Safeguarding Abila through Multiple Data Perspectives
VAST 2014 Grand Challenge Award: Effective Analysis and Presentation

Parang Saraf[*] Patrick Butler[†] Naren Ramakrishnan[‡]

Discovery Analytics Center
Department of Computer Science
Virginia Tech

Abstract

We introduce a system for visual analysis of news articles, emails, GPS tracking data, financial transactions and streaming micro-blog data. This system was developed in response to the 2014 VAST Grand Challenge and comprises of several interfaces for mining textual, network, spatio-temporal, financial, and streaming data.

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Interaction styles (e.g., commands, menus, forms, direct manipulation)

1 Introduction and Problem Overview

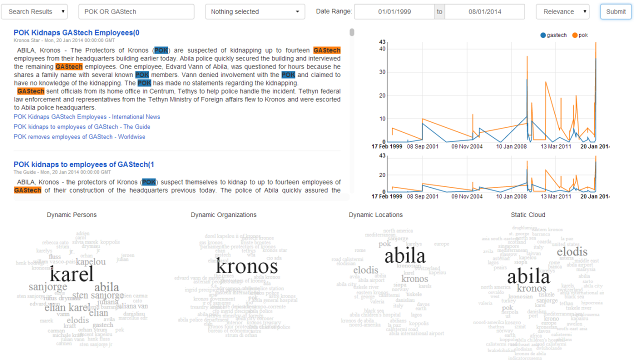
The VAST 2014 grand challenge describes a hypothetical scenario wherein some of the employees of an imaginary organization, GAStech have gone missing and it is speculated that an environmental activist group, Protectors of Kronos (POK) is responsible behind the disappearance. The challenge requires analysis of a variety of datasets in order to provide crucial leads to law enforcement agencies about suspicious persons, locations, organizations, events, etc. The provided dataset includes unstructured news articles, email headers from GASTech's company email, GPS tracking data of company cars assigned to employees, financial transaction data of employees' credit card & loyalty card, streaming data from a micro-blogging website and control room data of law enforcement agencies. The approach required to solve the challenge is akin to completing a jigsaw puzzle. Analysis of each of the datasets reveals partial information about the plot. An analyst needs to put all these pieces together in order to reveal the complete plot.

2 System Design

We developed a web-based visual analytics system that provides a set of tools and widgets specifically designed to analyze the given datasets. The system makes use of Google maps to display geo-spatial data and the Javascript-based graphical libraries `d3.js` [1] & `nvd3.js` for charts and graphs.

2.1 Unstructured News Articles

Figure 1 demonstrates the interface used for analyzing news articles. The interface makes use of a Python based search engine, Whoosh that allows logical queries and returns relevant news articles where similar articles are grouped together. The returned news articles are analyzed further using the Stanford's Named Entity Recognition tool [3], the results of which are displayed using word clouds. A time series plot shows keyword frequency over time, thereby giving insights into correlations between searched keywords.



POK Kidnaps GAStech Employees(0
Kronos Star - Mon, 20 Jan 2014 00:00:00 GMT

ABILA, Kronos - The Protectors of Kronos (POK) are suspected of kidnapping up to fourteen GAStech employees from their headquarters building earlier today. Abila police quickly secured the building and interviewed the remaining GAStech employees. One employee, Edvard Vann of Abila, was questioned for hours because he shares a family name with several known POK members. Vann denied involvement with the POK and claimed to have no knowledge of the kidnapping. The POK has made no statements regarding the kidnapping.

GAStech sent officials from its home office in Centrum, Tethys to help police handle the incident. Tethyn federal law enforcement and representatives from the Tethyn Ministry of Foreign affairs flew to Kronos and were escorted to Abila police headquarters.

POK Kidnaps GAStech Employees - International News
POK kidnaps to employees of GAStech - The Guide
POK removes employees of GAStech - Worldwise

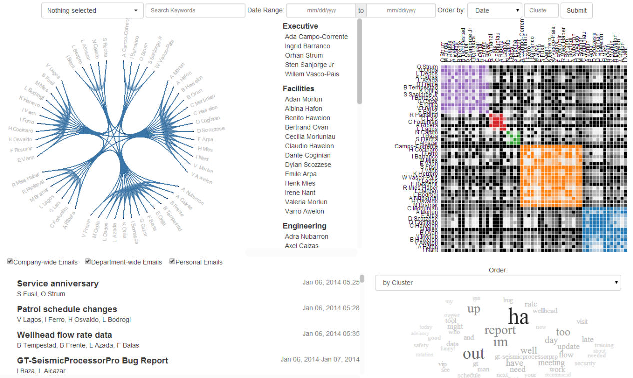
POK kidnaps to employees of GAStech(1
The Guide - Mon, 20 Jan 2014 00:00:00 GMT

ABILA, Kronos - the protectors of Kronos (POK) suspect themselves to kidnap to up to fourteen employees of GAStech of their construction of the headquarters previous today. The police of Abila quickly assured the

Figure 1: News Interface

2.2 Email Headers

The email header information (see Figure 2) is analyzed using the following three widgets: radial graph for examining the underlying email network structure, co-occurrence matrix for identifying users who frequently exchange emails between them and a search interface for querying the emails. The co-occurrence matrix uses spectral co-clustering [2] to group users together and was implemented using the Python based scikit-learn package [4].



Executive
Ada Campo-Corrente
Ingrid Barranco
Orhan Strum
Sten Sanjorge Jr
Willem Vasco-Pais

Facilities
Adan Morlun
Albina Hafon
Benito Hawelon
Bertrand Ovan
Cecilia Morluniau
Claudio Hawelon
Dante Coginian
Dylan Scozzese
Emile Arpa
Henk Mies
Irene Nant
Valeria Morlun
Varro Awelon

Engineering
Adra Nubarron
Axel Calzas

Service anniversary Jan 06, 2014 05:25
S Fusil, O Strum

Patrol schedule changes Jan 06, 2014 05:28
V Lagos, I Ferro, H Osvaldo, L Bodrogi

Wellhead flow rate data Jan 06, 2014 05:35
B Tempestad, B Frente, L Azada, F Balas

GT-SeismicProcessorPro Bug Report Jan 06, 2014-Jan 07, 2014
I Baza, L Alcazar

Figure 2: Email Header Interface

2.3 GPS Tracking Data

Three different interfaces were built to process GPS tracking data. The first interface (see Figure 3) allows for playback of GPS coordinates. Using this interface an analyst can visualize the movement of cars over time.

The second interface (see Figure 4) examines user's location information along with financial transaction data to geo-locate recreational establishments on a map.

[*] e-mail: parang@cs.vt.edu
[†] e-mail: pabutler@vt.edu
[‡] e-mail: naren@cs.vt.edu

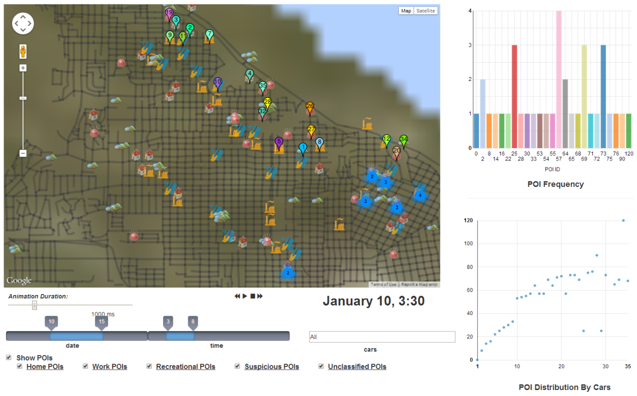
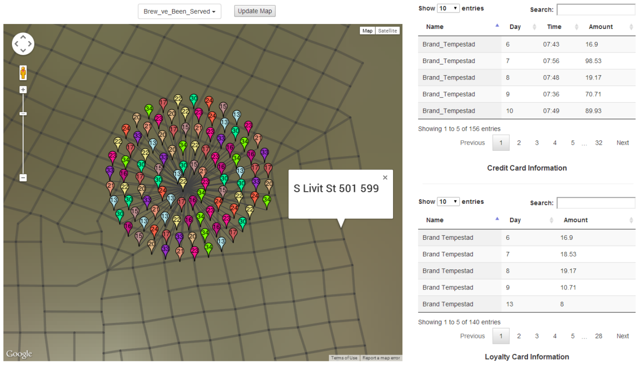


Figure 3: Location Playback Interface



Show 10 entries Search:

Name	Day	Time	Amount
Brand_Tempestad	6	07:43	16.9
Brand_Tempestad	7	07:56	98.53
Brand_Tempestad	8	07:48	19.17
Brand_Tempestad	9	07:36	70.71
Brand_Tempestad	10	07:49	89.93

Showing 1 to 5 of 156 entries

Previous 1 2 3 4 5 … 32 Next

Credit Card Information

Show 10 entries Search:

Name	Day	Amount
Brand Tempestad	6	16.9
Brand Tempestad	7	18.53
Brand Tempestad	8	19.17
Brand Tempestad	9	10.71
Brand Tempestad	13	8

Showing 1 to 5 of 140 entries

Previous 1 2 3 4 5 … 28 Next

Loyalty Card Information

Figure 4: Recreational POIs Interface

The third interface (see Figure 5) provides widgets to explore locations which are frequently visited by users. These locations, also known as Points of Interest (POI) can be classified as home, work, recreational and suspicious based on their distribution and frequency over time.

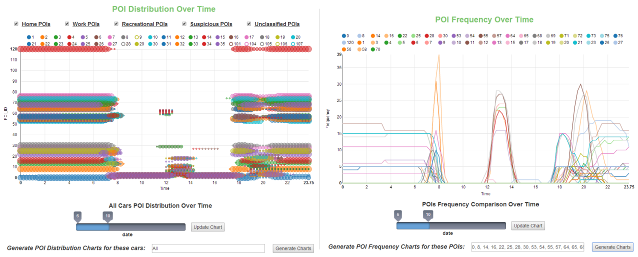


Figure 5: Points of Interest Analysis Interface

2.4 Financial Transaction Data

The interface for analyzing user spending data (see Figure 6) provides three widgets: employee vs employee spending comparison, employee spending distribution, and establishment sales distribution. As the name suggests, employee vs employee spending comparison compares employee spending across company as well as across departments that he/she works for. Employee spending distribution displays the spending of an employee across all establishments as well as across days. Similarly, establishment sales distribution compares the sales of a particular establishment across all employees as well as over days.

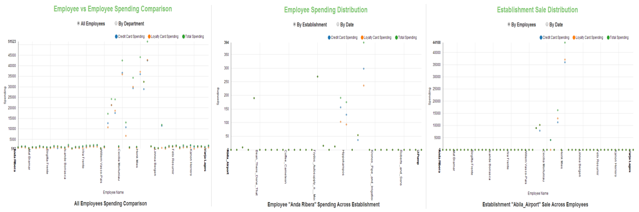
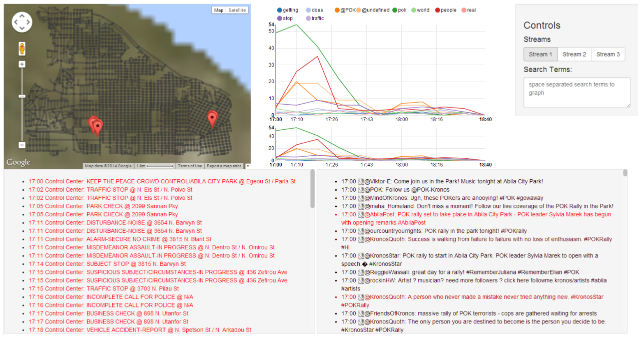

Employee vs Employee Spending Comparison
All Employees
By Department
Credit Card Spending
Loyalty Card Spending
Total Spending
Employee Name
All Employees Spending Comparison
Employee Spending Distribution
By Establishment
By Date
Employee "Anda Ribera" Spending Across Establishment
Establishment Sale Distribution
By Employees
By Date
Establishment "Abila_Airport" Sale Across Employees

Figure 6: Financial Data Interface

2.5 Streaming Micro-blogging and Control Room Data

Figure 7 presents the interface for visualizing streaming data. In this interface, micro-blog records are displayed in different shades of red, where the gradient of the color defines the frequency with which a message is re-posted. Darker the color higher the frequency of re-posted message. An analyst can visualize geo-tagged posts on a map and can filter the micro-blog and control room data by searching for keywords. A line chart shows the frequency of searched keywords over time.



- 17:00 Control Center: KEEP THE PEACE-CROWD CONTROL/ABILA CITY PARK @ Egeou St / Parla St
- 17:02 Control Center: TRAFFIC STOP @ N. Els St / N. Polvo St
- 17:02 Control Center: TRAFFIC STOP @ N. Els St / N. Polvo St
- 17:05 Control Center: PARK CHECK @ 2099 Sannan Pky
- 17:05 Control Center: PARK CHECK @ 2099 Sannan Pky
- 17:11 Control Center: DISTURBANCE-NOISE @ 3654 N. Barwyn St
- 17:11 Control Center: DISTURBANCE-NOISE @ 3654 N. Barwyn St
- 17:11 Control Center: ALARM-SECURE NO CRIME @ 3815 N. Blant St
- 17:11 Control Center: MISDEMEANOR ASSAULT-IN PROGRESS @ N. Dentro St / N. Omirou St
- 17:11 Control Center: MISDEMEANOR ASSAULT-IN PROGRESS @ N. Dentro St / N. Omirou St
- 17:14 Control Center: SUBJECT STOP @ 3815 N. Barwyn St
- 17:15 Control Center: SUSPICIOUS SUBJECT/CIRCUMSTANCES-IN PROGRESS @ 436 Zefirou Ave
- 17:15 Control Center: SUSPICIOUS SUBJECT/CIRCUMSTANCES-IN PROGRESS @ 436 Zefirou Ave
- 17:15 Control Center: TRAFFIC STOP @ 3703 N. Pilau St
- 17:16 Control Center: INCOMPLETE CALL FOR POLICE @ N/A
- 17:16 Control Center: INCOMPLETE CALL FOR POLICE @ N/A
- 17:17 Control Center: BUSINESS CHECK @ 898 N. Utanfor St
- 17:17 Control Center: BUSINESS CHECK @ 898 N. Utanfor St
- 17:18 Control Center: VEHICLE ACCIDENT-REPORT @ N. Spetson St / N. Arkadou St

- 17:00 @Viktor-E: Come join us in the Park! Music tonight at Abila City Park!
- 17:00 @POK: Follow us @POK-Kronos
- 17:00 @MindOfKronos: Ugh, these POKers are anooying! #POK #gowaway
- 17:00 @maha_Homeland: Don't miss a moment! Follow our live coverage of the POK Rally in the Park!
- 17:00 @AbilaPost: POK rally set to take place in Abila City Park - POK leader Sylvia Marek has begun with opening remarks #AbilaPost
- 17:00 @ourcountryourrights: POK rally in the park tonight! #POKrally
- 17:00 @KronosQuoth: Success is walking from failure to failure with no loss of enthusiasm. #POKRally #HI
- 17:00 @KronosStar: POK rally to start in Abila City Park. POK leader Sylvia Marek to open with a speech.� #KronosStar
- 17:00 @ReggieWassali: great day for a rally! #RememberJuliana #RememberElian #POK
- 17:00 @rockinHW: Artist ? musician? need more followers ? click here followme.kronos/artists #abila #artists
- 17:00 @KronosQuoth: A person who never made a mistake never tried anything new. #KronosStar #POKRally
- 17:00 @FriendsOfKronos: massive rally of POK terrorists - cops are gathered waiting for arrests
- 17:00 @KronosQuoth: The only person you are destined to become is the person you decide to be. #KronosStar #POKRally

Figure 7: Streaming Data Interface

3 Results

Using these different interfaces, we were able to unearth reliable evidence suggesting that Protectors of Kronos is indeed responsible for the disappearance of employees. Additionally, a few GAStech employees involved in suspicious activities were also identified.

Acknowledgements

We would like to thank Ritika Dokania for her creative inputs and feedback on the visualization, as well as for lending her voice to the explanatory video that describes the system. This work is partially supported by US NSF Grant CCF-0937133.

References

[1] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.

[2] I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 269–274. ACM, 2001.

[3] T. G. Jenny Rose Finkel and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pages 363–370, 2005.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.