IPM – A Framework for Visual Sense Making
VAST 2014 Mini-Challenge 1

Perakath Benjamin* Karthic Madanagopal Belita Gopal Kannan Swaminathan

Knowledge Based Systems, Inc.

{pbenjamin, kmadanagopal, bgopal, kswaminathan}@kbsi.com

ABSTRACT

Analysts are increasingly drowning in a flood of information. KBSI has addressed this challenge through the Intelligence Product Mosaic™ (IPM) analysis, visualization, and reporting tool that exploits the cognitive strength of analysts and enables them to discover information intuitively and quickly. This article provides an overview of the IPM analytic process and the visualizations that were used to enable better end user understanding of the information derived from multi-source text data. The main benefits of the IPM tool described in this paper include (i) significant reductions in 'data-to-decision' time through the use of semantic and collaborative visual analytics techniques and (ii) significant increases in the ability to exploit information and knowledge embedded within data through the use of semantic methods.

Keywords: Visual Analytics, Semantic Tagging, VAST 2014, Natural Language Processing.

Index Terms:

1 INTRODUCTION

The VAST 2014 Mini-Challenge 1 focused attention on the analyzing and making sense of the events surrounding the disappearance of several employees of a fictitious natural-gas company, GAStech. The data provided include 845 news articles, 2 historical documents on an organization named Protectors of Kronos (POK), resumes of 35 employees of GAStech, header data of emails sent over a two-week period surrounding the kidnapping, a spreadsheet containing employee information, factbooks on Kronos and Tethys, and a map of Kronos. The task was to describe the events surrounding the kidnappings, provide a visual representation of the leadership of POK and its evolution over time, and suggest possible reasons for the disappearance all with the goal of assisting the law enforcement.

Our analysis was supported by KBSI's Intelligence Products Mosaic™ (IPM) and the OSR Studio™ -- software tools that were developed by KBSI. First we used the KBSI Natural Language Processing (KNLP™) Pipeline to convert unstructured text into a form that is tagged with concepts. During this process, named entities such as people, location, organization, etc. are identified and tagged. A Query Interface provided by the OSR Studio™ is used to query for concepts such as *leadership* and *protest*. At this stage we also used Microsoft Excel® to perform simple ad-hoc queries to gain a better understanding of the data. Finally, IPM is used to visualize POK social networks, its leadership and changes in its organizational structure over time, email traffic, timeline of key events, etc. This article provides an overview of the analytic process and the visualizations that were used to enable better end user understanding of the information derived from the data.

* corresponding author email: pbenjamin@kbsi.com

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

2 ANALYTICAL PROCESS

The VAST data provided was processed through the KBSI Natural Language Processing (KNLP™) pipeline, which converted unstructured text into a form tagged with concepts from an ontology. During this process, named entities such as people, location, organization, etc. were also identified and tagged. Along with the social entities, our pipeline also identifies a rich set of events that are defined in the underlying ontology. The KNLP™ pipeline generated a common RDF store as shown in Figure 1.

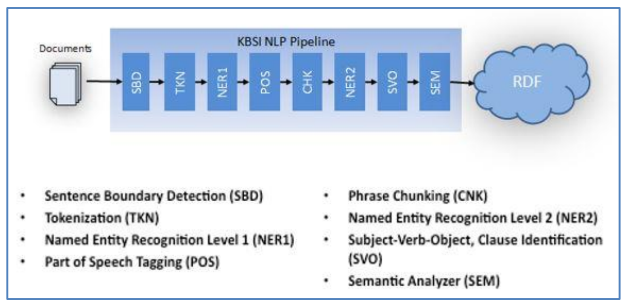

Figure 1: The KNLP™ Pipeline

2.1 Semantic Tagging

Semantic tagging refers to the activity of 'labeling' data with 'tags' that provide 'meaning' in the context of an application about the information contained in the data [1]. A unique aspect of the KNLP™ pipeline-based semantic tagging approach is the use of an ontology to enhance the semantic quality (depth of meaning represented) in the resulting tagged text. This also includes the capability to generate Resource Description Format (RDF) semantic labels from multi-source text data.

2.2 Semantic Search

After the RDF store was generated, it was queried through the OSR Studio™ to support the investigation process. This tool supports complex queries that makes uses of the richness of the semantic tags in the dataset. For example, to get a list of the POK leaders, queries were constructed to search for "POK" term and *leadership* concept. The result of this search provided an insight into who the leaders were and where the information came from.

3 VISUALIZATION

Very frequently, disparate data sources have to be analyzed collectively. The VAST 2014 MC1 data contained unstructured news articles and structured email headers. Our KNLP™ pipeline and the data-linking module help to combine the data from different sources and types into a single repository and advanced text analytics methods are used to explore the data. The IPM provides various visualization mechanisms that allow the analyst to ‘slice and dice’ the data for better sense making.

3.1 Social Network Viewer

A set of queries constructed using the OSR Studio™ was used to extract the social network from the MC1 dataset. The network extracted was then visualized using the IPM social network viewer as shown in Figure 2. The social entities are color coded to show their affiliations for example, POK people are shown in blue and GAStech people are shown in red.

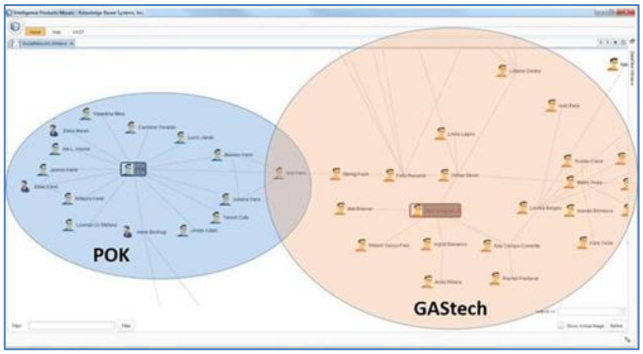

Figure 2: Intersection of Two Social Networks

The social network viewer superimposed with the timeline slider helps an analyst pinpoint key people who are active in each network during the selected time range. The playback feature helps in visualizing evolution of networks.

3.2 Email Explorer

One of the data sets provided was a series of emails between GAStech employees spanning 2 weeks. To support better data exploration and discovery, the graph viewer component of the IPM was used to visualize the communication network. This interface allows an analyst to filter the network by email metadata and message headers Figure 3.

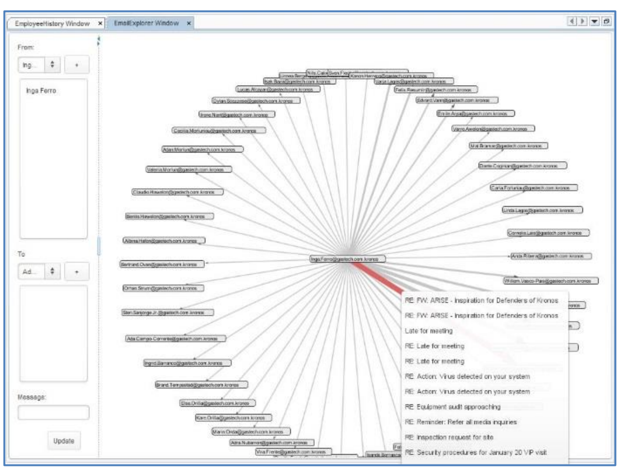

Figure 3: Visualizing Email Communications

A directed graph would be shown to the user with varying link strength and the link strength is determined by the frequency of communication that happens between the employees. Clicking a node (employee) displays the information associated with that employee and clicking on a link displays the headers of all the emails exchanged between the 2 nodes. An email with the subject “ARISE - Inspiration for Defenders of Kronos” was being circulated amongst a set of security personnel. Further analysis revealed that all of them shared the last name of known POK members.

3.3 Report Generation

Most data analysis platforms support only data exploration, forcing the analyst to use another tool for report generation. A key distinguishing feature of IPM is its advanced report generation capability which enables the analyst to create reports on the fly while performing the data exploration. The analyst can create templates for their reports and subsequently add details to each section. The analysis views may be captured using the built-in snapshot tool and added to the report (Figure 4).

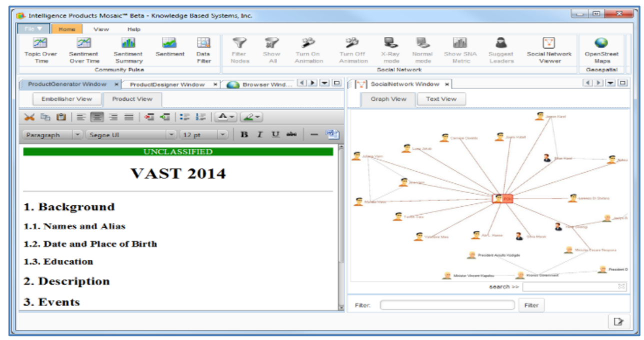

Figure 4: Snapshot and Reporting

This module maintains the pedigree of the data, allowing a quality report to be generated in a short period of time. The report can then be exported as a Microsoft Word® document or as an html page for publishing.

4 Conclusion

Analysts are increasingly drowning in a flood of information. KBSI has addressed this challenge through the Intelligence Product Mosaic™ (IPM) analysis, visualization, and reporting tool that exploits the cognitive strength of analysts and enables them to discover information intuitively and quickly. The main benefits of the IPM tool described in this paper are (i) significant reductions in ‘data-to-decision’ time through the use of semantic and collaborative visual analytics techniques; (ii) significant increases in the ability to exploit information and knowledge embedded within data through the use of ontology-driven semantic methods; and (iii) superior abilities to leverage and better use increasingly scarce and time-constrained human cognitive and decision making skills, through the use of interactive sense making techniques.

References

[1] P. Benjamin, K. Madanagopal, K. Vadakkeveedu, S. Ramachandran and P. Koola, "A Framework for Multi-source Semantic Information Extraction & Fusion for Collaborative Threat Assessment (SIFT)," in *WorldComp2013 - ICAI*, Las Vegas, 2013.

[2] J. Heer and B. Shneiderman, "Interactive Dynamics for Visual Analysis," *Commun. ACM,* pp. 45-54, 2012.