VAST 2006 Contest - First Place, Corporate Category

Applied Analysis Using nSpace and GeoTime

Pascale Proulx, Lynn Chien, Adam Bodnar, Kaleb Ruch and William Wright*

Oculus Info Inc.

ABSTRACT

GeoTime and nSpace are new innovative visualization tools that enable the analysis of large, multi-dimensional datasets. The analytic capabilities supported by these tools were used by improvised analysts to investigate the 2006 VAST contest. This paper describes how these tools were used to create an analytical environment that enabled novice analysts to examine the scenario, assess evidence, and visually represent meaningful hypotheses.

CR Categories: H.5.2 [Information Interfaces & Presentations]: User Interfaces – Graphical User Interfaces (GUI); I.3.6 [Methodology and Techniques]: Interaction Techniques.

Keywords: visual analytics, human information interaction, sense making, geo-spatial information systems, temporal analysis

1 OVERVIEW OF THE ANALYTIC TOOLS

GeoTime and nSpace are two novel visual analytic applications that have been developed in collaboration with analysts to support the investigation of large and complex datasets. GeoTime supports the visualization and analysis of entities and events over time and geography [3] and is currently in transition to deployment for analysts to use on a day-to-day basis. nSpace is used for triaging massive data and for analytical sense-making [2], [4], and is currently undergoing experimental evaluation and pilot deployment.

nSpace combines the multi-dimensional linked views found in TRIST (The Rapid Information Scanning Tool) with the visible, cognitive mechanisms of the Sandbox. TRIST provides query planning, rapid scanning over thousands of search results in one display, and includes multiple linked dimensions for result characterization and correlation. Analysts work with TRIST to triage their massive data and to extract information into the Sandbox evidence marshalling environment. The Sandbox is a flexible and expressive thinking environment that supports both ad-hoc and more formal analytical sense making. Key capabilities for the Sandbox include "put-this-there" cognition, automatic layout of evidence with analytical templates, gestures for the fluid expression of thought, assertions with supporting/refuting evidence gates and scalability mechanisms to support larger analysis tasks. Additionally, nSpace integrates advanced computational linguistic functions [1] using a web services interface and protocol.

GeoTime improves perception of entity movements, events, relationships, and interactions over time within a geospatial (or any conceptual) context. Events are represented within an X,Y,T coordinate space, in which the X,Y plane shows geographic space and the T axis represents time. Events are arrayed in time along time tracks, which are located wherever events occur within the spatial plane. For analysts, GeoTime's single view representation of a combined temporal-spatial three-dimensional space amplifies concurrent cognition of entity relationships and behaviours in both space and time [3].

2 ANALYSIS PROCESS WITH VAST DATASET

2.1 Setting up the Stage for Scanning

The analysts began their investigation by importing the entire corpus of 1200 news stories into TRIST, and exploring it using the multi-dimensional linked views supported by the tool. Custom dimensions (key issues, players, and locations) were created as the analysts learned about the particular of the task.

2.2 Scanning with Multi-Dimensional Linked Views

TRIST supports the exploration of data from any particular line of thought. For example, starting in the *Key Issues* dimension, the analyst selected all results in the *laboratory* bin, automatically highlighting all people, places and organizations mentioned in the corpus related to the laboratory. The analyst then switched their attention to the *Date Published* dimension, which revealed the time range of the articles mentioning the laboratory, and proceeded to scan the lab results chronologically. With the relevant entities seen in context, they can be added to the *Key Players* dimension using intuitive drag and drop gestures. This process was repeated for any interesting topic found.

2.3 Saving and Organizing in the Sandbox.

Throughout the triage process, relevant types of information were discovered, including events, key people, relationships, key sources, etc. The analyst saves these relevant fragments of information by dragging them into the Sandbox. A short 'to-the-point' fragment is enough since nSpace inherently supports linking of fragments to full sources. In the Sandbox, the analyst quickly organized relevant content using 'put-this-there' interaction. For example, the analyst grouped events about the laboratory in chronological order, while another group was created to classify the voting behavior of council members. From these classifications, observations regarding alliances, voting frequency, and decision making were derived.

* email: {pascale.proulx, lynn.chien, adam.bodnar, kaleb.ruch, bill.wright}@oculusinfo.com

2.4 Creating Social Network and Detailed Profiles

The analyst continued their investigation by exploring relationships between the key players. Keeping track of relationships was accomplished by manually drawing links between them, each with a label and evidence for the relationship. Each key player within the social network was expanded to embed a detailed profile based on information discovered by scanning with the *Key Players* dimension. During this process, the analyst developed a profile structure that included groups for professional background, personal background, positive and negative relationships, and their personal and professional dispositions. The analyst then decided to save this structure as a template, which could then be reused and shared. This template was then applied to all the other key players, which automatically organized the existing fragments according to that same structure.

Figure 1. Key Players Network and Profiles in the Sandbox.

2.5 Capturing Questions, Thoughts and Hypotheses.

As various plot elements became more tangible, the visual representation of the Sandbox began to match the mental model of the analyst. The Sandbox supported the analyst in the development and assessment of meaningful hypotheses, which were captured as assertions. Evidence was associated with an assertion using simple drag and drop gestures through the refuting or supporting gate of the assertion. The analyst generated several competing hypotheses, to determine patterns of behaviour within the scenario. These were generated using both a bottom up (interpretations of the evidence) and top down (assumptions about what is implied by the top level assertion) fashion.

2.6 Importing Data into GeoTime

Entities and events identified as potentially relevant in nSpace were imported in GeoTime using a combination of automated import using an Excel-GeoTime plug-in, as well as manually entering data using the appropriate built in wizards. Each event was linked to the related data source, such as a news clipping, for access to the full context from within GeoTime.

2.7 Adding Visual Structure with Custom Layout

Once in GeoTime, the analyst added structure and context by classifying related entities into groups. Places, which in this case were not constrained by specific lat/long coordinates, were simply moved around with drag and drop gestures. Various shapes were created using the *Ink* tool to help visualize the structure of these groups, and text was added to supplement this. Colour was also used to convey the intended relationship of the groups.

2.8 Isolating Patterns and Capturing Hypotheses

GeoTime visualization makes patterns of events in time and space easier to recognize. Various built-in tools in GeoTime also facilitate the detection and recognition of interesting events. For example, in an attempt to root out the suspicious phone log findings, the GeoTime *find* tool was used to execute searches for "Switzerland". Results turned in a host of events, targets and places all with links to Switzerland. These were dragged to a *Switzerland layer*. Layers are containers used to categorize and organize interesting observations and evidence for quick future referrals. At any point, the insights provided by the visualization can be captured in annotated graphical reports. These reports can then be integrated with the other arguments in the Sandbox and subsequently prompt the analyst to go back to TRIST to look for additional supporting or refuting evidence for these hypotheses.

3 Strengths

Using nSpace, the analyst was able to freely brainstorm, identify important entities, find evidence, generate and refine hypotheses, marshal evidence, and report findings. Combined with intuitive interaction, such as "put-this-there", nSpace supported both ad-hoc and more formal analytical sense making techniques.

GeoTime provided a geo-temporal environment that supported the analyst in understanding entities and events, such as communications, money flows, positions in organizations, relationships and physical movements of key people. GeoTime's one-data-view system eliminated the typical analyst's dependence on multiple spreadsheets, tables, maps and other simultaneous cross-referenced views that can cause cognitive confusion.

4 Limitations

Currently, nSpace does not scale well to over 40,000 entities. This means that some pre-processing of the data was necessary before using nSpace to explore the voter registry. Additionally, nSpace does not currently support a specialized timeline view, requiring the analyst to manually lay out events chronologically.

While GeoTime inherently supports temporal analysis, it is currently laborious to import events into GeoTime. One of our current research efforts aims at supporting fluid creation of GeoTime scenarios within GeoTime, eliminating the need for external plug-ins or laborious wizards.

5 Conclusion

The VAST analysis contest demonstrated that nSpace and GeoTime can effectively enable novice analysts to take on real-world data analysis problems and work them out using their full reasoning and imaginative capabilities, not bogged down in rigid workflows or difficult to use tools.

References

[1] Freitag, D., M. Blume, J. Byrnes, R. Calmbach and R. Rohwer, "A Workbench for Rapid Development of Extraction Capabilities", In Proc. International Conference on Intelligence Analysis, 2005.

[2] Jonker, D., W. Wright, D. Schroh, P. Proulx and B. Cort, "Information Triage with TRIST", In Proc. International Conference on Intelligence Analysis, 2005.

[3] Kapler, T., and W. Wright, "GeoTime Information Visualization", IEEE Symposium on Information Visualization, 2004.

[4] Wright, W., D. Schroh, P. Proulx, A. Skaburskis and B. Cort, "The Sandbox for Analysis – Concepts and Methods", ACM CHI 2006.