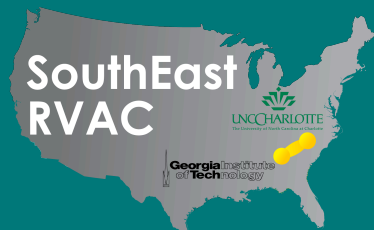# Visual Analytics for Investigative Analysis

John Stasko (PI), Carsten Görg, Zhicheng Liu

Information Interfaces Research Group

School of Interactive Computing

Georgia Institute of Technology

SouthEast
RVAC

UNCCHARLOTTE
The University of North Carolina at Charlotte

Georgia Institute
of Technology

Georgia
Tech

College of
Computing

School of Interactive Computing

Regrets

- John had committed to giving a keynote lecture in Germany this week

Background

- Through support of NVAC, we have been developing Jigsaw, a tool for investigative analysis on document collections

FODAVA Research Goals

- Explore how to integrate more automated, computational analysis with the human-centered document exploration of Jigsaw

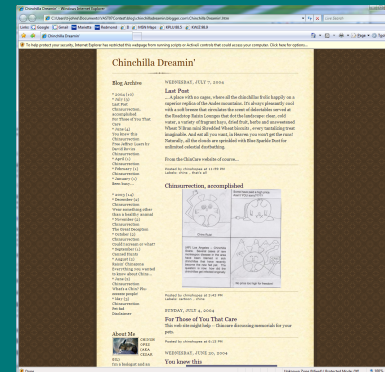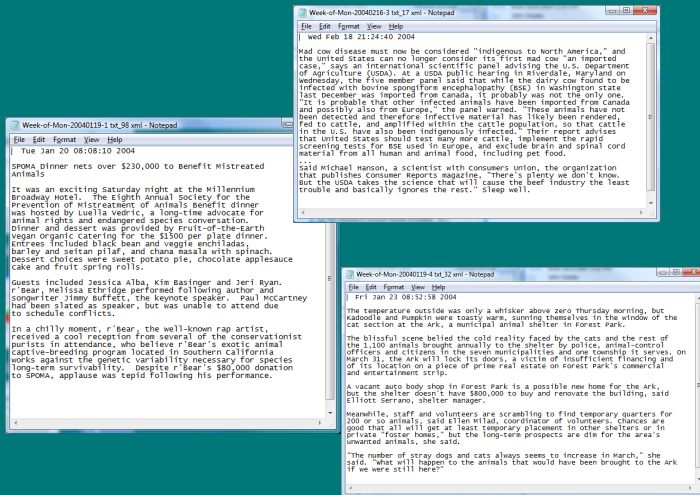- Handle larger and more varied forms of data

Step Back

- Quick overview of Jigsaw to provide the context for future research

Jigsaw Goals

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

Blogs

Documents/
case reports

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

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

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

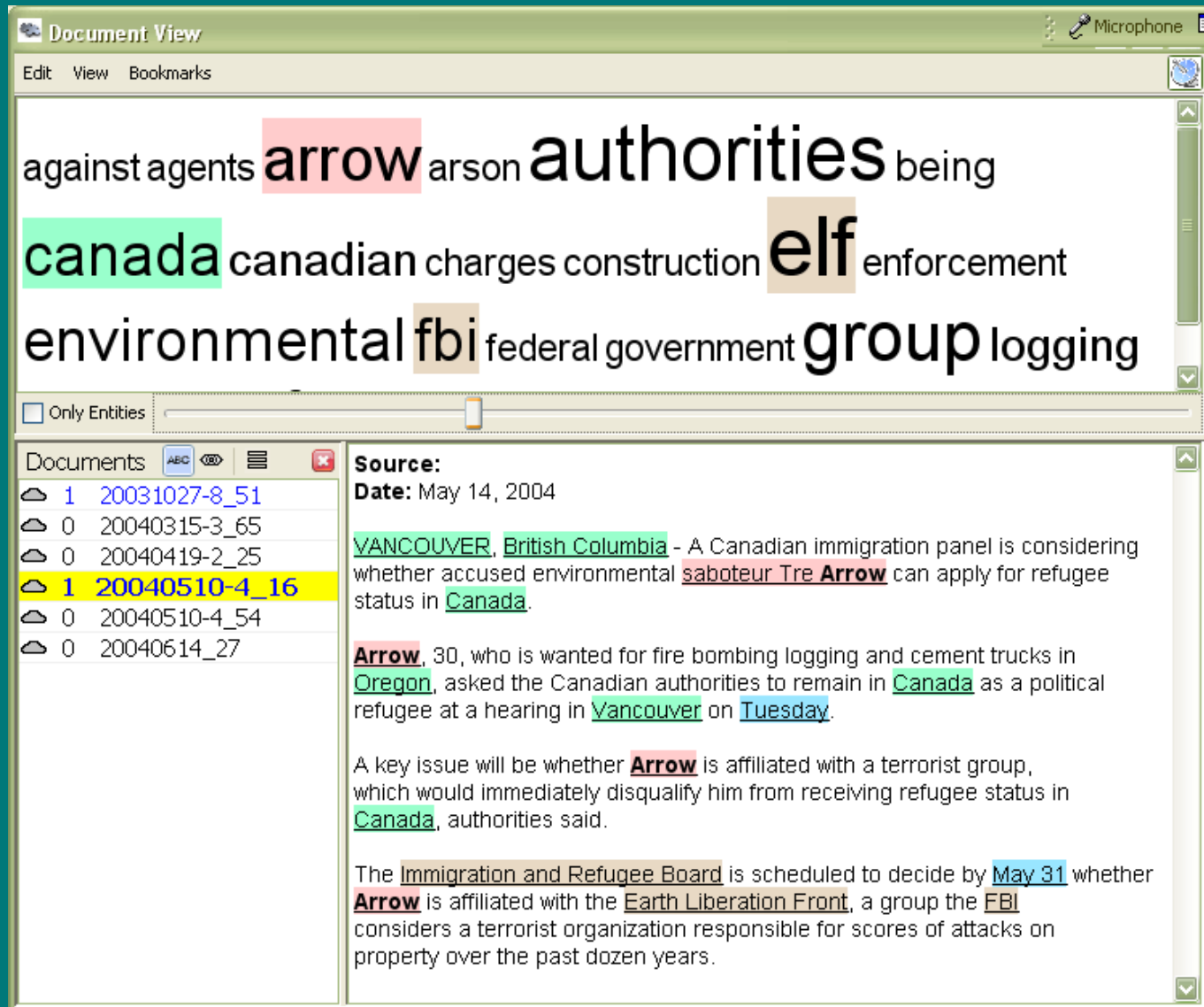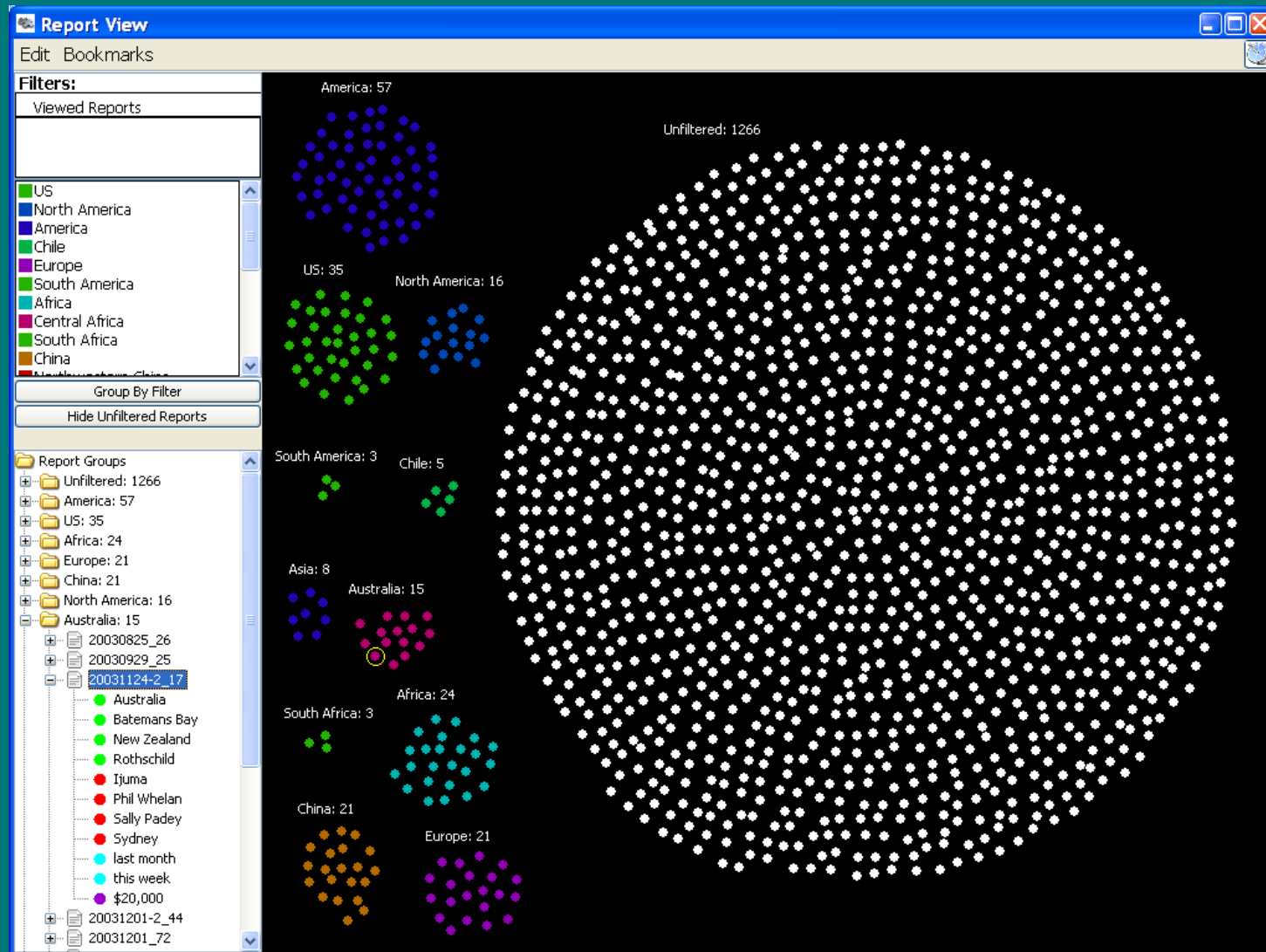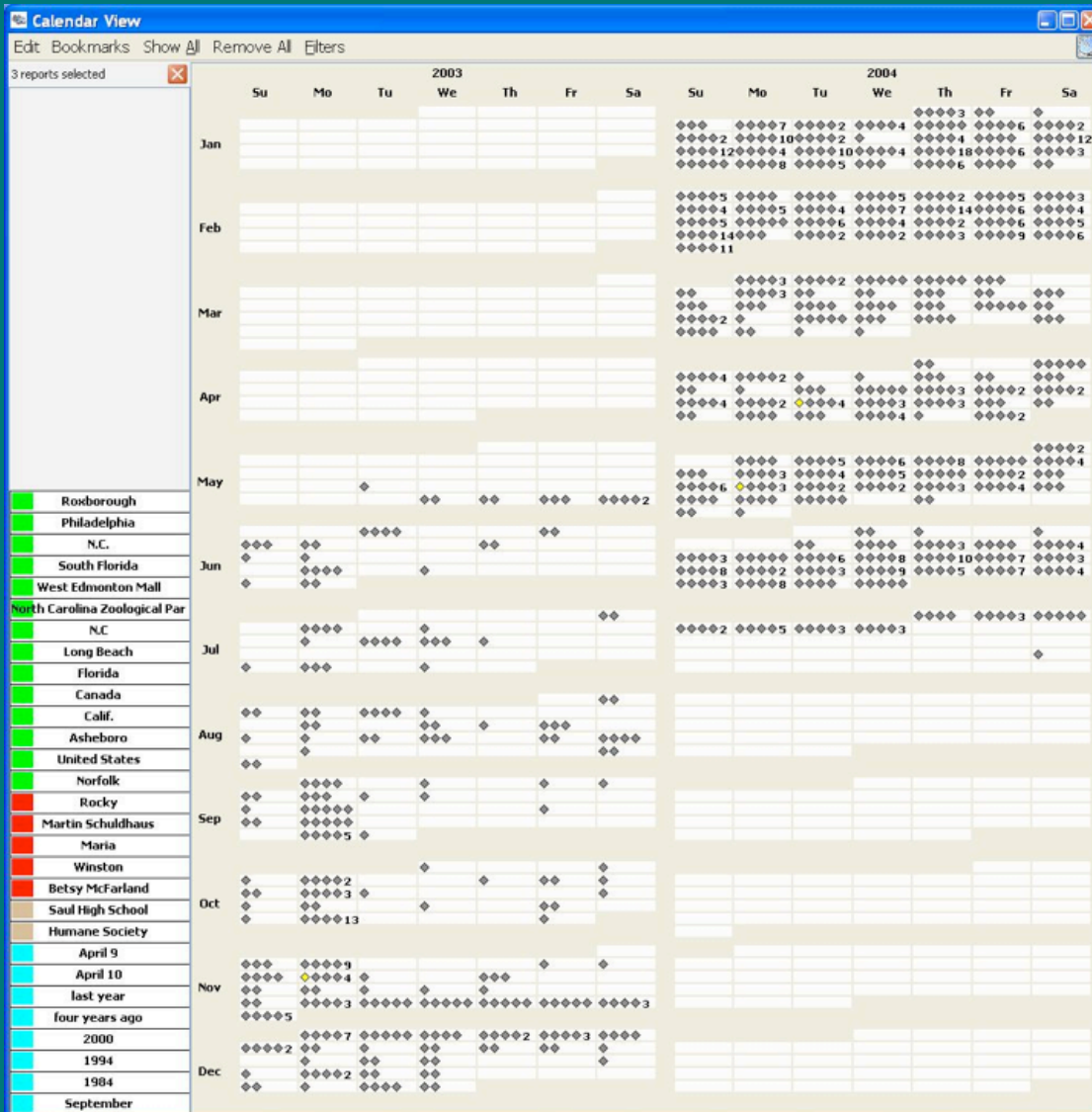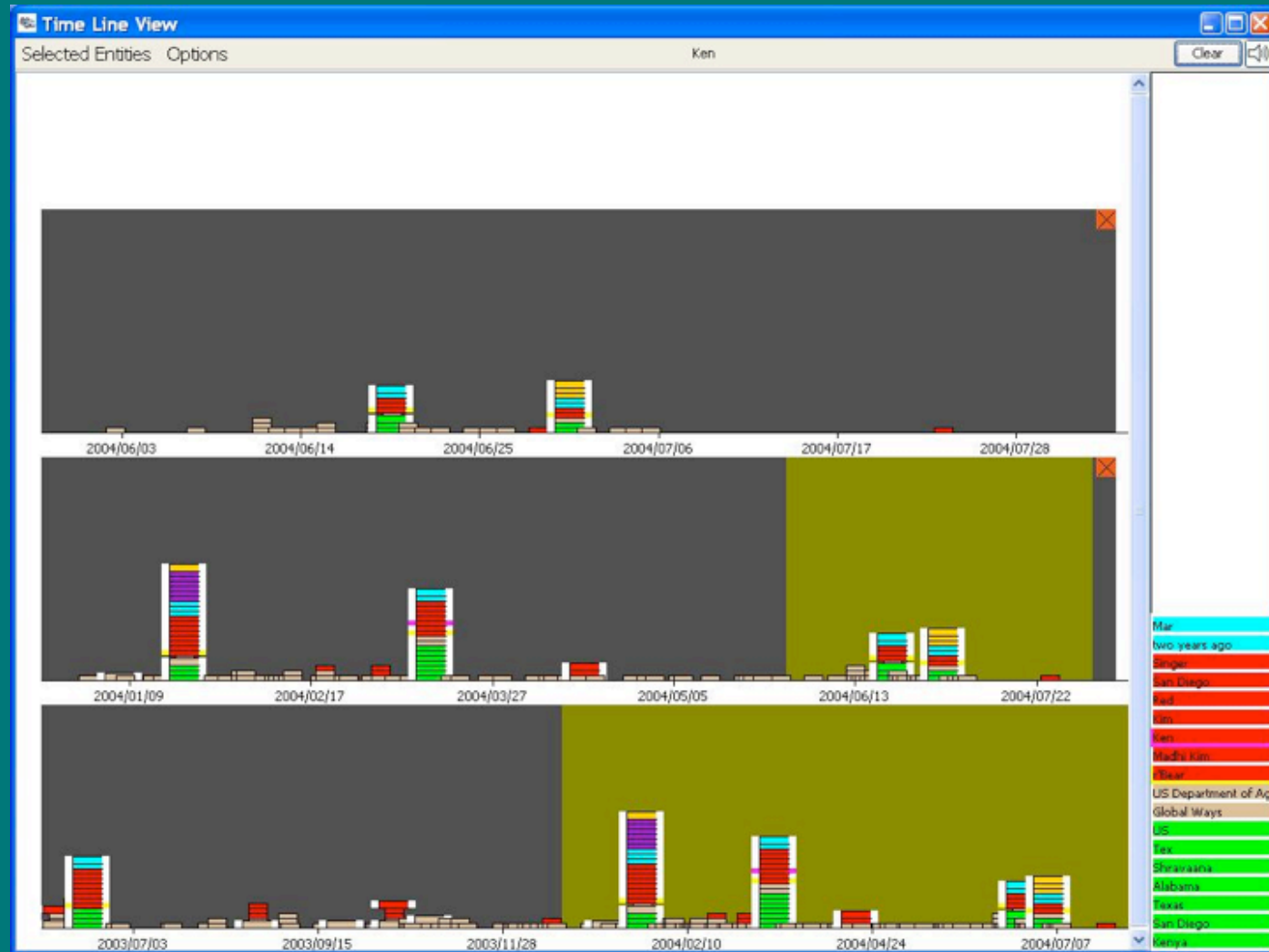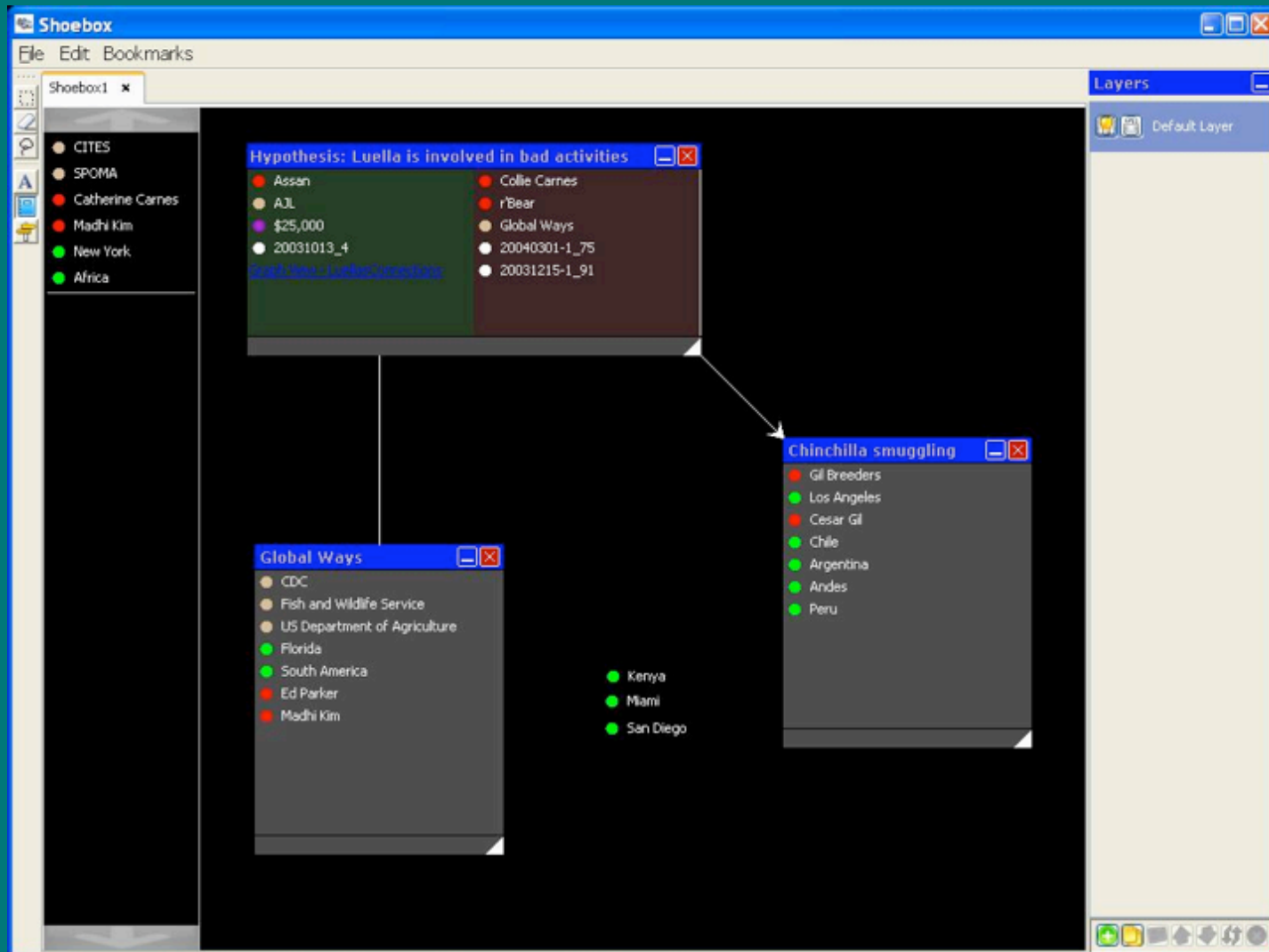- User actions generate events that are transmitted to and (possibly) reflected in other views

"Putting the pieces together"

System Views

The Need for Pixels

Video

Document View

Report Cluster View

Calendar View

Timeline View

Shoebox

FODAVA Goals

- Jigsaw can handle a few thousand documents, but how about a few hundred thousand?

- What about other types of data, eg, spreadsheets, audio, video, etc?

- Work with colleagues on new analysis techniques for larger data sets (text first focus)

Influence on FODAVA

- Investigative analysis on these types of data sets is widespread

  - Our success will provide techniques and tools for others to adopt and use

Developing FODAVA

- John will work "backwards"
  - Most FODAVA researchers have the data analysis foundations, but are new to the data domains and problems (and visual analytics)
  - John will bring the domain and VA knowledge, and address the need for better data analysis techniques

- Will act as a bridge to the RVAC/NVAC communities

To Learn More

- ## http://www.gvu.gatech.edu/ii/jigsaw
  - ### – Video available there

Acknowledgments

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