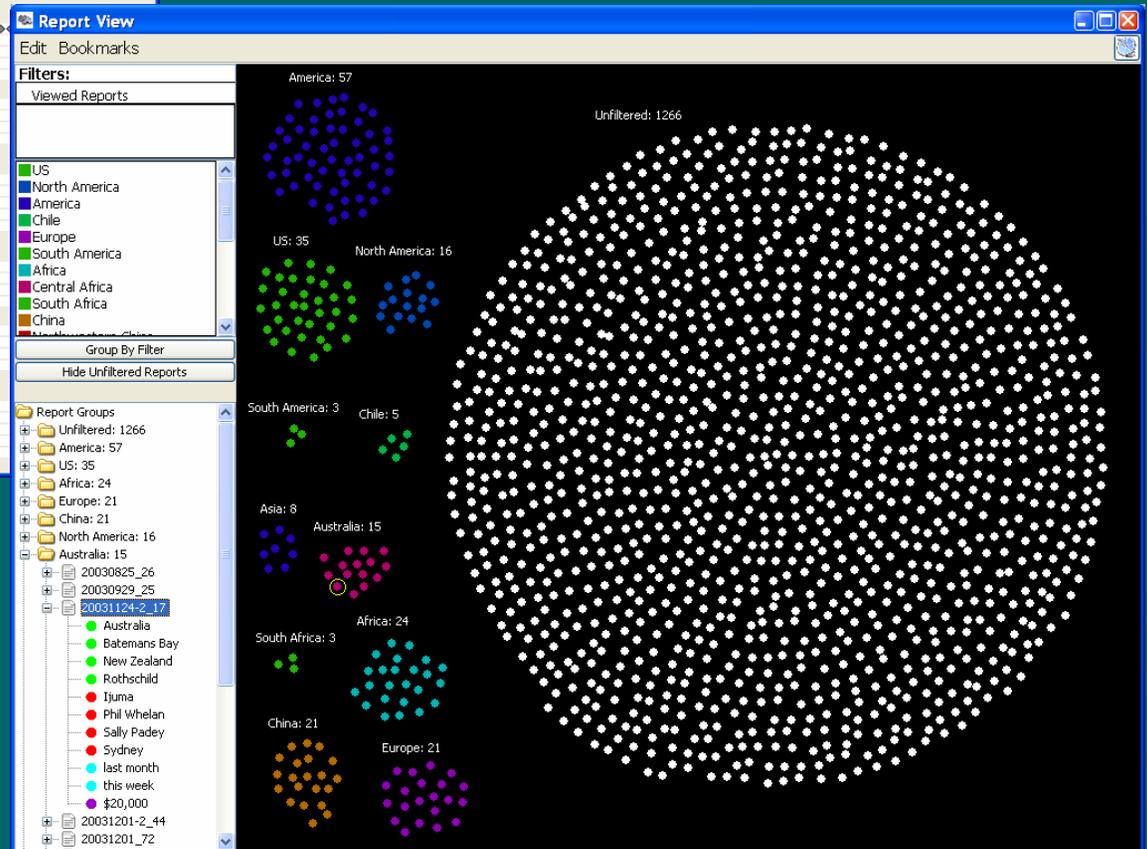
# Future Work/Current Status

- In progress
  - Integrate entity analysis and dynamic updates
- Much done
  - Improve view representations & interactions
- Yes and more
  - Other/more views?
- Ongoing
  - Better evidence marshalling and documentation support
- Ongoing
  - Evaluation and trials by analysts
- Ongoing
  - Other domains of data





# Calendar view



# Report view



# Shoebox

The Shoebox application interface consists of a sidebar on the left, a main workspace, and a top navigation bar. The sidebar contains a list of items categorized by 'UP' and 'DOWN'. The main workspace features a top navigation bar with a search bar and several floating windows. The search bar contains the text 'Kim visited', 'New York last', and 'spring'. The floating windows include:

- Hypothesis: Luella is involved in bad activities**: A window with a list of items including Assan, Collie Carnes, AJL, r'Bear, \$80,000, Global Ways, 20030526-2\_57, 20031013\_4, and 20040119-1\_98. A link 'Graph View - LuellasConnections' is highlighted.
- Global Ways**: A window with a list of items including CDC, Fish and Wildlife Service, US Department of Agriculture, Florida, South America, Ed Parker, and Madhi Kim.
- Chinchilla smuggling**: A window with a list of items including Gil Breeders, Los Angeles, Cesar Gil, Chile, Argentina, Andes, and Peru.

Additional items in the main workspace include 'chinchilla Flocke ???', 'Miami', 'San Diego', and 'Kenya'.



# Larger To-Do

- Visualize themes in collection
- Linguistic/Semantic analysis
- Visualizing uncertainty
- Collaboration



# To Learn More

- Video is on DVD
- Visit
  - <http://www.gvu.gatech.edu/ii/jigsaw>



# Acknowledgments

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



- Supported by NSF IIS-0414667



# End

- Thanks for your attention!
- Questions?

