Jigsaw: Visualization for Investigative Analysis

Suppose that you are given a big box full of the pieces from many different jigsaw puzzles and you are asked to put the pieces together from one or two of the most "interesting" puzzles and describe what you see. Oh, by the way, not all the pieces of those "interesting" puzzles are in the box.

Investigative analysts, particularly those in fields such as law enforcement or intelligence, frequently confront this kind of challenge in their work. They are given large collections of seemingly unconnected documents and are tasked with identifying a plot or threat that is hinted at, but not clearly communicated, by a small subset of the documents in the collection.

Researchers at Georgia Tech and the SouthEast RVAC believe that clever visual analytic technologies can provide significant assistance to investigators faced with such challenges. They have developed a system called Jigsaw to do just that. Jigsaw provides visualizations of documents and the entities within them to help analysts explore the document contents and develop hypotheses about plots and stories suggested therein.

Investigative Focus

Systems such as PNNL's IN-SPIRE™ provide investigative analysts with rapid overviews of large document collections, highlighting themes across the collection and sets of documents that cluster together. Jigsaw, instead, focuses on the individual entities within the documents: the people, places, organizations, etc., that are the actors and events in narratives to be discovered. Jigsaw allows the analyst to search for a particular entity, and then the system visually communicates the context of that entity, such as the documents in which it appears and the other entities to which it is connected.

Jigsaw provides a suite of visualizations that depict different perspectives on the documents and their entities. Each visualization (called a "view" in Jigsaw) communicates a different aspect of the documents and how the different entities relate to each other.

The figure to the left shows the most basic view within the system, the Document View. This visualization presents a subset of documents from the collection (listed in the left-hand panel) and shows one selected document in the main window. Entities of varying types, indicated by the words with different background colors, are shown in the context of the document's source text. The tag cloud above highlights some of the key terms from subsets of the documents.

The first figure on the next page illustrates the List and Graph Views in Jigsaw. The List View presents lists of entities organized by entity type. Selected entities are drawn with a yellow background; entities logically connected to them are shown in orange, with darker shades indicating stronger connections. Entity lists can be sorted by frequency of occurrence across the document collection, connection strength, or alphabetical order on the entity names. Thus, the List View provides rapid exploration of the connections between sets of entities in the documents.

The Graph View provides a node-link representation of the documents and entities. Documents are white rectangles and entities are small colored circles connected by lines to the documents in which they appear. Jigsaw allows an analyst to display a particular document or entity and then interactively expand that item to show connected entities and documents, repeating the process to explore more of the "interesting" aspects of the collection. Jigsaw provides layout assistance when the view becomes cluttered so that the analyst can focus on the entities and documents of most interest to her/him.
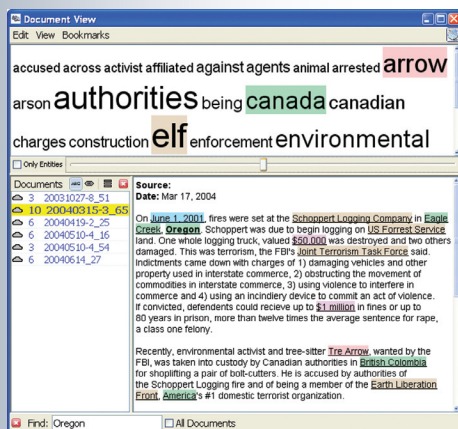
Document View showing a report with identified entities. The tag cloud summarizes key terms from a subset of documents.

This notion of interactive exploration is a key part of Jigsaw's design. The system's user interface provides a simple-to-learn model: a single mouse click selects an item and a double-click expands the item. Analysts can always query items through the system's search capability, but most investigation and analysis occurs by exploring the different representations in the system views.

Jigsaw includes a number of other views as well, including calendar-based temporal perspectives of the documents and entities and cluster-based collections of documents. Further, Jigsaw includes a Shoebox window to help analysts develop hypotheses, identify supporting evidence and organize their thoughts in general.

Because the system provides many different views of the document collection, researchers have found it fruitful to run Jigsaw on a multiple monitor computer that allows many views to be shown concurrently without overlap. The figure below shows a computer with four monitors in the Georgia Tech lab used to run the system.

Jigsaw in Action

Using Jigsaw as an analytic aid, Georgia Tech entered the 2007 IEEE Visual Analytics Science and Technology Symposium Contest and won the university division. The team is presently incorporating feedback from practicing analysts to refine Jigsaw's capabilities and views. Future planned research will explore collaborative versions of the system and support for reflection on the investigative process. Additionally, the Jigsaw team is actively seeking organizations that would like to explore the use of Jigsaw in their investigations.
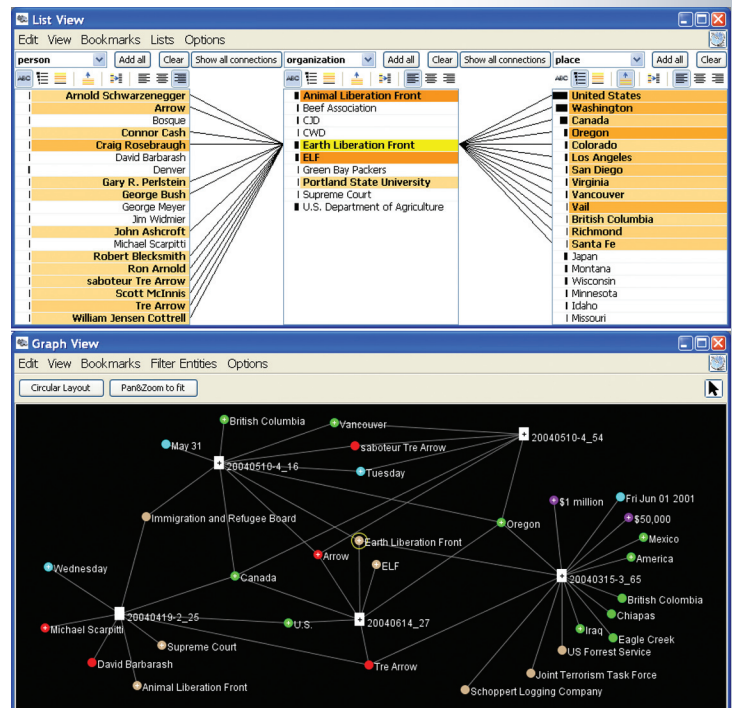
For more information about the Jigsaw project, see
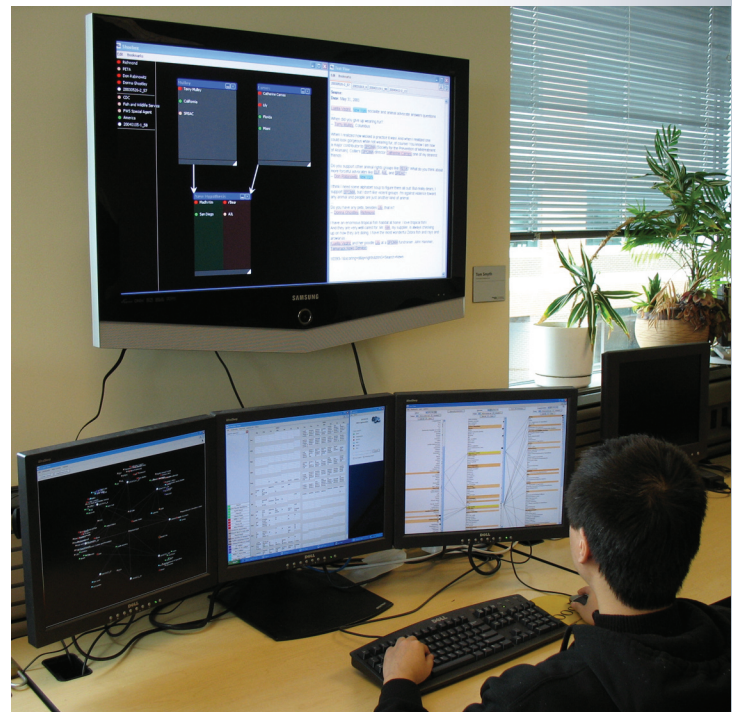http://www.gvu.gatech.edu/ii/jigsaw

Team Members

Georgia Tech: Meekal Bajaj, Carsten Görg, Zhicheng Liu, Vasilios Pantazopoulos, John Stasko, Gennadiy Stepanov, Sarah Williams

Point of Contact

John Stasko, Professor, Georgia Institute of Technology, stasko@cc.gatech.edu

List and Graph Views illustrating documents, entities and the connections.

Jigsaw running on a multiple-monitor computer.