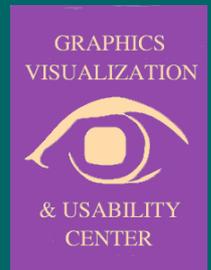


Investigative Analysis – Putting the Pieces Together with Jigsaw and Visual Analytics

John Stasko

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



The Problem

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

Analogy



Example 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.



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



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

“Putting the pieces together”



The Need for Pixels



Demo

Document View

Edit: Bookmarks

20030526-2_57 | 20030602-1_66 | 20030609_4

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**.

Jigsaw

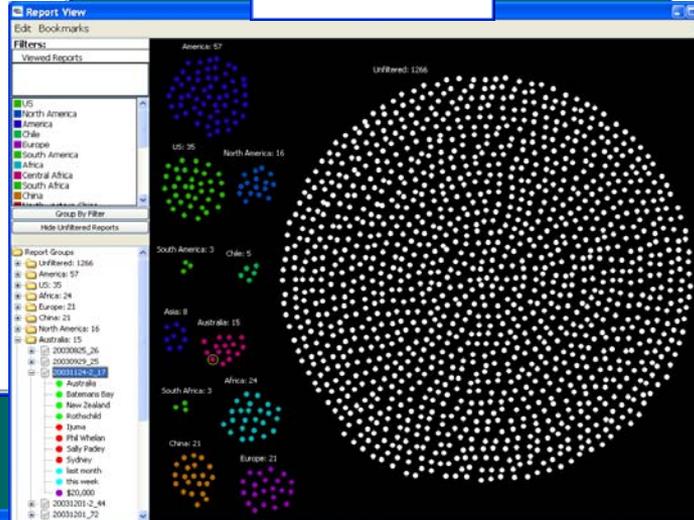
Blue Iganodon

Color Legend:

- place
- organization
- person
- date
- money
- time

vedic Search

Text search | Substring search



List View

person | place | organization

Value | Move active up | Unselect | Add All | Clear

- A. Hassan
- Abu Hassan
- Abu Hassan
- Arthur Sweredale
- Catherine Carnes
- Brandon Tropical Fish
- Collic Carnes
- Collic Carnes
- Don Rabinowitz
- Donna Chavaley
- Ed Parker
- Emma Green
- Griffin Vulture
- Jeri Ryan
- Jessica Alba
- Jillan
- Kim
- Kim Basinger
- Lily
- Luella Vedic
- Madhi Kim
- Marcus James
- Melissa Etheridge
- Michael Jackson
- Paul McCartney
- Richard
- Richmond
- Songwriter Jimmy Buffett

Africa: Alabama, Arkansas, California, Florida, Iowa, Kenya, Liberia, Madagascar, Mali, Mauritania, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Tanzania, Zimbabwe

United States: Abu Hassan aka, A.J.L., Broadway Hotel, CDC, Centers for Disease Control and Prevention, CITES, Delta Society, Department of Health, DOA, Eighth Annual Society, ELF, Fish and Wildlife Service, Florida Department, FWS, FWS Special Agent, Global Ways, Miami Beach Convention Center, Ministry of Agriculture, Northfield Elementary School, Pet Partners, PETA, Sanchez, Shrivastava, SPEAC, SPOMA, Tamarack News Service, U.S. Fish and Wildlife Service, University of California Medical Center, US Department of Agriculture



Shoebots

Report: Florida is involved in bad activities

- Assam
- AJL
- \$80,000
- 20030526-2_57
- 20030602-1_66
- 20030609_4

Global Ways

- CDC
- Fish and Wildlife Service
- US Department of Agriculture
- Florida
- South America
- Kenya
- Madison

Chickadee conspiracy

- GBirders
- Los Angeles
- Central AZ
- Chile
- Argentina
- Ed Parker
- Pariz

Map: Miami, San Diego, Kenya

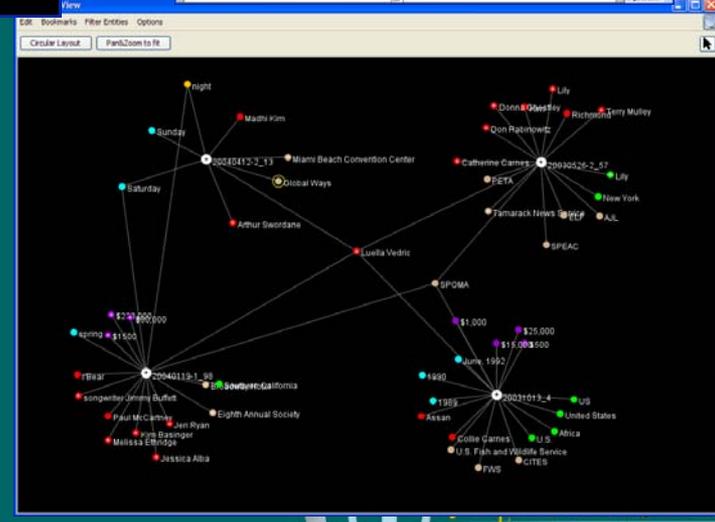
Calendar View

2003

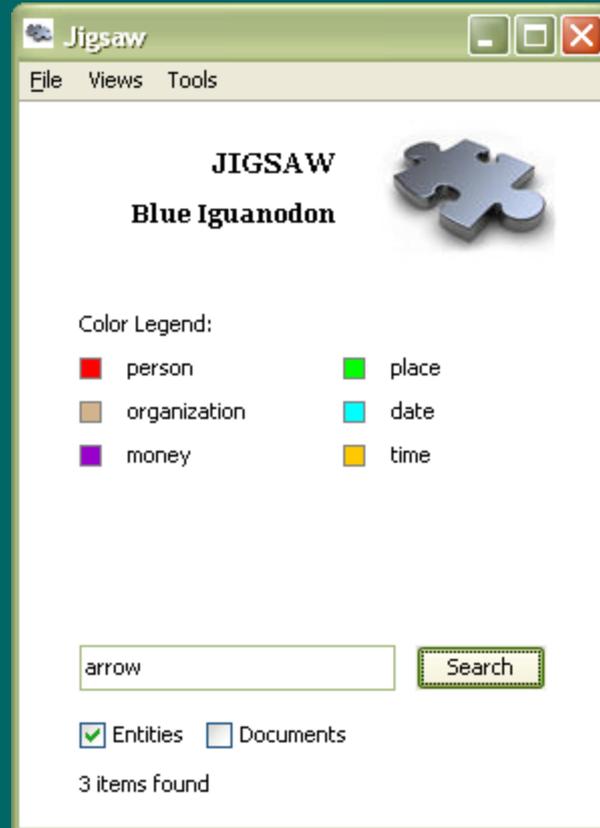
1 recently selected

Month	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Jan							
Feb							
Mar							
Apr							
May							
Jun							
Jul							
Aug							
Sep							
Oct							
Nov							
Dec							

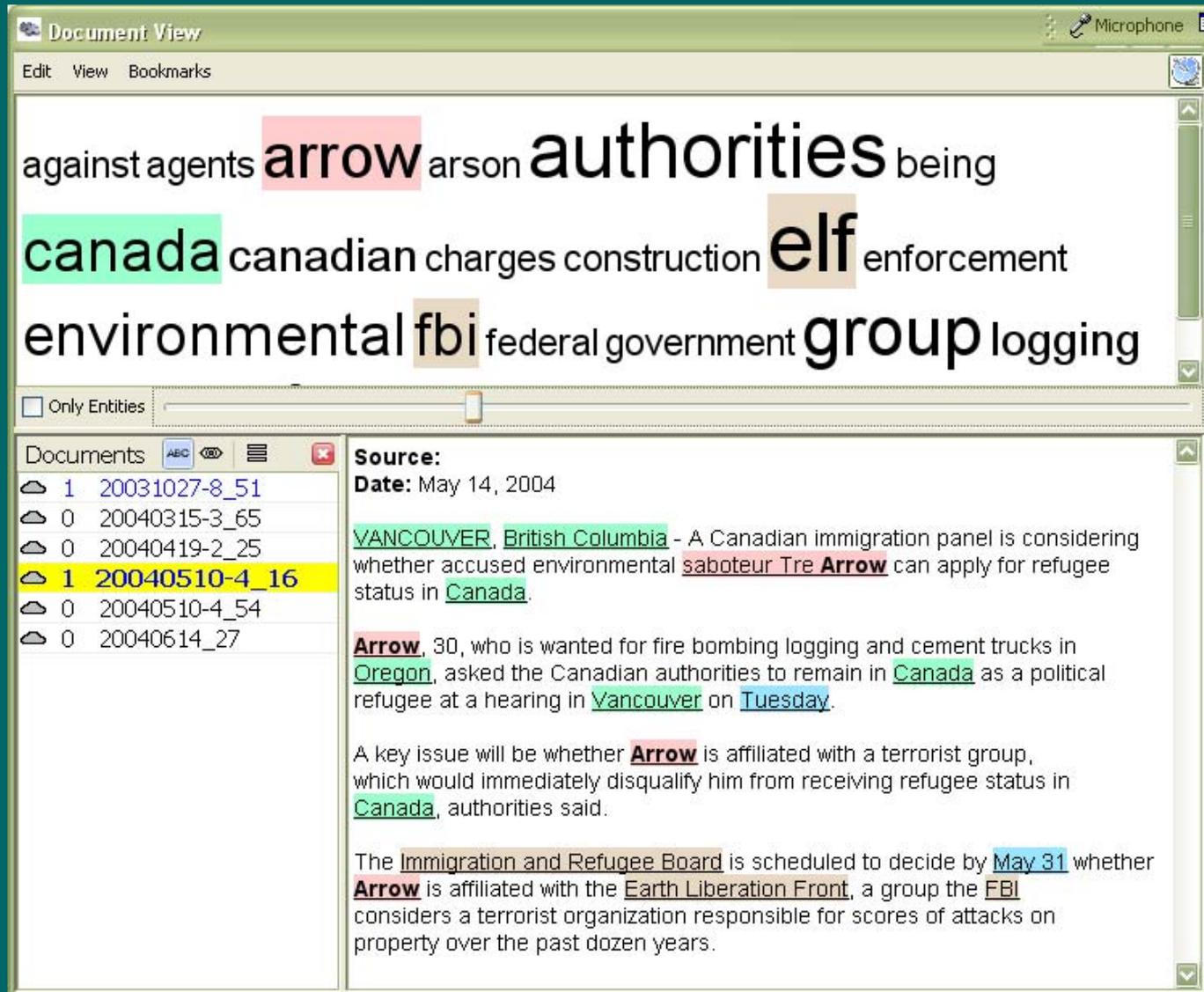
Legend: night, madison Africa, Florida, Gender National Park, England, George River, Africa, Robert Fisher, Michael Wilton, New Zealand, Sam Patterson, Madagascar, Paul McCartney, Anne Paury, University of Tennessee and University of Cambridge, Mexican Museum of Natural History, today, 4 million years ago, 1.6 billion years ago



Console



Document View



Document View

Edit View Bookmarks

Microphone

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.

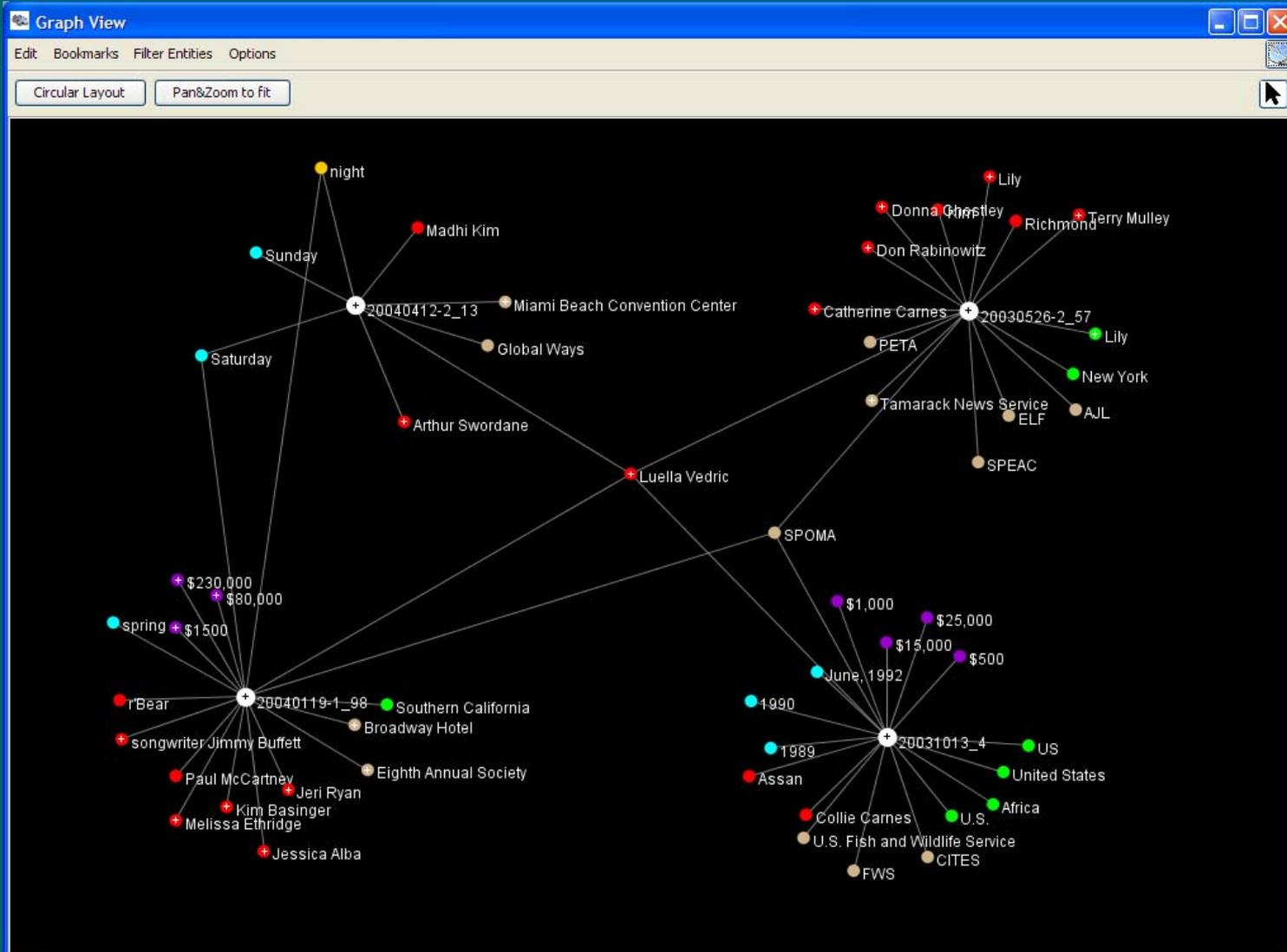


List View

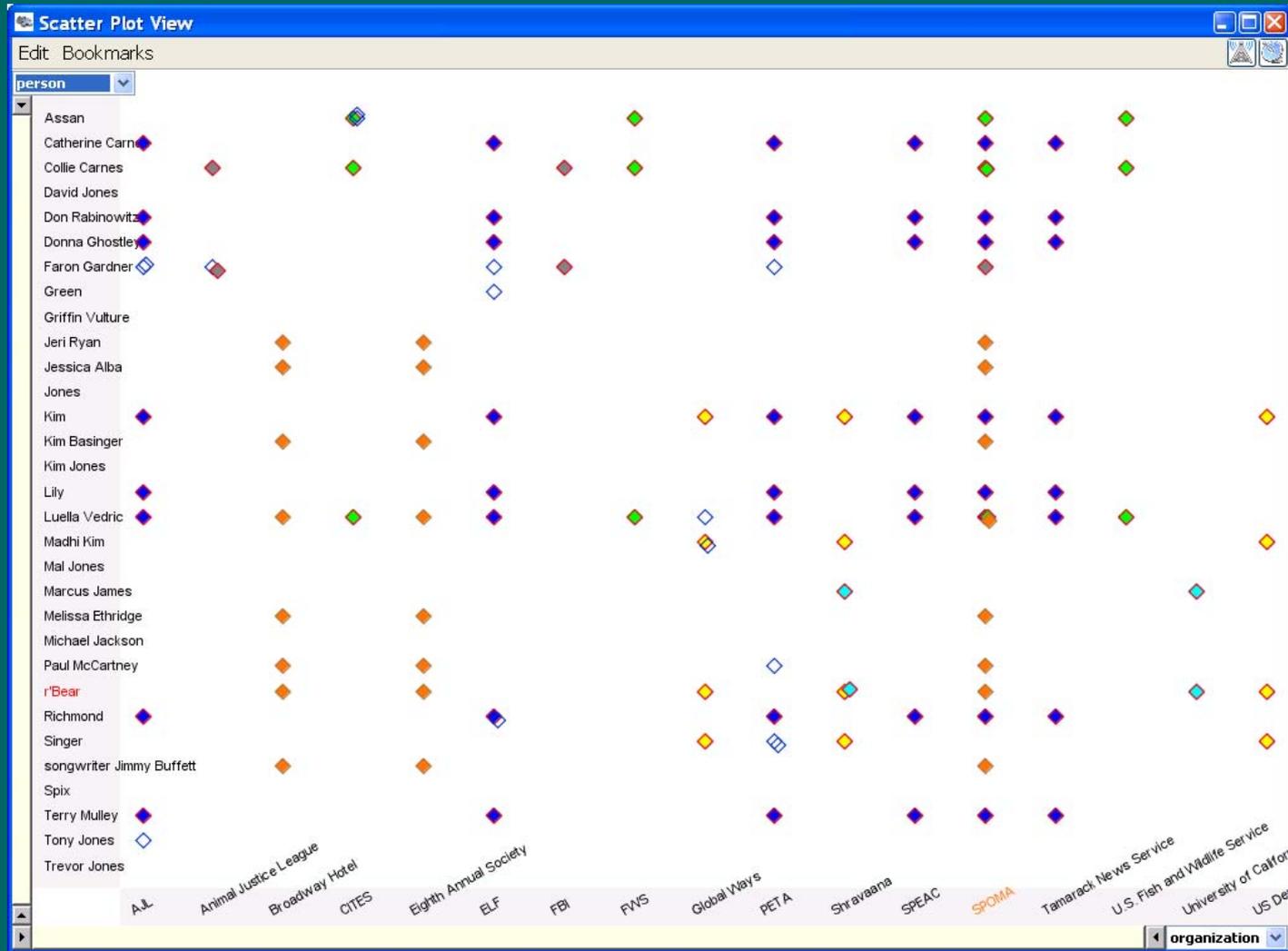
The screenshot shows a 'List View' application window with three main panels, each displaying a list of items. The panels are labeled 'place', 'person', and 'organization'. Each panel has a toolbar with options like 'Add all', 'Clear', and 'Show all connections'. The 'place' list includes items like 'Abbotsford', 'America', 'British Columbia', 'Canada', 'Chiapas', 'Colorado', 'Columbia', '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', and 'Puget Sound'. The 'person' list includes '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', and 'Wolf'. The 'organization' list includes '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', and 'WHO KILLED JANE'. Lines connect items across the lists, showing relationships. The 'Tre Arrow' person is highlighted in yellow.



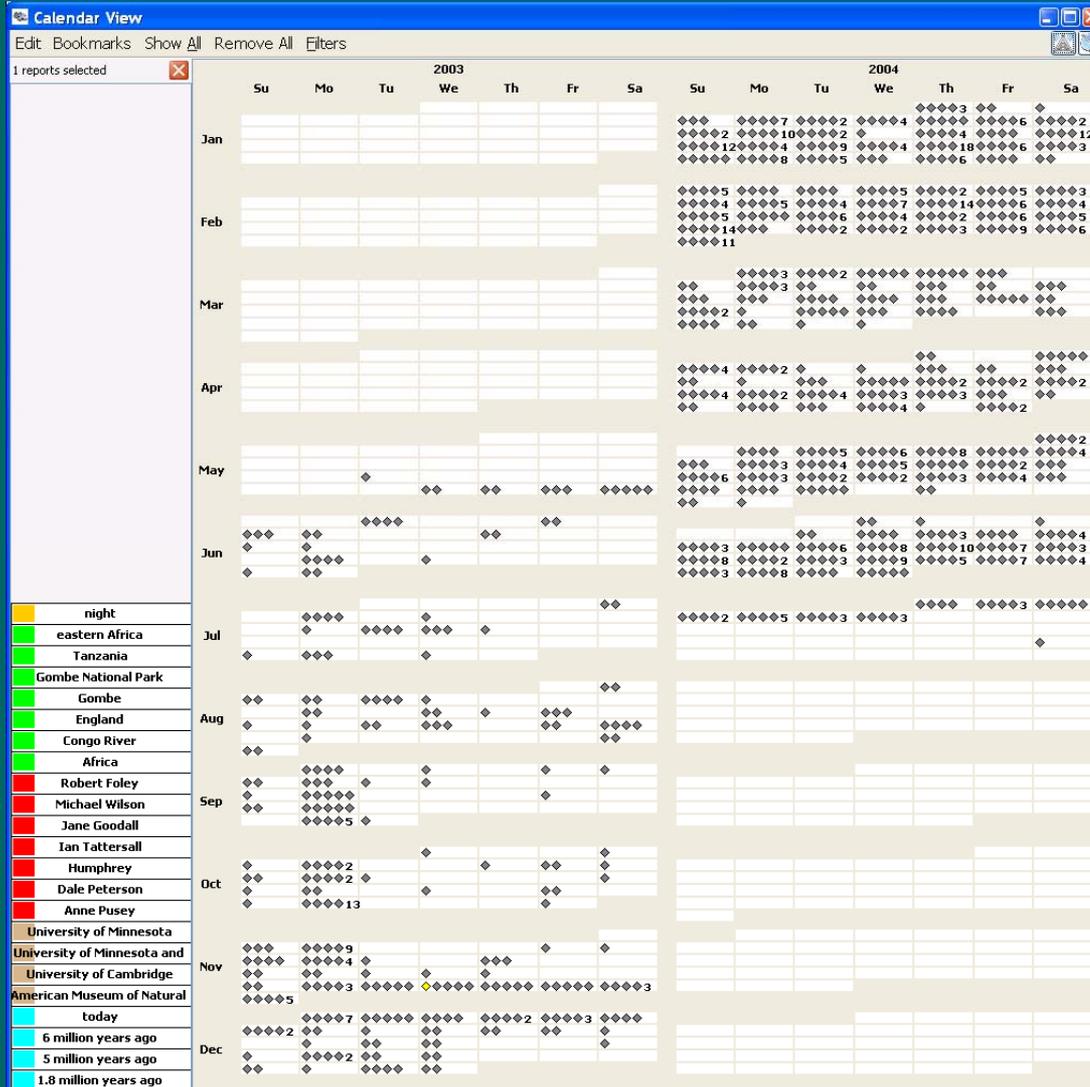
Graph View



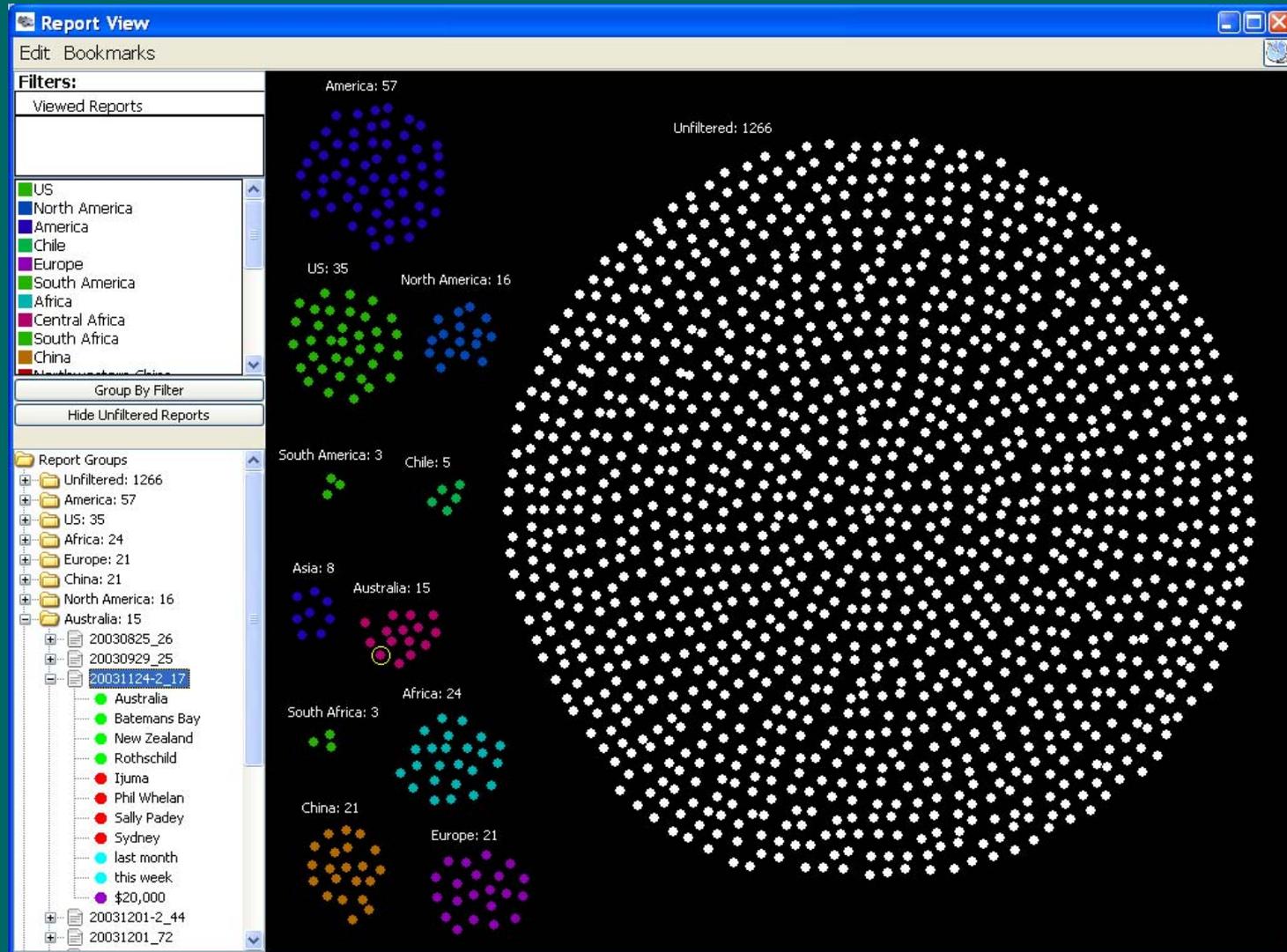
Scatterplot View



Calendar View



Report Cluster View



Shoebox

The screenshot shows the Shoebox application interface. On the left is a sidebar with a list of filters under the heading "UP" and "DOWN". The filters include names (Lily, Madhi Kim, South America, Florida Department, FWS, Global Ways, Sanchez, Aquatics, Arthur Swordane, Kim, Leslie, Luella Vedric, r'Bear, Singer, Alabama, America, California, New York, Texas, United States, Zimbabwe), dates (1989, 1990), and monetary values (\$15,000, \$25,000). The main workspace contains several floating windows:

- 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, 20031013_4, 20030526-2_57, www.jedidiaspokane.com, and 20040119-1_98.
- Global Ways**: A window listing CDC, Fish and Wildlife Service, US Department of Agriculture, Florida, South America, Ed Parker, and Madhi Kim.
- Chinchilla smuggling**: A window listing Gil Breeders, Los Angeles, Cesar Gil, Chile, Argentina, Andes, and Peru.

At the top right, there is a navigation bar with buttons for "Kim visited", "New York last", and "spring". A search bar contains the text "chinchilla Flocke ???". At the bottom right, there are additional filters for "Miami", "San Diego", and "Kenya".

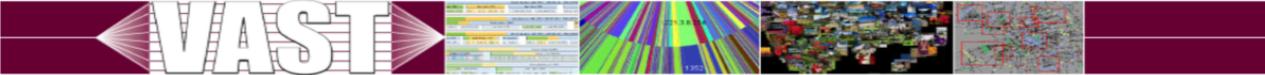


http://www.cs.umd.edu/hcil/VASTcontest07/ - Windows Internet Explorer

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

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%





Next Steps

- Looking for organizations to test the system
 - Hoping to cooperate
- Further research
 - Collaborative version
 - Reliability & uncertainty representation
 - Further text analysis



To Learn More

- <http://www.gvu.gatech.edu/ii/jigsaw>
 - Video available there



The screenshot shows a Windows Internet Explorer browser window displaying the website <http://www-static.cc.gatech.edu/gvu/ii/jigsaw/>. The page features the "information interfaces" logo at the top. The main content area is titled "Jigsaw: Visualization for Investigative Analysis" and includes a navigation menu on the left with links for About, People, Projects, Publications, Resources, Talks, and Videos. The main text describes the system's purpose, listing team members (John Stasko, Carsten Goerg, Zhicheng Liu, Gennady Stepanov, Vasili Panatazopoulos, Sarah Williams, and Alumni Neel Parekh, Kanupriya Singhal) and a list of publications. A central graphic shows a magnifying glass over puzzle pieces. A summary section explains that the system helps analysts connect small bits of evidence to uncover larger plans and narratives.

ACTIVE DOWNLOADS
Video - Introduction to system (20 meg wmv)

PAPERS
Jigsaw: Supporting Investigative Analysis through Interactive Visualization - VAST 2007
Jigsaw meets Blue Iguanodon - The VAST 2007 Contest (Invited Contest Paper) - VAST 2007
Visual Analytics with Jigsaw (Invited Poster Paper) - VAST 2007
Using Jigsaw, we won the university division of the VAST 2007 Contest

Summary
Investigative analysts acquire clues and connect small bits of evidence to uncover larger plans, stories, or narratives. Often the individual bits of evidence are short text documents, and analysts must examine large collections of such documents in order to "put the pieces together" and formulate a well-supported hypothesis about actions that may occur in the future. As the number of documents to examine rises, it becomes more and more challenging for analysts to uncover the embedded plans.

We are creating Jigsaw, a visual analytics system to help analysts better assess, analyze, and make sense of such document collections. Our specific objective is to help analysts reach more timely and accurate understandings of the plot(s) embedded throughout textual reports. Jigsaw provides a collection of visualizations that each portray different aspects of the documents. We particularly focus on presenting the identifiable important entities (people, places, organizations, etc.) and their direct or indirect connections. Textual processing extracts the important entities from the documents and then the visualizations help an analyst to explore the relationships and connections among the entities. The system includes graph, calendar, scatterplot and and tabular connections-based views, as well as views of individual document's text and the report collections as a whole. Jigsaw essentially acts as a visual index onto the document collection, helping analysts identify particular documents to read and examine next.

To learn more about the system's details and its different visualizations, please examine the Jigsaw views page.



Acknowledgments

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



- Supported by NSF IIS-0414667

