

Visual Analytics with Jigsaw

Carsten Görg*

Zhicheng Liu†

Neel Parekh‡

Kanupriya Singhal§

John Stasko¶

School of Interactive Computing & GVU Center
Georgia Institute of Technology

ABSTRACT

This article briefly introduces the *Jigsaw* system and describes how we used it in analysis activities for the VAST '07 Contest. *Jigsaw* is a visual analytic system that provides multiple coordinated views to show connections between entities that are extracted from a collection of documents.

Keywords: Visual analytics, investigative analysis, intelligence analysis, information visualization, multiple views

Index Terms: H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces

1 INTRODUCTION

The *Jigsaw* system provides multiple views that show connections between entities across documents. We designed *Jigsaw* with usability and simplicity as key design features – its operations should be intuitive and easy to use. For a detailed description of the system, we refer the reader to a regular paper [3] about *Jigsaw* in the VAST '07 proceedings and to a video on the project website [2] that shows interaction with the system. Further details about the analytic process we used to work on the VAST '07 contest are described in [1].

2 JIGSAW SYSTEM

Jigsaw presents information about documents and their entities through six different types of views. Therefore, the availability of significant screen space is very beneficial – we used a computer with four monitors while exploring the contest data set.

Jigsaw's views do not show the entire dataset at once but use an incremental query-based approach to show a subset of the dataset. At start-up the views are first empty and the analyst can populate them either by querying for entities or later by expanding entities (showing other entities they are connected to) that are already visible in one of the views. This approach allows *Jigsaw* to handle larger datasets. The following views, which can be instantiated multiple times, are part of the system:

The **List View** draws links and uses colors to show connections between entities. It allows users to specify the number of lists that are displayed, the type of entity shown in each list, and it offers multiple sorting options.

The **Graph View** displays connections between entities and reports in a node-link diagram. It provides the possibility to explore the network step-by-step by expanding and collapsing nodes to show or hide their connected entities or reports, respectively.

*e-mail: goerg@cc.gatech.edu

†e-mail: zcliu@cc.gatech.edu

‡e-mail: justneel@gmail.com

§e-mail: ksinghal@cc.gatech.edu

¶e-mail: stasko@cc.gatech.edu

The **Scatter Plot View** highlights pairwise relationships between any two entity types. It supports focusing on a specific subset of entities by range sliders.

The **Text View** displays the actual reports, highlights the entities within them, and shows how often a report already has been displayed.

The **Time Line View** shows entities that are connected to a date on a time line. It supports zooming in on a specific time range.

The **Calendar View** provides an overview of the reports and the entities within them according to the publication date of the reports.

The 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. *Jigsaw* also has an option to turn event listening off in case a specific view shows an interesting data subset that should not be changed.

Figure 1 shows four different views that display data from the VAST '07 contest dataset. The List View shows the organizations to which *Luella Vetric* is connected (the lines from the person to the organization column) and also with whom else she has connections (the orange colored entities in the person column). The Text View shows the four reports in which *Vetric* is mentioned. The Graph View displays *Vetric*'s connection network: the reports she is mentioned in as well as the entities within those reports. Finally, the Calendar View gives an overview of reports published in June, July, August, and September 2003 and shows the entities of one selected report on the side.

3 ANALYTIC PROCESS WITH JIGSAW

We used *Jigsaw* to explore the VAST '07 contest data set and identify an underlying plot. Since *Jigsaw* does not present themes within a report collection, it did not remove the need for us to read the reports. Instead, *Jigsaw* acted as a visual index onto the report collection and helped us find the most important reports to read. While working on the contest dataset we realized that repeatedly reading reports is important to incrementally expand knowledge about the dataset and to understand details in the underlying plot.

The possibility of having multiple instances of one view was one feature of *Jigsaw* that helped us during the analysis. We could follow different threads in parallel during our investigation and when two threads seemed to merge or at least seemed to be related, we could create a new view containing the entities of both threads and investigate further with this combined and extended set of entities.

Having multiple instances of one view was also useful in another context: we used multiple instances of the Text View to group reports (e.g. we had multiple Text Views open where each showed the reports about a key person in the plot). A missing feature in this approach was to be able to easily see the reports that are present in more than one of the Text Views.

The Graph View helped in finding highly connected entities. This view provides a layout algorithm that arranges the reports on

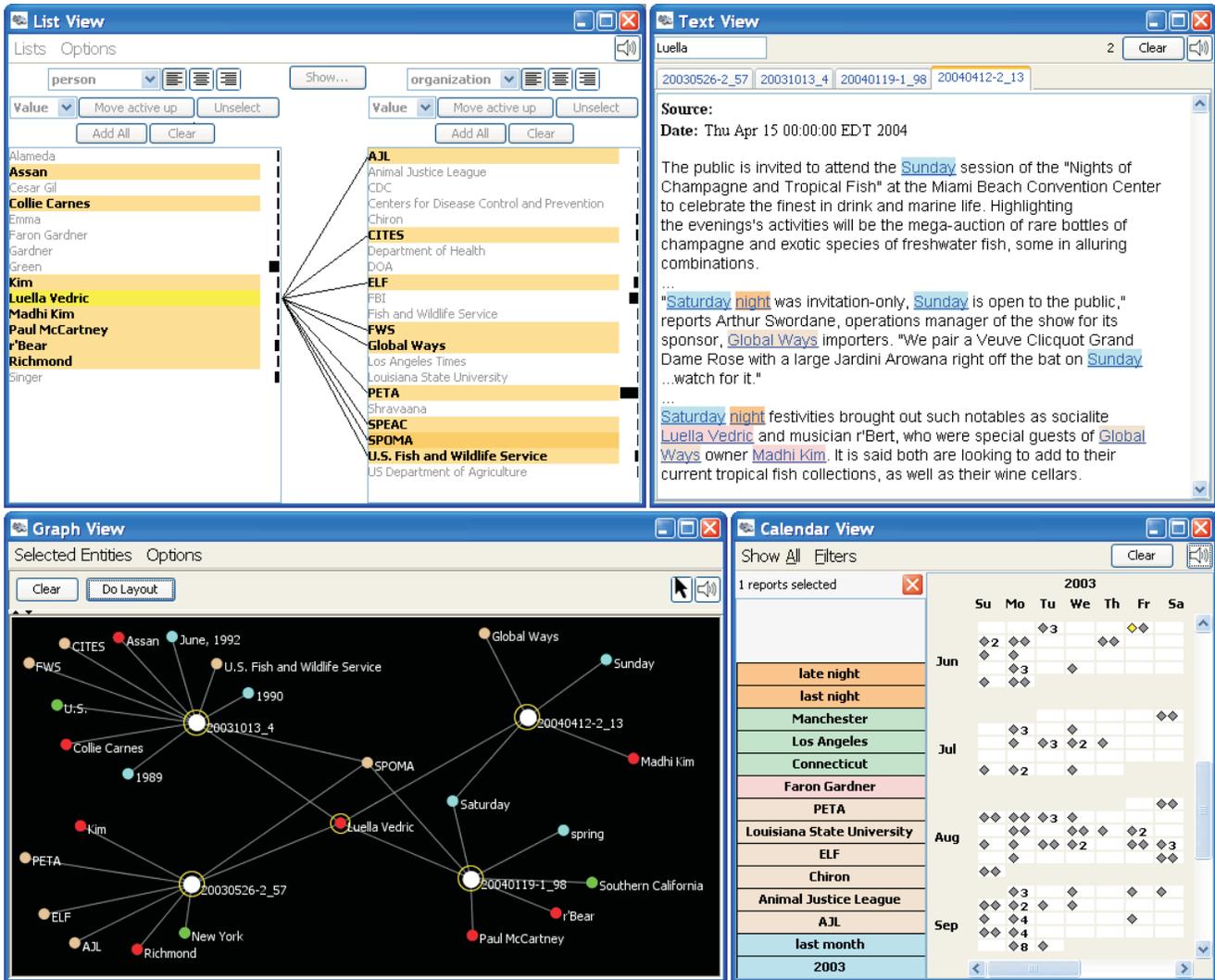


Figure 1: The List View, Text View, Graph View, and Calendar View showing data from the VAST '07 dataset.

a circle, puts all entities that are connected to only one report in the outer area of the circle, and all entities that are connected to multiple reports in the inner area of the circle. This representation allows analysts to quickly find and focus on entities of a subset of the document collection that are highly connected.

Another useful approach we applied was to increase the context of an entity collection step-by-step. We began by putting a small entity collection in the Graph View, expanded all these entities to get the reports in which they occur, and then expanded all those reports to get their entities. Next, we explored this new entity and report set in more detail in one of the other views and by reading the reports in order to decide which of the new entities would be relevant. We removed the ones that were not of interest and then we continued to expand the context again.

One drawback of Jigsaw found during our investigation was the missing functionality of being able to change the extracted entities on the fly. Since the system uses the co-occurrence of entities to build the connection network, it is crucial that the entities are properly identified. Missing or unidentified entities result in a knowledge gap: connections that are not there cannot be visualized. To overcome this shortcoming we will extend Jigsaw's function-

ality to allow the user to add, remove, or change extracted entities dynamically.

ACKNOWLEDGEMENTS

This research is supported in part by the National Science Foundation via Award IIS-0414667 and the National Visualization and Analytics Center (NVAC™), a U.S. Department of Homeland Security Program, under the auspices of the Southeast Regional Visualization and Analytics Center. Carsten Görg was supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD).

REFERENCES

- [1] C. Görg, Z. Liu, N. Parekh, K. Singhal, and J. Stasko. Jigsaw meets Blue Iguanodon - The VAST 2007 Contest. In *IEEE Symposium on Visual Analytics Science and Technology*, 2007.
- [2] Jigsaw project. <http://www.gvu.gatech.edu/ii/jigsaw/>.
- [3] J. Stasko, C. Görg, Z. Liu, and K. Singhal. Supporting Investigative Analysis through Interactive Visualization. In *IEEE Symposium on Visual Analytics Science and Technology*, 2007.