Interactive Data Analysis with nSpace2®

VAST 2011 Mini Challenge #3 Award: "Good Analysis & Support Debrief"

Casey M. Canfield and David Sheffield*

Oculus Info Inc.

ABSTRACT

nSpace2 is an innovative visual analytics tool that was the primary platform used to search, evaluate, and organize the data in the VAST 2011 Mini Challenge #3 dataset. nSpace2 is a web-based tool that is designed to facilitate the back-and-forth-flow of the multiple steps of an analysis workflow, including search, data triage, organization, sense-making, and reporting. This paper describes how nSpace2 was used to assist every step of the analysis process for this VAST challenge.

KEYWORDS: visual analytics, human information interaction, analysis workflow, sense-making.

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

1 INTRODUCTION

nSpace2 is an open-source and multi-source analysis application designed to support multiple analytical tasks, styles, and workflows within a web browser-based system. By combining human information interaction, computational services, and innovative visualization techniques, nSpace2 enhances the analyst's ability to retrieve, comprehend, and organize large amounts of data. The nSpace2 environment is a system-of-systems that includes two main components: TRIST and the Sandbox.

1.1 nSpace2 TRIST®

The Rapid Information Scanning Tool, or TRIST, uses advanced information retrieval and characterization methods to enable the analyst to work with more information without increasing cognitive load. TRIST provides query planning, rapid scanning of results, and multiple linked dimensions for result characterization and correlation. Analysts use TRIST to triage massive amounts of data and extract information into the Sandbox marshalling environment [2].

1.2 nSpace2 Sandbox®

The Sandbox is the evidence-marshalling and sense-making component of nSpace2. It is a flexible and expressive thinking environment that supports both ad-hoc and formal analytical sense making tasks. The Sandbox provides a visual alternative to paper or text editors for analysis activities, and includes tools for hypothesis creation and analysis, analysis of competing hypotheses (ACH) [1], corroboration, grouping data, annotating evidence, and prioritizing information [3].

2 ANALYSIS PROCESS WITH MINI-CHALLENGE #3 DATASET

The analytical process was iterative. It began with broad concepts gathered in a brainstorming session in the Sandbox, and then focused on specifics as data gathering, triage, and sense-making progressed.

2.1 Sandbox: Brainstorming and Template Creation

To begin, the analyst created a Sandbox to gather ideas and develop a consistent framework for analysis. Since the challenge focused on discovering potential terrorist threats, the analyst developed a Sandbox structure that contained ideas for terror methods, motives, targets, and suspects. The analyst also created a timeline structure to contain gathered data in chronological order. With this infrastructure in place, the analyst started to gather data with TRIST.
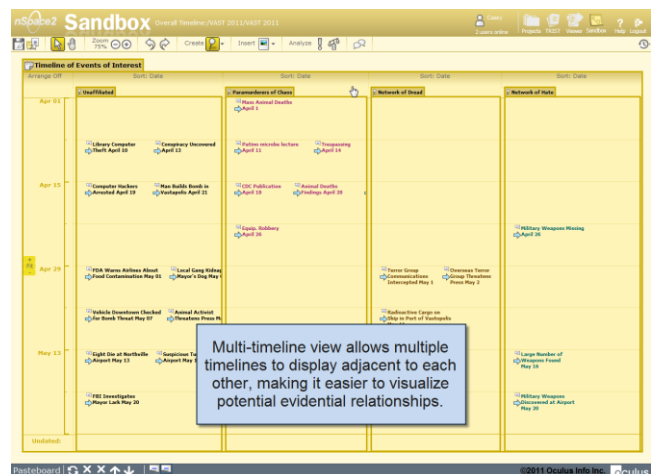

Figure 1. Timeline structure populated with evidence.

2.2 Information Triage in TRIST

2.2.1 Preparing for Scanning

The analyst imported the corpus of news stories into TRIST, and began queries against the corpus using the terrorism-related terms gathered in the brainstorming Sandbox. The initial query was a broad term designed to capture as many relevant results as possible in a single pass. This allowed for more effective result characterization and helped develop query terms for follow-up searches. The analyst then created custom dimensions to contain terrorism-related terms of interest.

* e-mail: {ccanfield, dsheffield}@oculusinfo.com

2.2.2 Scanning with Multi-Dimensional Linked Views

TRIST supports the exploration of data in accordance with the analyst's line of thought. User specific dimensions can be easily created and used. For example, in the Terrorism Terminology dimension, the analyst selected the "bomb" keyword, automatically highlighting all results and all other keywords contained in those results. By selecting "terror" and "group" keywords, the analyst highlighted the results containing both of those terms. This allowed the analyst to quickly find documents with the most relevance to the task.
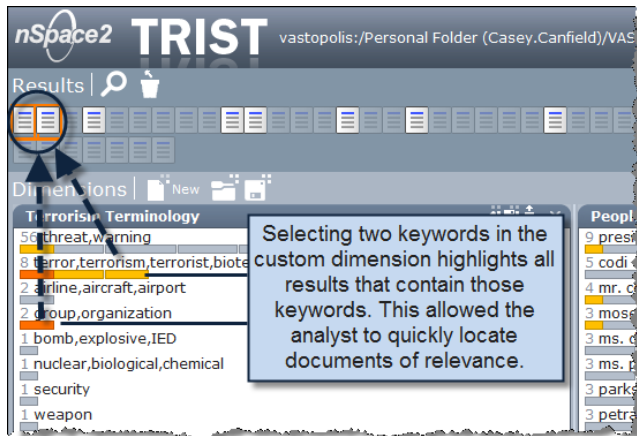

Figure 2. TRIST provided rapid location of relevant articles.

2.2.3 Targeted Queries and Iterative Data Triage

Based on the results of the initial query, the analyst made subsequent targeted queries to delve into topics of interest. For example, the analyst discovered a document about terrorism expert "Jose Thom" in the first triage. That document mentioned three other extremist groups by name, so the analyst queried the corpus for those names to retrieve additional results about those groups. The analyst also added the group names to a custom dimension so that results containing those names were highlighted when each name was selected.

2.3 Evidence Marshalling in the Sandbox

2.3.1 Saving and Organizing Data in the Sandbox

After triage of the initial query results, the analyst discovered relevant types of information, including events, key people, terrorist groups, etc. The analyst saved these pieces of information by using the nSpace2 Pasteboard to move them to the Sandbox. nSpace2 preserved the provenance of the data source as metadata for each information piece.

The analyst moved data to the Sandbox in the form of snippets. Small snippets of data were enough since the Sandbox supports linking of fragments to full sources. nSpace2 does this automatically when fragments are captured via the Pasteboard. The analyst grouped data in the Sandbox according to the nature of the information. All events deemed relevant were placed on a master timeline.

2.3.2 Creating Group Profiles to Evaluate Threats

Over time, the analyst completed "dossier" Sandboxes for groups that posed an apparent threat. Each profile contained both a timeline of events related to that group and a structure to facilitate the analyst's evaluation of the gathered evidence. The structure provided potential motives, typical methods, and potential targets of the group at a glance. The use of a common template for these profiles made the analysis significantly more efficient. Each time that the analysis cycle progressed from query to triage to Sandbox, the various Sandboxes evolved as a model of the analyst's thought processes.

2.3.3 Managing Questions, Thoughts, and Hypotheses

The Sandbox allowed the analyst to develop and assess meaningful hypotheses, represented by Assertions. As the process developed, the analyst created and evaluated Assertions with the available evidence. For example, the analyst suspected the Paramurderers of Chaos as a biological attack threat. This became the premise of the Assertion. To evaluate, the analyst used drag-and-drop gestures to assign evidence to that Assertion. The analyst then assigned a weight to each piece of evidence, represented visually by a symbol next to the evidence. The Assertion represented the combined weights with a bar graph that displayed whether the net weighting of the evidence supported or refuted the hypothesis. This ability of nSpace2 allowed the analyst to approach the task from a more objective standpoint, relying more on the evidence gathered than on possible cognitive bias. Using these tools, the analyst was able to determine the most immediate threat as outlined in the task.
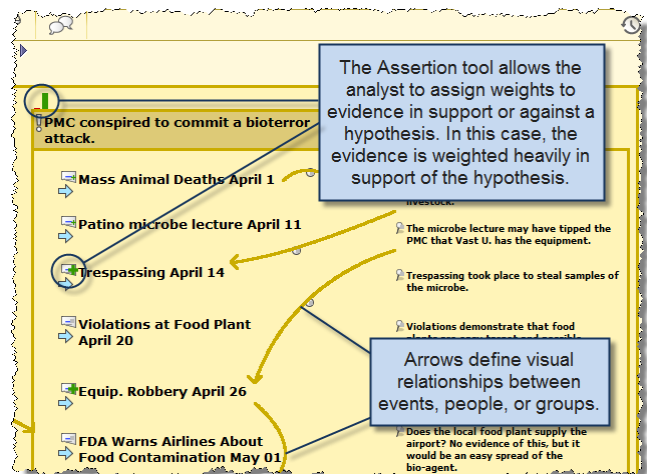

Figure 3. Evaluating evidence in an Assertion.

3 CONCLUSION

nSpace2 has proven to be a powerful system that allowed the analyst to efficiently search, triage, organize, and make sense of the large Mini-Challenge #3 dataset. The task presented by the 2011 VAST Mini-Challenge #3 was the ideal fit for nSpace2's array of capabilities, and demonstrates nSpace2's ability to greatly improve efficiency when analyzing massive amounts of data.

REFERENCES

[1] Heuer, Richard J. Jr., Psychology of Intelligence Analysis, Central Intelligence Agency: Center for the Study of Intelligence, 1999.
[2] Jonker, D., Wright, W., Schroh, D., Proulx, P. and Cort, B. Information Triage with TRIST, In Proceedings of. International Conference on Intelligence Analysis, 2005.
[3] Wright, W., Schroh, D., Proulx, P., Skaburskis, A., and Cort, B. The Sandbox for Analysis – Concepts and Methods, paper accepted for ACM CHI 2006.