**Jigsaw: a Visual Index on Large Document Collections**

Carsten Görg          John Stasko  
School of Interactive Computing & GVU Center  
Georgia Institute of Technology  
{goerg,stasko}@cc.gatech.edu

**Abstract**

Investigative analysts who work with collections of text documents connect embedded threads of evidence in order to formulate hypotheses about plans and activities of potential interest. As the number of documents and the corresponding number of concepts and entities within the documents grow larger, sense-making processes become more and more difficult for the analysts. We have developed a visual analytic system called *Jigsaw* that represents documents and their entities visually in order to help analysts examine reports more efficiently and develop theories about potential actions more quickly.

The *Jigsaw* system provides multiple coordinated views that show connections between entities (like people, places, organizations, dates, etc.) across documents. A connection between two entities is defined as a co-occurrence in at least one document.

To allow *Jigsaw* to handle large datasets, the system does not show the entire dataset at once but uses an incremental query-based approach to show a subset of the dataset. The query window allows analysts to search for entities and also provides a text search within the documents. *Jigsaw’s* query approach is different from traditional search engines. Getting a list of ranked documents as a result of a query would not be sufficient for analysts’ tasks because their activities go beyond just looking for a set of documents. Analysts also care about understanding what is inside of a document and how those entities are connected to entities in other documents. To support that task *Jigsaw* acts as a visual index on the document collection: the query results, consisting of entities and documents, are sent to multiple views that show different perspectives on the connections between those elements. The analysts can interact with the views, apply filters, or expand the context to gain more insight about the document collection. This exploration then spurs further queries and retrieves other documents and entities. While acting as a visual index, *Jigsaw* guides the analysts to related documents and facilitates the information retrieval process.

*Jigsaw* presents documents and entities resulting from queries through six different types of views. Therefore, the availability of significant screen space is very beneficial. The Text View shows document text, allowing analysts to validate connections, providing their context, and giving access to information that is not extracted as an entity. The List and Graph Views display connections between entities and allow analysts to explore the connection network. The Scatter Plot View highlights pairwise relationships between any
two entity types. The Time Line and Calendar Views organize entities and reports by date to ease the search for time patterns.

Figure 1 shows three different views after querying for “Faron Gardner” and exploring the query result. The Text View shows documents related to the query in three tabs. Entities within those documents are color coded accordingly to their type. The List View shows people and organizations connected to Gardner, with a darker shade of orange indicating a stronger connection. The Graph View displays the documents in which Gardner is mentioned, as well as the entities within these documents. Thus, it is easy to see which entities are mentioned in multiple documents.

Figure 1: The Text View, List View, and Graph View showing different perspectives after querying for “Faron Gardner” and exploring the query result.

Jigsaw’s views are coordinated using an event mechanism: interactions with one view (selecting, adding, removing, or expanding entities) are transformed into events that are then broadcast to all other views. Thus, the views of the system stay consistent and provide different perspectives on the same data.

For a detailed description of the system, we refer the reader to an article about Jigsaw in the VAST ’07 proceedings and to a video on the project website that shows interaction with the system.