

Supporting Investigative Analysts using Visual Analytics

Carsten Görg

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



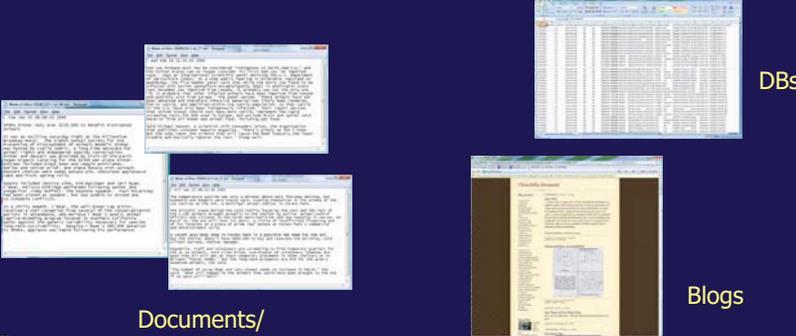
Investigative Analysis - Our Goals

- Assist analysts (police/intelligence, lawyers, reporters, scientists) confronting “information overload” problems while doing exploratory data analysis
- Assist analysts in information foraging and sense-making given a multitude of documents (reports, emails, blogs, abstracts, etc.)



Problem Addressed

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



The image displays four screenshots of software interfaces. On the left, three windows show document text, labeled "Documents/ case reports". On the right, a window shows a data table, labeled "DBs". Below the table, another window shows a web page, labeled "Blogs".

Documents/
case reports

DBs

Blogs



3

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.



4

Our Focus

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



5

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



6

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.



7

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



8

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”



9



System Views

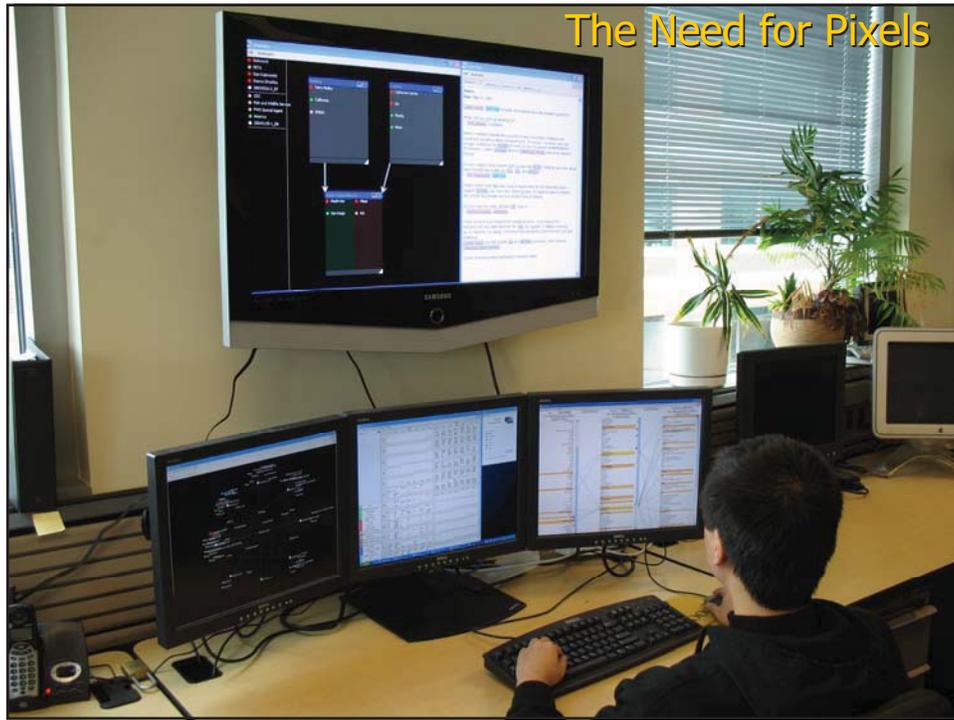
The image displays a collection of overlapping windows from a software application, illustrating different views of a system. The windows include:

- A text editor window showing a document with various colored highlights and a search bar.
- A window titled "Jigsaw" with a legend for "Color Legend" (Red, Green, Blue, Yellow, Purple) and "Shape Legend" (Circle, Square, Triangle, Diamond, Star).
- A window showing a network graph with nodes and edges, overlaid on a large circular visualization of many small nodes.
- A window showing a hierarchical tree structure of data.
- A window showing a detailed network graph with nodes and edges.
- A window showing a list of data entries with columns and rows.
- A window showing a detailed view of a single data entry.



10





Jigsaw Demo



12



VAST 2007 CONTEST
Blue Iguanodon

News
 We posted the BLUE I.G. and solution
 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 contests and FAQs](#).

Overview of the event
 The VAST Contest is a participation category of the [IEEE VAST 2007 Symposium](#) (part of [VIST 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/courier 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) displaying 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 ability 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 pre-processed data or the raw data.

Data set, tasks, and judging
 - [RF INFORMATION](#) about the data set, tasks and judging criteria. (File 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](#) (PDF) read the [RF INFORMATION](#) page above first

~1500 articles

Week-of-Mon-20040119-1.txt_98.xml - Notepad
 Tue Jan 20 08:08:10 2004
 SPOMA Dinner nets over \$230,000 to benefit mistreated animals
 It was an exciting Saturday night at the Millennium Broadway hotel. The Eighth Annual Society for the Prevention of Mistreatment of Animals benefit dinner was hosted by Luella Vedic, a long-time advocate for animal rights and endangered species conservation. Dinner and dessert was provided by Fruit-of-the-Earth vegan organic catering for the \$1300 per plate dinner. Entrees included black bean and veggie enchiladas, barley and seitan pilaf, and chana masala with spinach. Dessert choices were sweet potato pie, chocolate applesauce cake and fruit spring rolls.
 Guests included Jessica Alba, Kim Basinger and Jeri Ryan. r'Bear, Melissa Etheridge performed following author and songwriter Jimmy Buffett, the keynote speaker. Paul McCartney had been slated as speaker, but was unable to attend due to schedule conflicts.
 In a chilly moment, r'Bear, the well-known rap artist, received a cool reception from several of the conservationist purists in attendance, who believe r'Bear's exotic animal captive-breeding program located in southern California works against the genetic variability necessary for species long-term survivability. Despite r'Bear's \$80,000 donation to SPOMA, applause was tepid following his performance.

Week-of-Mon-20040216-3.txt_17.xml - Notepad
 wed Feb 18 21:24:40 2004
 Mad cow disease must now be considered "indigenous to North America," and the United States can no longer consider its first mad cow "an imported beast," says an international scientific panel advising the U.S. Department of Agriculture (USDA). At a USDA public hearing in Riverdale, Maryland on Wednesday, the five-member panel said that while the dairy cow found to be infected with bovine spongiform encephalopathy (BSE) in Washington state last December was reported from Canada, it probably was not the only one. "It is probable that other infected animals have been imported from Canada and possibly also from Europe," the panel warned. "These animals have not been detected and therefore infective material has likely been rendered fed to cattle, and amplified within the cattle population, so that cattle in the U.S. have also been indigenously infected," their report advises that United States should test many more cattle, implement the rapid screening tests for BSE used in Europe, and exclude brain and spinal cord material from all human and animal food, including pet food.
 said Michael Hanson, a scientist with Consumers Union, the organization that publishes Consumer Reports magazine. "There's plenty we don't know. But the USDA takes the science that will cause the beef industry the least trouble and basically ignores the rest." Sleep well.

Week-of-Mon-20040119-4.txt_33.xml - Notepad
 Fri Jan 23 08:52:58 2004
 The temperature outside was only a whisker above zero Thursday morning, but Kadoosie and Mupkin were toasty warm, sunning themselves in the window of the cat section at the Ark, a municipal animal shelter in Forest Park.
 The blissful scene belied the cold reality faced by the cats and the rest of the 1,100 animals brought annually to the shelter by police, animal-control officers and citizens in the seven municipalities and one township it serves, on March 31, the Ark will lock its doors, a victim of insufficient financing and of its location on a piece of prime real estate on Forest Park's commercial and entertainment strip.
 A vacant auto body shop in Forest Park is a possible new home for the Ark, but the shelter doesn't have \$800,000 to buy and renovate the building, said Elliott Serrano, shelter manager.
 Meanwhile, staff and volunteers are scrambling to find temporary quarters for 200 or so animals, said Ellen Milad, coordinator of volunteers. Chances are good that all will get at least temporary placement in other shelters or in private "foster homes," but the long-term prospects are dim for the area's unwanted animals, she said.
 "The number of stray dogs and cats always seems to increase in March," she said. "What will happen to the animals that would have been brought to the Ark if we were still here?"



17



Deployment and Evaluation

- King County Police Department in Seattle
- Formative evaluation to refine the views
- Now comparative studies



18



To Learn More

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



Thanks for your attention!



Acknowledgments

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



- Supported by NSF IIS-0414667

