Health GeoJunction: tracking infectious disease threats and related science

Michael Stryker (in collaboration with Scott Pezanowski, Ian Turton, and Alan MacEachren)
Department, North-East Visualization & Analytics Center, Penn State University

RVAC Regional Visualization and Analytics Centers

NEVAC

Introduction:

Public health risks can become public health crises if they are not recognized, assessed, and responded to early. Effective situation monitoring, assessment and response requires well informed analysts and requisite information is often scattered and diverse.

Monitoring outbreak events, collaborating in research and public health practice, and making decisions about how to structure and carry out surveillance and response activities requires an understanding of information that may be buried within the text of numerous documents at different locations and in different formats.

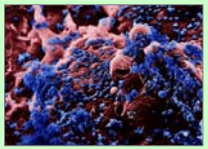


Image of H5N1 virus

1918 Influenza Pandemic

Research Objectives:

To build a visual analytic environment for quickly making sense of a collection of documents through interactive visual representations of the ***geographic***, ***thematic***, and ***temporal*** references within the text. And thus, support:

(1) **situation assessment** enabling an analyst to quickly become apprised of the current and evolving state of science and public health policy related to an infectious disease threat (e.g., an avian flu pandemic threat).

(2) **situation awareness** by providing access to key indicators of an evolving threat for an analyst with epidemiological expertise.

(3) leveraging the **social network** of researchers in this domain by topical coverage, geographic focus, and research facility locations.

Monitoring the threat posed by avian flu entails tracking the interaction between wild and domestic bird populations, response activities to outbreaks, and research on the transmission, immunity, and disease outcome.

Methods and Materials:

The Health GeoJunction is a highly interactive visual analytic web portal, that extracts information from, integrates, and geographically contextualizes scientific literature, public health reports, and event data to enable situation assessment and surveillance for infectious disease and related threats.

Data sources supported by GeoJunction:

- weekly disease reports from World Organization for Animal Health (OIE)
- scientific publication abstracts from the PubMed library
- RSS (Really Simple Syndication) news feeds from the World Health Organization (WHO)

Geographic and thematic entity extraction:

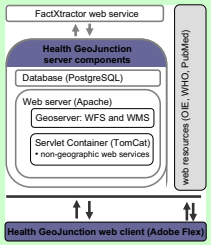
- abstracts returned from a PubMed query on 'avian flu' are processed using FactXtractor (Pan and Mitra, 2007) to identify keywords and geographic entities
- resulting data are made accessible as web feature services (WFS) and web map services (WMS)

Geovisualization of spatiotemporal data GeoJunction supports:

- queries conceptualized as having place, time, and attribute components (Andrienko et al., 2003)
- multiple coordinated views that include an interactive map, timeline, and linked tag clouds

Implementation:

- GeoJunction is a client-server application developed for the Flash player web browser plug-in using Adobe Flex™ featuring a port of Open Layers to Flex to support Open Geospatial Consortium Standard WMS & WFS.



Results:

1 What are the global temporal, geographic, and attribute components of Avian Flu events and science?

temporal (constraint) — *spatial (result)* — *thematic (result)*

Query by Time

Overview Map

Country Details

Keyword Tag Clouds

Sequence (1) shows a stacked time line filtered for a ten month period resulting in a graduated symbol map of disease events and publications and a tag cloud of key terms for these articles. Conceptually, this is an example of constraining the temporal dimension to view the resulting spatial and thematic dimensions.

Graduated symbol map: The first map shows paired graduated symbols for each country representing the number of avian flu disease incident reports and the number PubMed articles by author location. The second map image shows the result of navigating and zooming to a larger scale and then placing the mouse of a symbol to review a summary of the associated documents.

Filtered Tag Clouds: The upper tag cloud lists the top 100 most common keywords from all abstracts in the PubMed document set about avian flu. More common terms are displayed in a larger font. Those displayed in dark blue exceed a user defined threshold for the number of articles using a given keyword. The lower tag cloud limits results to selected time range.

2 What documents use this concept?

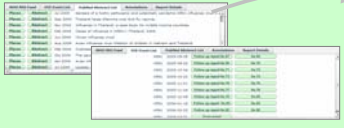
thematic (constraint)

Query by Concept

3 Where are the authors and what places are they writing about?

thematic (result) | thematic (constraint) — *spatial (result)*

Document List



Document Footprint

Document Footprint: Arcs originate at the location affiliated with the author(s) and terminate at the location of place entities extracted from the abstract using FactXtractor (Pan and Mitra, 2007) and georeferenced with GeoNames.org web services.

4 What reports have been issued about Avian Flu for this place? Articles?

spatial (constraint) — *thematic (result) | thematic (constraint)*

Query by Place

Article or Report List

Article or Report

Example queries: The sequences listed above (1-4) are examples of some the ways the data are represented visually and how interacting with these allows the user to iteratively query across temporal, spatial, and attribute (thematic) dimensions.

Conclusions:

A web-based geovisualization interface has been designed to leverage Geographic Information Retrieval and entity extraction technologies by integrating their output with information from other sources made accessible through OGC compliant web services.

The goal is to empower analysts to observe and interrogate geographic, temporal, and thematic relationships embedded in textual information, thus to aid them in the related tasks of situation assessment, situation awareness, and the discovery of research activities relevant to preparedness for infectious disease threats.

Future Research:

Planned research includes support for query by example, annotation of views resulting from queries, and for additional geographic and text data sources. Text processing techniques will be extended to take advantage of the National Library of Medicine's controlled vocabulary thesaurus of Medical Subject Headings (MeSH). Other planned research includes developing additional methods to represent and interact with the complex, interconnected information available as well as usability testing of the results.

Literature Cited:

Pan, C.-C. and Mitra, P., 2007. FemaRepViz: Automatic Extraction and Geo-Temporal Visualization of FEMA National Situation Updates. In: P. Mitra (Editor), Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on, pp. 11-18.

Andrienko, N., Andrienko, G. and Gatalsky, P., 2003. Exploratory spatio-temporal visualization: an analytic review. Journal of Visual Languages and Computing, 14: 503 - 541.

Acknowledgements:

This project was funded through the Regional Visualization and Analytics Centers (RVACs) by the Department of Homeland Security, Science and Technology Directorate, Office of University Programs.

For Further Information:

Contact Michael Stryker (mzs114@psu.edu) for further information. More information on this and related projects is available from http://www.geovista.edu/NEVAC'
An online, PDF-version of the poster can be found at:
http://www.geovista.psu.edu/dl/new_library/library.html