

Something's "Fishy" at Global Ways and Gill Breeders – Analysis with nSpace and GeoTime

Lynn Chien

Annie Tat

William Wright

Oculus Info Inc.

ABSTRACT

GeoTime and nSpace are two interactive visual analytics tools that support the process of analyzing massive and complex datasets. The two tools were used to examine and interpret the 2007 VAST contest dataset. This poster paper describes how the capabilities of the tools were used to facilitate and expedite every stage of an analyst workflow.

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

Additional 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 analytical visualization applications that facilitate the process of massive data analysis through supporting different facets of knowledge discovery, knowledge cultivation and analytical reasoning. Both applications were developed in collaboration with analysts and are being used and evaluated by analysts on a day-to-day basis.

1.1 nSpace – Unifying Automatic and User-driven Visual Analysis

nSpace combines TRIST (The Rapid Information Scanning Tool) and the Sandbox, while integrating advanced computational linguistic functions using a Web Services interface and protocol [3]. Analysts alternate between TRIST and the Sandbox to continually develop and refine their analyses.

TRIST is a massive data triaging tool with capabilities such as planned query execution, automatic information extraction, and customizable multi-linked dimensions, which help provide rapid scanning, result characterization and correlation [1]. Information gained from TRIST can be transferred to the Sandbox for evidence marshalling and further analysis.

The Sandbox is a flexible and expressive thinking environment that allows analysts to visualize their cognitive mechanisms. It supports both ad-hoc and more formal analytical sense-making through capabilities such as “put-this-there” cognition, automatic organization of evidence, assertions and evidence assembly, and analysis of competing hypotheses [3].

1.2 GeoTime – Unifying Temporal and Spatial Visual Analysis compete

GeoTime focuses on interactions between entity movements, events and relationships over time within a geospatial (or any conceptual) context to amplify the concurrent cognition of time and space. Entity behaviors and relationships, along with their patterns, in both space and time are easily identified. Entities and events are charted in a single interactive three-dimensional view [2]. The ground plane is the geographic space represented by the X and Y axes, the vertical T-axis represents time. Events can be animated in time.

2 WORKFLOW OF ANALYTICAL PROCESS

2.1 Query - Setting the Stage with Proficiency

Before executing the analysis, the analysts used the expressive thinking environment provided by the Sandbox to plan out the process. Contest instructions were gathered and organized; questions and keywords for querying were generated; notes, thoughts, and prior knowledge were annotated. Since TRIST is able to show thousands of results in one display, the entire data corpus of 1600 documents was indexed and loaded into nSpace and a general framework of the dataset was quickly observed without any work from the user. Icon rows representing results showed the number and the different types of documents; the date published dimension provided a time range of the dataset; automatic dimensions showed the topics. User-tailored dimensions evolved to contain key issues and entities.

2.2 Scan – Triaging Data at a Glance

Relationships between key issues, players, organizations, and locations that were pertinent to answering the contest challenges were tracked by comparing dimensions side by side. A new Links and Nodes visualization in TRIST allowed associations to become more apparent (see middle of Figure 1). The Animal Import database was transferred into GeoTime to further investigate connections between suspicious players and organizations. Using GeoTime's link analysis tools, the history of any individual importer was quickly reviewed.

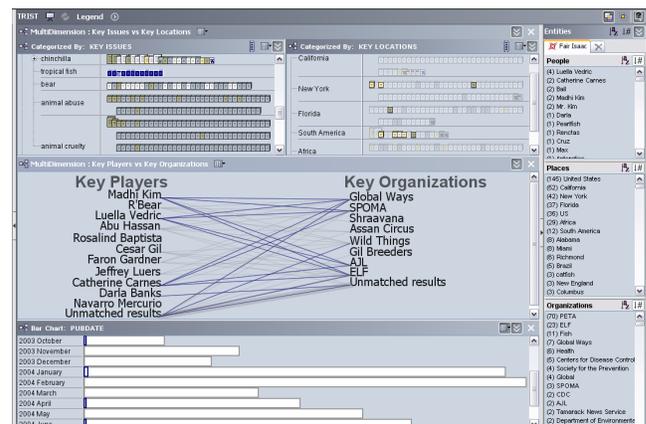


Figure 1. Triaging key topics and relationships in TRIST

2.3 Read – Exploring Relevant Issues in Depth

When pertinent players, organizations, and locations were identified, the analysts began to read relevant documents. As soon as a document was opened, entities and search terms were automatically extracted and then highlighted to facilitate quick identification of key contents.

Once important events had been discovered, a timeline created by the analyst was transferred into GeoTime and events were

plotted in the three-dimensional view and were examined for pattern interpretation (Figure 2).

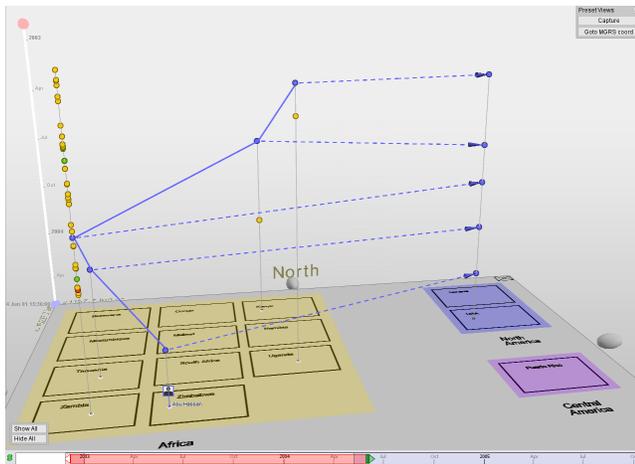


Figure 2. Plotting important events with GeoTime timeline

2.4 Assemble – Organizing and Interpreting Findings

Throughout each step of the analysis, discovered insights and supporting evidence were saved and annotated in the Sandbox. Tools in the Sandbox, such as links, entities, and groups, helped organize the gathered pieces of data. The analysts created assorted layouts that corresponded with their mental models, such as various social networks and hierarchies (Figure 3a).

The Sandbox supported the analysts in the development and assessment of meaningful hypotheses, which were captured as assertions [3]. Evidence for assertions were marshaled through evidence gates by simply dragging and dropping supporting or refuting evidence from the left and right side of an assertion (Figure 3b). Conflicting evidence and competing assertions were analyzed using the new Analysis of Competing Hypotheses tool, which helped clarify and diagnose the strength of each evidence and hypothesis.

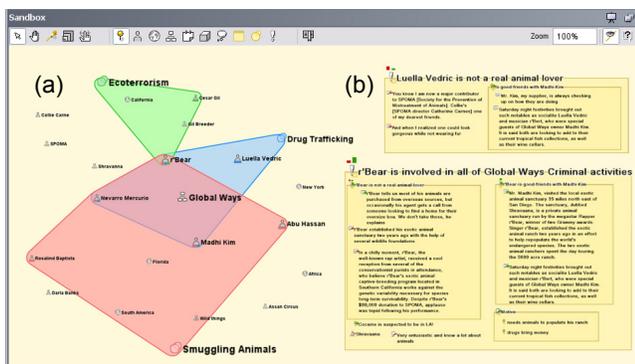


Figure 3. (a) Creating a social network using overlapping colour coded sets (b) Analyzing hypotheses using Assertions

Normally, evidence is organized in a specific way to support each analyst's put-this-there cognition and may be susceptible to pre-dispositions. To ensure all perspectives are considered, different models of organization were applied to existing evidence. The evidence automatically re-organized to conform to a new analytical framework, which helped the analyst exercise a different point-of-view.

2.5 Collaborate – Verifying Conclusions

Multiple analysts were working on the dataset simultaneously and often needed to get together to share their work at different times. Various elements of content in nSpace were exported from one workspace and imported into another. User-defined dimensions, such as Key Players and Key Locations, and templates were exported and shared between analysts and were modified to help uncover unexpected patterns.

Analysts were able to corroborate answers quickly because findings were presented easily using nSpace. When in presentation mode, nSpace menu items became hidden and both TRIST and Sandbox showed findings in a single display. Tools such as the Powerful Finger temporarily enlarged items when the Sandbox was at low zoom and bookmarks allowed desired views to be set and located quickly. Final reports were generated easily by dragging and dropping objects from TRIST and the Sandbox and sources for documents and fragments were automatically added.

3 LESSONS LEARNED

nSpace is a system for supporting analytical work-in-progress. It can be tailored to adapt to both the task at hand, as well as cognitive mechanisms of the analysts. Every step of an analyst's workflow is assisted with automated and user-driven analytical visualizations. Judges from the contest found the automatic cluster dimension a great place to start. nSpace easily suited various types of data provided by the contest and the new ACH tool added utility.

GeoTime enabled the analysis of information connectedness over time and geography within a single, highly interactive 3D view. [2]. Expected relationships were observed and investigated from various view points and unexpected patterns were discovered with the elimination of multiple spread-sheets, tables, maps, and other cross-referenced data that are often simultaneously needed when doing analyses.

Records in CSV files, when imported in nSpace, were processed as a single field rather than separate records with multiple fields. This is a feature that may be implemented in the near future. nSpace is a rich tool with many capabilities. For a novice user, a simpler version would be less overwhelming.

The connection between nSpace and GeoTime included some manual steps and required multiple analysts to work independently. Future goals include linking the two applications for more efficient collaboration.

4 CONCLUSION

nSpace and GeoTime proved to be powerful systems that enabled novice analysts to perform more proficiently full analysis of real-world problems. The VAST 2007 contest allowed the systems to work together and the tools proved to work well both independently and collectively.

REFERENCES

- [1] Jonker, D., W. Wright, D. Schroh, P. Proulx and B. Cort, Information Triage with TRIST, In Proc. International Conference on Intelligence Analysis, 2005.
- [2] Kapler, Thomas, and William Wright, 2004. GeoTime Information Visualization, IEEE Symposium on Information Visualization.
- [3] Wright, William, D. Schroh, P. Proulx, A. Skaburskis and B. Cort, The Sandbox for Analysis – Concepts and Methods, paper accepted for ACM CHI 2006.