College of Information Science and Technology

Drexel E-Repository and Archive (iDEA)
http://idea.library.drexel.edu/

Drexel University Libraries
www.library.drexel.edu

Please direct questions to archives@drexel.edu

Visual Analysis of Terrorism Events Extracted From the Public Knowledge Bases

Weizhong Zhu, Chaomei Chen
College of Information Science and Technology, Drexel University

Introduction

Understanding the conceptual, temporal and social properties of terrorism is essential to unlock terrorists' minds, track terrorists and make policies for emergency responses to terrorists' attacks. The primary goal of this poster [1] is to introduce an integrative approach to study the ideology of terrorism, social structures of terrorists' organizations and consistent patterns of terrorism events hidden in the public knowledge bases. Storylines [2] provides a novel storytelling framework to examine our approach and to help user visually and interactively explore textual information from multiple resources with diverse perspectives. Our system has been applied to the 2006 VAST contest data that is a synthesized dataset. The first test case here includes 948 real-world news from 1997-2006 and 1427 attack incidents from 1990-2002 extracted from a concentrated terrorism resource ICT (http://www.ict.org.il/) . The second test case has 275 historical documents extracted from the Avalon project of Yale Law School, a web resource related to Cuban Missile Crisis.

Research Questions

Our approaches aim to solve the tasks with the questions like:
What are the major terrorists' organizations at the Middle East? And what is the ideology for the organization?
Who are the key leaders of a terrorists' group?
When are the peak periods of terrorism attacks? And How and Why?
Are there emergent patterns connecting attacks with certain groups?
What is the lesson learned from an emergency case, such as Cuban Missile Crisis?

System Design and Architecture

Storylines enables user study a body of unstructured text without prior knowledge of its thematic structure and automatically find Who, When, What and Where in a salient story. It integrates natural language processing, latent semantic indexing and social network analysis. Natural language processing extracts data signatures, such as single key words, n-grams and named entities (People, Location, Organization and Time). Latent semantic analysis supports visually systematic exploration of the data signatures with reference to quantitative measures of importance and their associations. The importance of the data signatures is measured by their statistical contributions to the global latent concept space or a local latent concept dimension. Social network analysis combines co-occurrence analysis of named entities and importance measures such as degree centrality and betweenness centrality. Named entity inter and intro relationships networks aware user the social structure and key players or main locations in the stories.
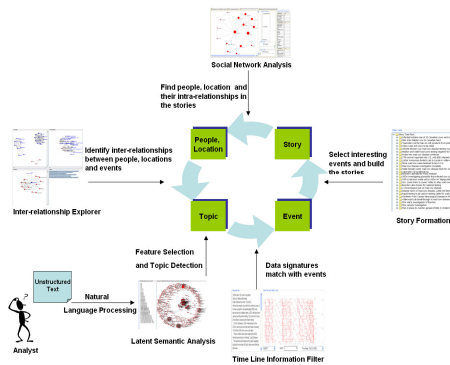

Fig.1 Storylines provides six coordinated views: latent concept explorer, document view, incremental event timeline filter, term view, story lines and associative named entity network. The diagrams are generated based on 1,182 news articles in the VAST contest dataset.

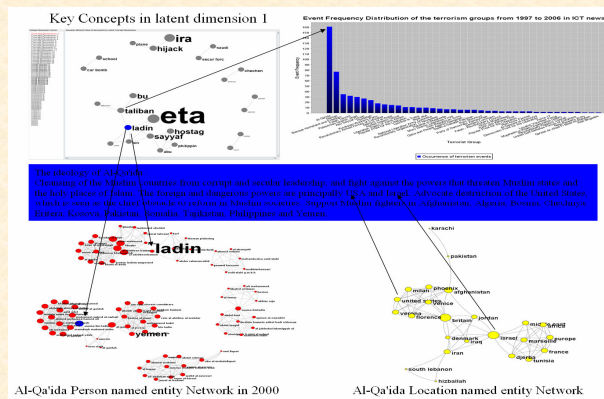## Test Case One: ICT News and Terrorism Attack Incidents

Fig.2 An integrative view of the key concepts (Bin Ladin, etc. al), ideology of Al-Qa'ida, person named entity networks and location named entity networks extracted from 161 news as one story that is related to the Al-Qa'ida terrorist group from 1997-2006.
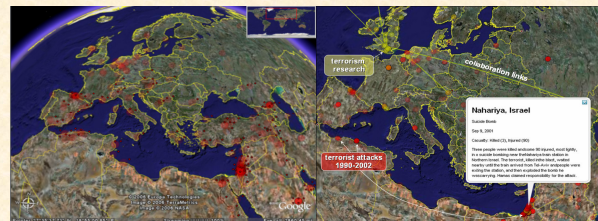

Fig.3 The Google Earth map left depicts geospatial distributions of 1427 terrorist incidents such as suicide bombings, shooting, and kidnappings from 1990-2002 [3]. The map right shows the details of a terrorist incident in Nahariya, Israel and a few collaboration links in the same timeframe, 2001. Terrorist incidents are shown as red markers.
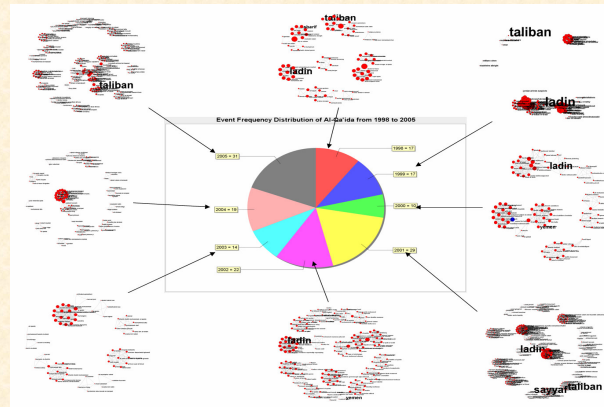

Fig.4 This diagram shows the temporal variations of Al-Qa'ida attacking frequency and the social structures of the group from 1998 to 2005. The evidence indicates that before 9/11, 2001 (a peak period of the terrorism attacks), Bin Ladin had been a key player in the terrorists' social networks. It suggests there is possibility to predict the attacks by monitoring the social structure changes of the terrorists' social networks extracted from the public knowledge bases.

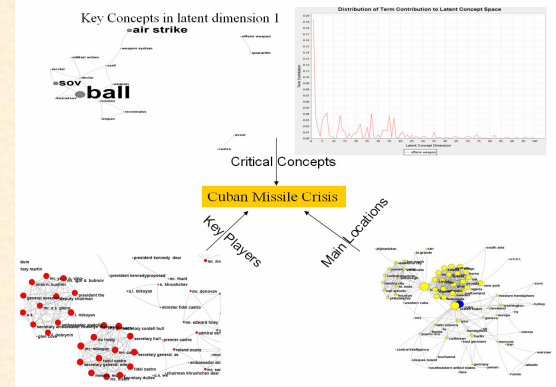## Test Case Two: Cuban Missile Crisis

Fig.5 An overview of Cuban Missile Crisis according to the critical concepts (for instance, offensive weapon), Person and Location named entity networks extracted from 275 documents as one story. These views enhance the understanding to the main arguments and historical context involved in that famous crisis.

Conclusion

1. A novel storytelling framework for visual analytics systems is designed and built to help user understand unstructured text.
2. Storylines provides a unique way to investigate the nature of the latent concept spaces within unstructured text through term-contribution networks.
3. The system identifies key concepts, key actors and main locations with reference to quantitative measures of importance and the hidden links between them.
4. The integrated analysis process, the multiple-dimension views of the latent concept space and the dynamics of the social networks show great potential to accelerate user sense making processes involving high-complexity and high-dimensionality problems.

Future Work

1. Study the conceptual and geospatial knowledge diffusion [3] of terrorism and predict the trends.
2. Extend the work to large scale real-world datasets such as news archives, live news feeds, email archives, citation records and web blogs.
3. Improve the features on usability and user collaboration towards an on-line version.

Reference

[1] The partial content of this poster was represented at the First Annual DHS University Network Summit on Research and Education as a student poster.
[2] Zhu, W., Chen, C. (2007) Storylines: Visual exploration and analysis in latent semantic spaces. *International Journal of Computers and Graphics*. Special Issue on Visual Analytics. In Press.
Video Demo: http://cluster.cis.drexel.edu/~wzhu/swf_output/storylines_full.htm
[3] Chen, C., Zhu, W., Tomaszewski, B., MacEachren, A. (2007) Tracing conceptual and geospatial diffusion of knowledge. HCI International 2007. Beijing, China. July 22-27, 2007
[4] Zhang, J., Zhu, W., Chen, Y., Vogeley, M., Chen, C., Analyzing the Impact of Sloan Digital Sky Survey on Astronomical Literature: A Multiple Perspective Approach, a poster paper in ISSI 2007. Storylines is applied partially.

Acknowledgement