

# AnnotatedTimeTree: Visualization and Annotation of News Text and Other Heterogeneous Document Collections

Jing Xia, Jieqiong Zhao, Isaac Sheeley, Joseph Christopher, Qiaoying Wang, Chen Guo, Jiawei Zhang, David S. Ebert, Yingjie Victor Chen, Zhenyu Cheryl Qian

**Abstract**—Investigation of critical events requires dealing with heterogeneous document collections, and is usually done case by case. We address the case of VAST Challenge 2014 Mini-Challenge 1 with the two views of our *AnnotatedTimeTree* system: *the document view* and *the GASTech view*. The document view is a novel AnnotatedTimeTree design for current and historical news reports in Kronos; and the GASTech view is implemented specifically for GASTech employee profiles and email activities. Together the two views facilitate rapid summarization of both crucial events in Kronos history and GASTech employees in Mini Challenge 1. The AnnotatedTimeTree design provides semantic interactions such as searching, highlighting, annotating and expanding for maintaining reasoning process and reasoning results to the documents.

**Index Terms**—Temporal visualization, document analysis

## 1 INTRODUCTION

Investigation of critical events requires putting together heterogeneous data collections. It requires human intelligence integration and thus is usually performed case by case. Good visual analytical tools can integrate human interactions and promote reasoning from existing data resources.

The VAST Challenge 2014 Mini-Challenge 1 is about investigation of missing employees in GASTech, a natural gas company. The suspect POK (protectors of Kronos) were originated from a small village, which had been long suffering from dirty water polluted by GASTech. Participants are given several datasets including GASTech employee profiles, resumes and email conversations, historical news reports regarding the conflicts between GASTech and POK. The tasks are to find out the development of POK, the event timeline, and to identify possible explanations why those employees were missing.

To address the problem of VAST Challenge 2014 Mini-Challenge 1, we developed AnnotatedTimeTree, a two-view online system to integrate the datasets: the document view for analysis of historical reports, and the GASTech view for exploration of GASTech employees. Specifically, the AnnotatedTimeTree in the document view is an innovative design we created to use a timeline to organize reports and analysts' annotations of these reports. It provides semantic interactions such as searching, highlighting, annotating and expanding for maintaining reasoning process and reasoning results to the documents. The GASTech view enables interactive investigations of employees based on given profiles, resumes and email conversations. Together with the two views, the AnnotatedTimeTree system promotes users reasoning process in discovering possible explanations of missing employees.

## 2 VISUAL EXPLORATION

AnnotatedTimeTree is composed of a news report timeline, an annotation area, a navigation bar and a report reader. Interactions such as highlighting, navigating, filtering are provided for faster locating of crucial persons, locations etc. The GASTech view consists of an e-mail explorer, an employee resume viewer and an employee network visualization.

## 2.1 AnnotatedTimeTree Design

AnnotatedTimeTree (see Fig. 1) is a day-based report timeline with each block (the unit on the timeline) being green color coded to the number of reports concerning a certain person on a certain day. We picked 11 names that have the highest frequencies in the reports, with all the remaining persons falling into the "Other" category. Grey areas are applied to those days with no reports since the news report dataset given is not continuous. Because the overall timeline is too long to show, a zoomed-out overview is provided to the right for preview and navigation. To the left is the report pile area, where users can browse the reports one by one and make annotation. The annotations being made are aligned next to its timeline location for browse.



Fig. 1. The document view of AnnotatedTimeTree. From left to right there are the report pile area with search bar at its top (a), the timeline (b) with annotations attached to its left and right and the thumbnail navigation bar (c).

**Navigate** With the timeline preview, users are able to find out days of interest by locating dark green blocks. We also notice a pattern of two-column green blocks. After inspection of those blocks, it turns out to be all conflicts between POK and GASTech in memory of POK leader Elian Karel and the girl Juliana Vann. Users can also navigate to those days by clicking on the timeline preview.

**Search** If users want to narrow the reports down to a smaller amount, they may search on the left with duration of time and keywords. All results are highlighted with the given keywords.

**Highlight** Other than highlights of keywords, names of persons, locations and organizations are also highlighted with different colors

• Jing Xia is with State Key Lab of CAD&CG, Zhejiang University in Hangzhou, China. E-mail: xiajing@zjucadcg.cn.  
• Jieqiong Zhao, Isaac Sheeley, Joseph Christopher, Qiaoying Wang, Chen Guo, Jiawei Zhang, David S. Ebert, Yingjie Victor Chen, Zhenyu Cheryl Qian are with the University of Wisconsin, Stevens Point, WI, USA. E-mail: jzhao@uwsp.edu, isaelee@uwsp.edu, jchristo@uwsp.edu, qwang1925@uwsp.edu, guo171@uwsp.edu, zhan1486@uwsp.edu, eberd@uwsp.edu, yingjie@uwsp.edu, zhenyu@uwsp.edu.  
IEEE Symposium on Visual Analytics Science and Technology 2014, November 8-14, Paris, France, 978-1-4799-6227-3/14/\$31.00 ©2014 IEEE

to draw users attention.

**Annotate** Users can make annotations or modify annotations at the end of reports. Based on the critical level of the annotations, they are color coded differently. Annotations to the same person of the same day are aligned and ordered in a string. Once changed, the corresponding annotation will be created or modified to maintain users' investigation on the report.

**Expand** When there are large numbers of reports accumulated in one day, users can expand the day by double clicking on a certain green block. Corresponding reports will be expanded and sorted according to time of day (see Fig. 2). Additionally, the annotations will be re-aligned correspondingly.

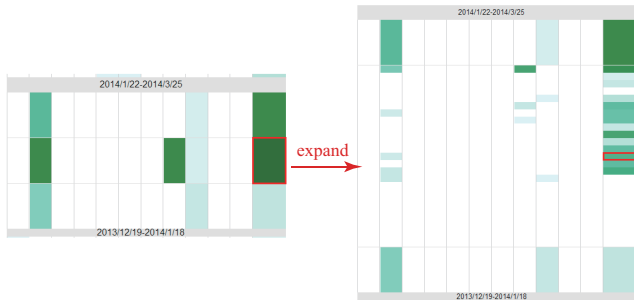


Fig. 2. Timeline of one day is expanded to 24 hours.

## 2.2 GAStech Profile Design

The GAStech view (see Fig. 3) integrates GAStech organization chart, employee resumes, profiles and their email conversations altogether in one view. A network based on the email conversation is built to explore employee relations. Employees are color coded according to the departments they belong. Users can mark emails as important and thus narrow the email network down to senders and receivers of important emails only. Users can also browse employees' profiles, resumes and car activities (for Mini-Challenge 2).

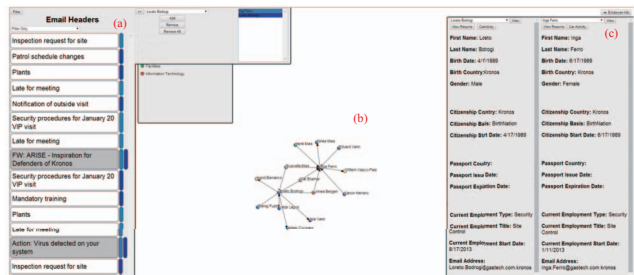


Fig. 3. The GAStech view of AnnotatedTimeTree. From left to right there are the email conversation viewer (a) where users can sort, mark and check details of the emails, the employee network (b) based on email conversations, and the employee resume viewer (c) where users can compare resume or profiles of any two employees.

## 2.3 Analysis Process

Fig. 4 shows that there had been several conflicts between POK and GAStech ever since the death of ten-year old girl Juliana Vann. The situation became worse after former POK leader Elian Karel's death. The annual rally was held and ended in riot.

By investigating the reports, users are able to efficiently locate those reports to summarize a timeline of what happened (see Fig. 1). First at 09:26 January 20th, GAStech CEO Sten Sanjorge was celebrating

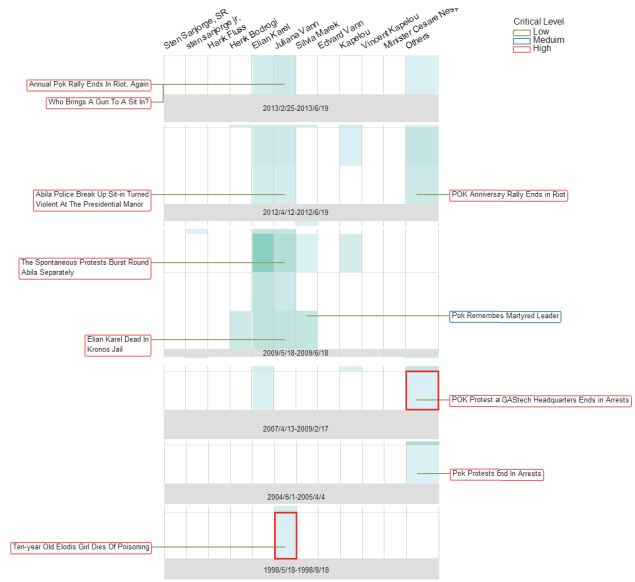


Fig. 4. Summary of historical conflicts between POK and GAStech.

the successful initial public offering of his company. Later the celebration was interrupted by a false fire alarm. Then at 12:20 the police of Abila were called to GAStech for an unspecified emergency. At 15:00 Edvard Vann was interrogated for hours. The last name "Vann" is shared by several key POK members. Though unofficially, a government spokesman stated that GAStech employees had been kidnapped at 20:20. At 22:00 Tethys federal law enforcement officials and staff from the Tethyn Ministry of Foreign Affairs arrived in Kronos to assist police in the investigation. On January 21st, 2014 POK announced the kidnapping and demanded 20M.

### 2.3.1 Findings

A reasonable explanation of employees' disappearance in GAStech might be they were kidnapped by GAStech employees who are POK members indeed. Two evidences support the finding. First, there had been a long history of conflicts (see Fig. 4) between POK and GAStech. Second, some GAStech employees from security department share the same family names with leaders or founders of POK, and they communicated frequently via emails (see Fig. 3). The responsible GAStech employees are Inga Ferro, Loreto Bodrogi, Isia Vann, Hennie Osvaldo, Minke Mies and Ruscella Mies, all from security department.

We also found something interesting (though not reflected in the figures), that different news media tends to have different opinions towards the same events. "Homeland Illumination" seemed to speak for POK while "The Abila Post" stood for governments' opinion.

## 3 CONCLUSION

AnnotatedTimeTree is an online system with two views for visualization of historical reports and GAStech profiles respectively. The AnnotatedTimeTree design in the document view allows users to make annotations on reports dynamically, which not only records and communicates users' understanding but also maintains their reasoning process towards the overall story. This design can be easily extended to analysis of time-critical documents.