Health GeoJunction: Geovisualization of news and scientific publications to support situation awareness

Michael Stryker Ian Turton Alan MacEachren
Geography Department
Penn State University

GeoVISTA Center

Outline

- Background
  - Problem and research questions
  - Text extraction and visualization to support geospatial temporal data analysis
  - Geovisual Analytic approach to the surveillance of the Avian Flu pandemic threat
- The Health GeoJunction application
  - Conceptual underpinnings and implementation
  - Web portal for visualizing text extraction service output
  - Deriving insight from the space-time-object conceptual framework
- Future objectives: capturing and sharing the analytic process
  - Capturing, sharing, and revisiting insights as contextualized artifacts in the geovisual analytic environment
  - Annotation as a means to capture evolving user perspectives
  - Visual overview and filtering of annotation
  - Booking marking the visual state associated with an insight

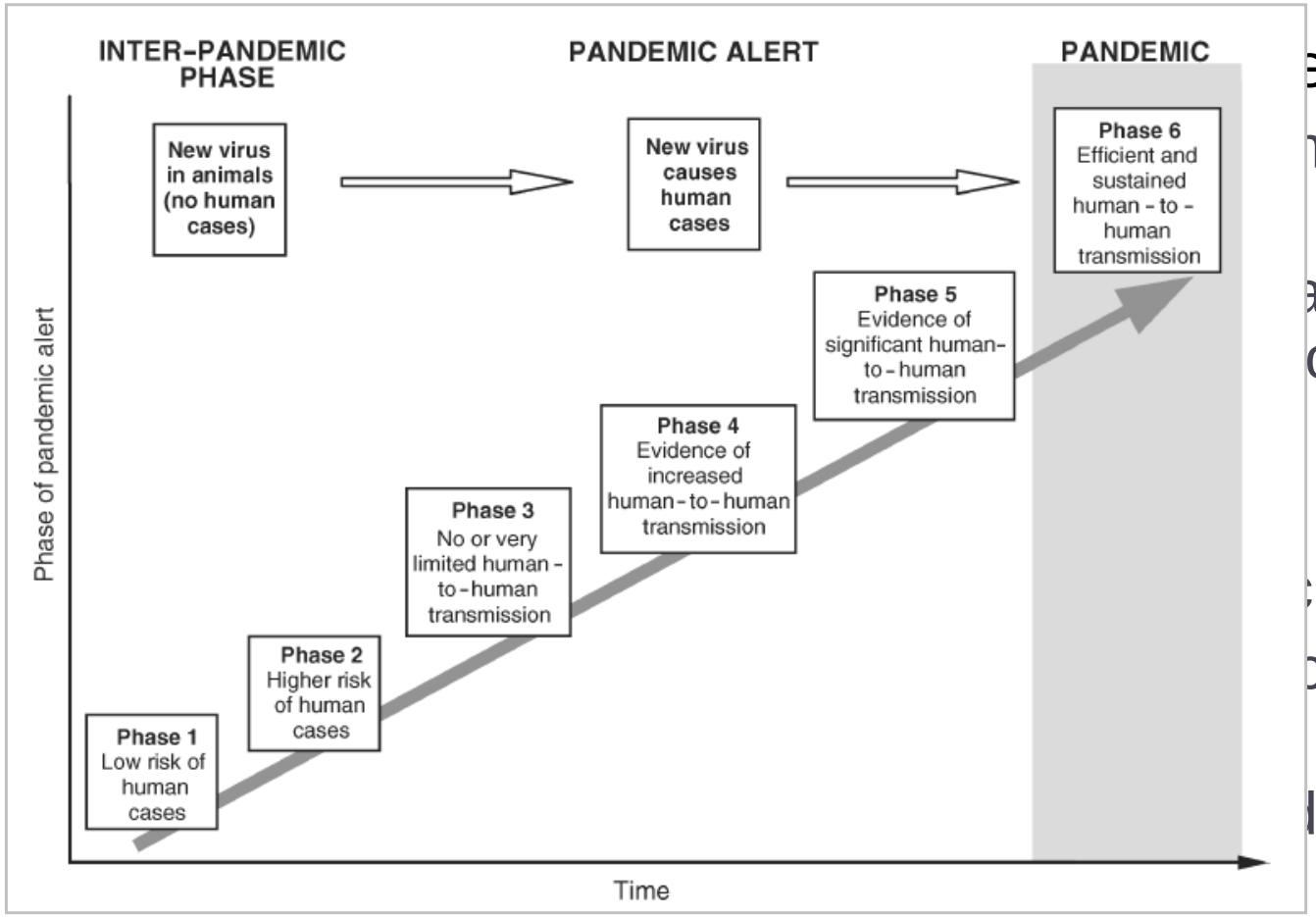
Background: Geovisual Analytic Approach visualizing document collections

Geovisual Analytic approach

- Problem: Large collections of time sensitive reports and a growing body of specialized academic publications requires new tools for identifying relevant information particularly when the geographic context of this content is necessary
- Research questions:
  - How can you design a scaleable web service based and highly interactive environment for searching documents based on geographic content in addition to temporal and key word criteria?
  - Approach: text extraction services to populate a data store of entities accessible through OGC compliant web services and computational tools for determining document relevance
  - What are effective visual interfaces for understanding the content of large document collections from entities extracted from this text?
  - Approach: provide coordinated views of document set conforming to the conceptual dimensions of place, time, and attribute (keyword)
  - How can you capture this process so that analysts can build upon prior insights derived within the interface and share these contributions in a collaborative fashion to build upon this knowledge over time?
  - Approach: enable analysts to annotate the visual analytic environment and discover and search these annotation stores

Avian Influenza surveillance: Data and Tasks

- WHO protocol for monitoring zoonotic infectious disease threats provides a model for processing events



- eness'
- nt within a given
- at hand with a
  ocesses at work
- journals
  ction of incident
- lication) feed

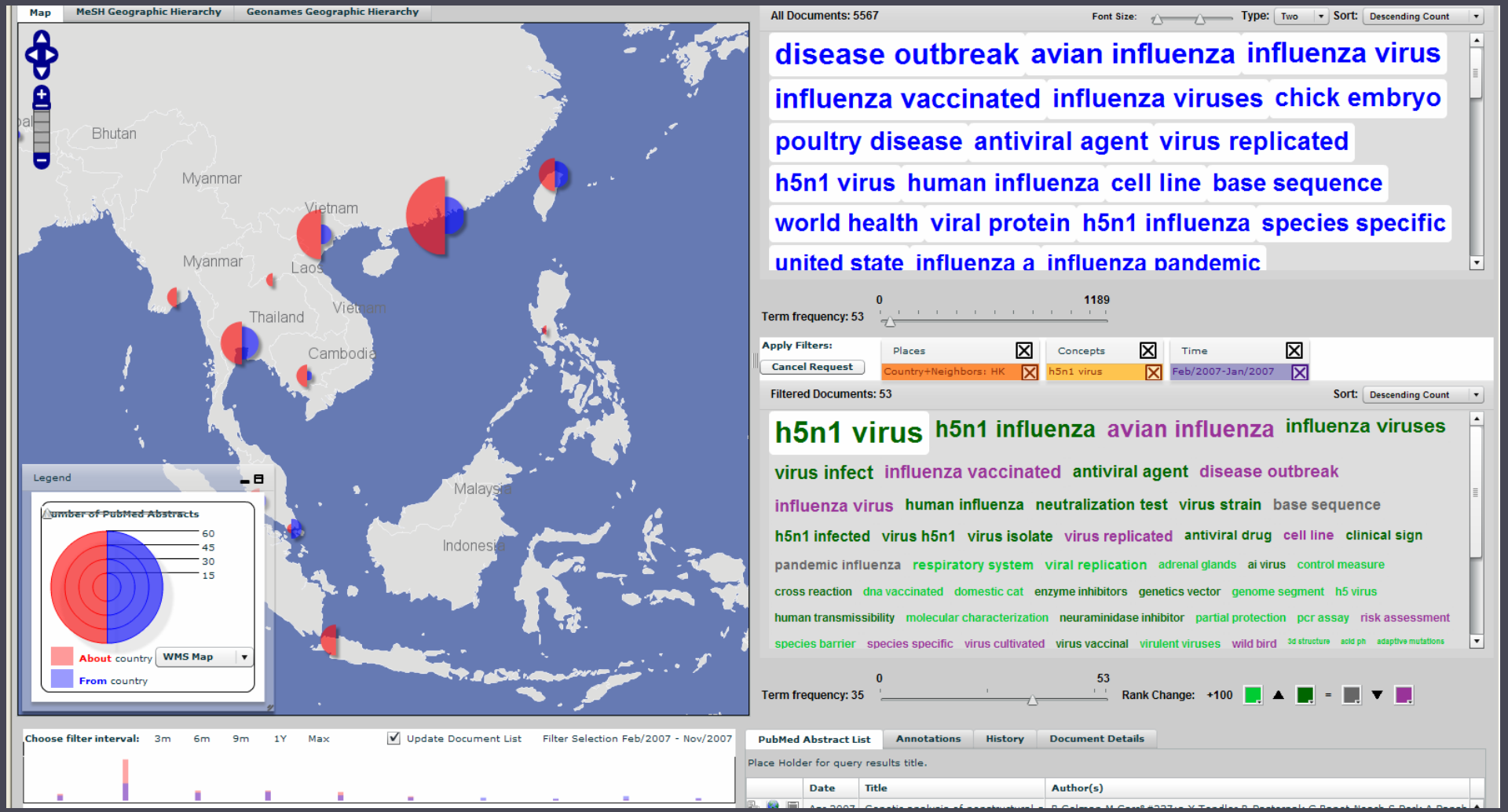
Graphic from Smallman-Raynor, M. and A. D. Cliff (2008)

Health GeoJunction

Geovisual Analytic Web Application

Health GeoJunction Interface

The Health GeoJunction interface provides multiple coordinated views organized around the conceptual dimensions of place, concepts, and time.



Health GeoJunction:

Geovisual Analytic approach → Computational tools + Highly Interactive Visual interface

Server Functionality

- Data sources:
  - PubMed abstracts
  - OIE situation reports
  - WHO Avian Flu bulletins
- Text processing (FactXtractor)
  - Extracting keywords
  - Extracting, disambiguating, and geocoding geographic entities
  - Storing entities and bibliographic metadata in a PostGres database
- Web services:
  - Geographic queries via Geoserver
  - non-spatial queries
  - WMS reference map

User Interface:

Developed with Adobe® Flex® technologies for the Flash® player web browser extension

- Map

  Port of OSGeo OpenLayers to ActionScript3
- Hierarchical gazetteer
- Comparative Tag clouds
- Time series
- Document result set
- User annotation

Health GeoJunction:

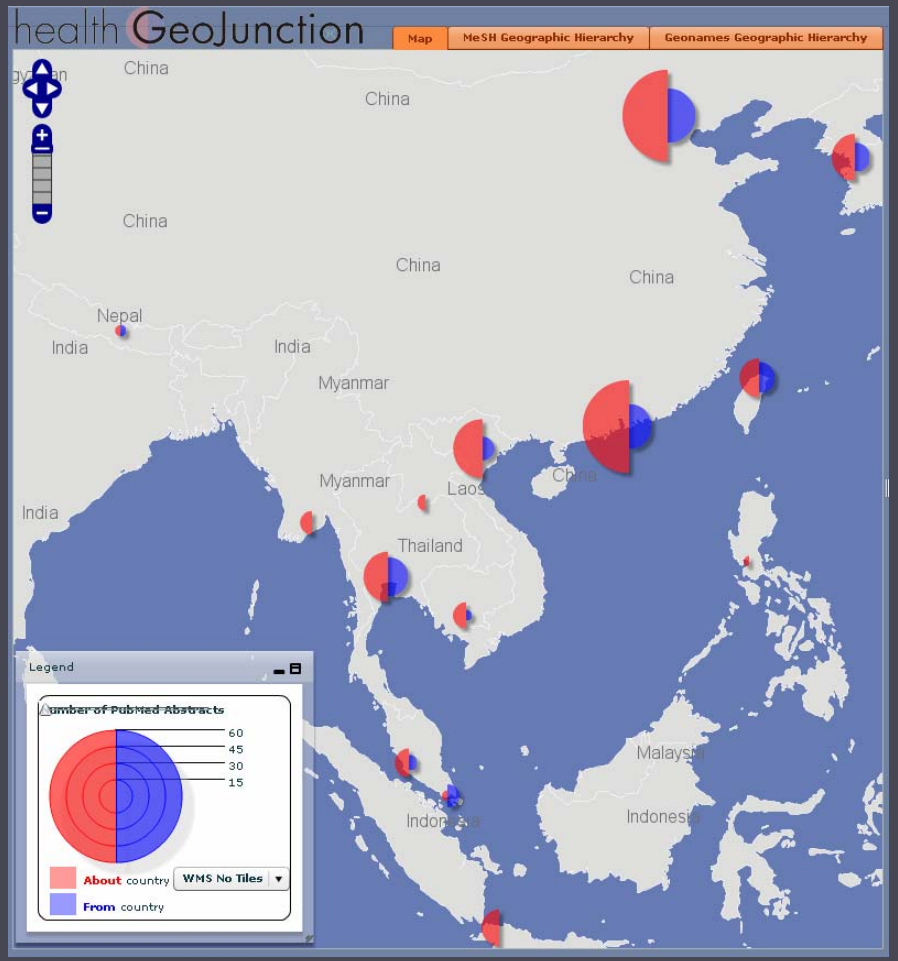
Geovisual Analytic approach → Computational tools + Highly Interactive Visual interface

Primary functionality

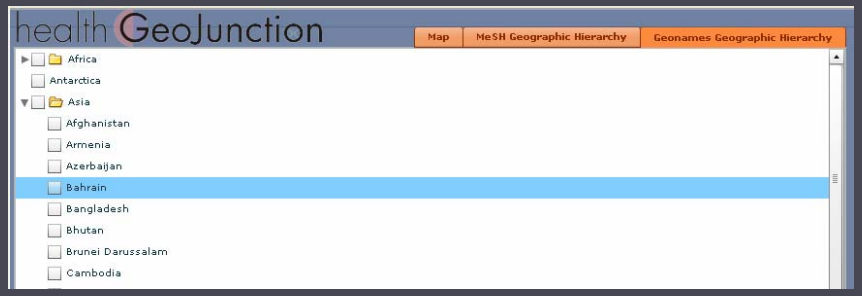
- Space-time-concept query:
  - Space: ‘about’, ‘from’, within country, within neighboring countries
  - Time: user defined intervals
  - Concept:
    - MeSH (Medical Subject Headings)
    - Automated tagging: one term, two term, Yahoo tag service
  - Facet-like tracking of query parameters
- Overview + detail comparative tag clouds
  - ‘Overview’: tag cloud for full document set
  - ‘Details’: tag cloud for filtered set
  - Rank comparison, sorting, highlighting by frequency
- Query-by-example:
  - Return related documents ordered by relevance (Lucene indexing technology)
- Annotate
  - Insert graphics within time series referencing short text entries

Client Functionality

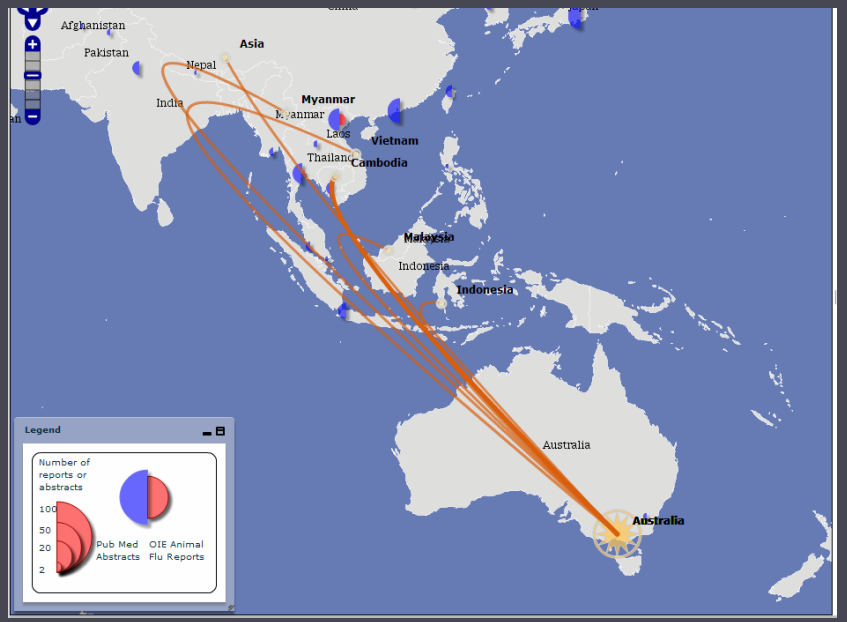
Map and Hierarchical Gazetteer



Overview map



Gazetteer: hierarchal place name list



Geographic ‘footprint’ map

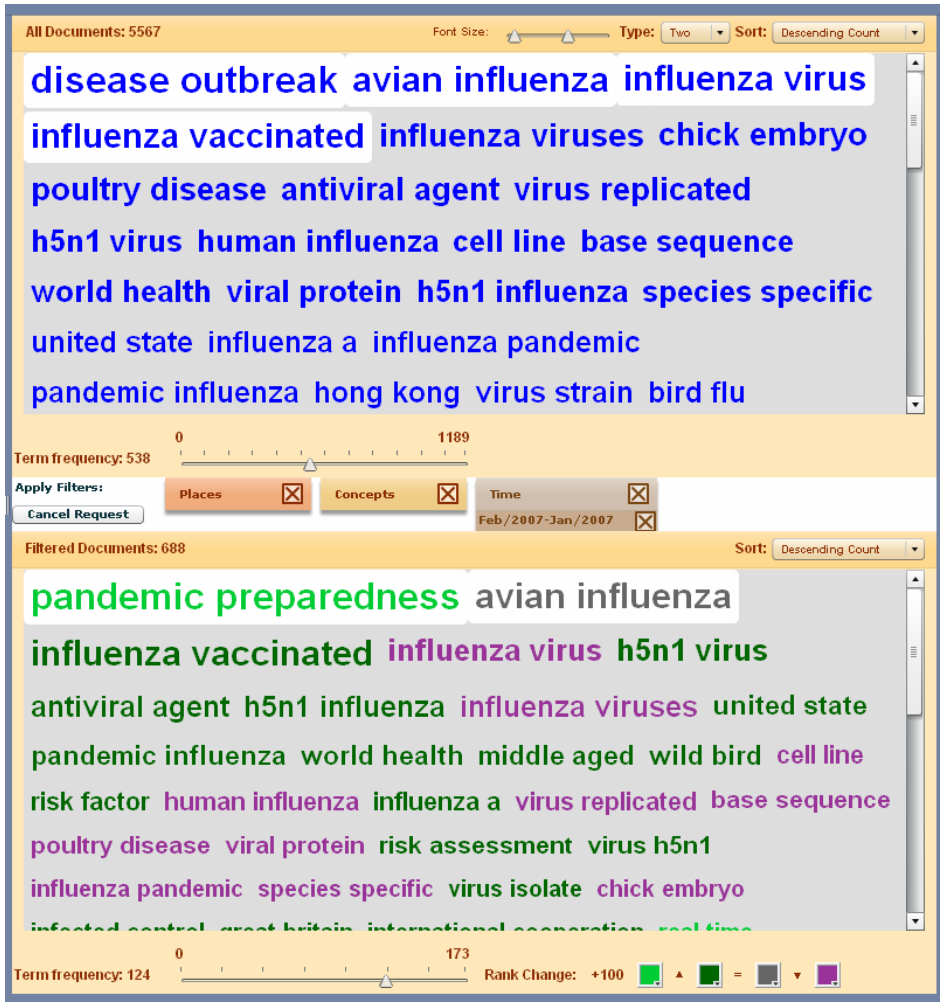
Comparative Tag Clouds and Facet-like Query

Tag Cloud

- All documents vs. filtered set
- Dynamic font size weighting to handle large and small variations in term frequency
- Term highlighting by frequency threshold
- Color coded categories for the change in rank order of terms
- Progressive filtering document set by selection of terms in either overview or filtered tag cloud

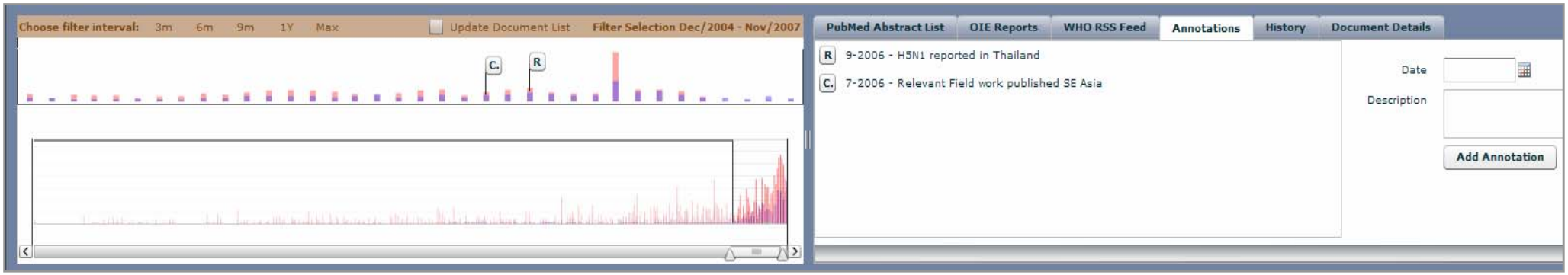
Facet-like Panels

- Concept
- Place
- Time



Time series and Annotations

- Preset Filter intervals
- User defined interval
- Stacked time series: Overview+detail
- Mouse-over details
- Paired bar chart of document counts with ‘from’ and ‘about’ locations
- Time series annotation

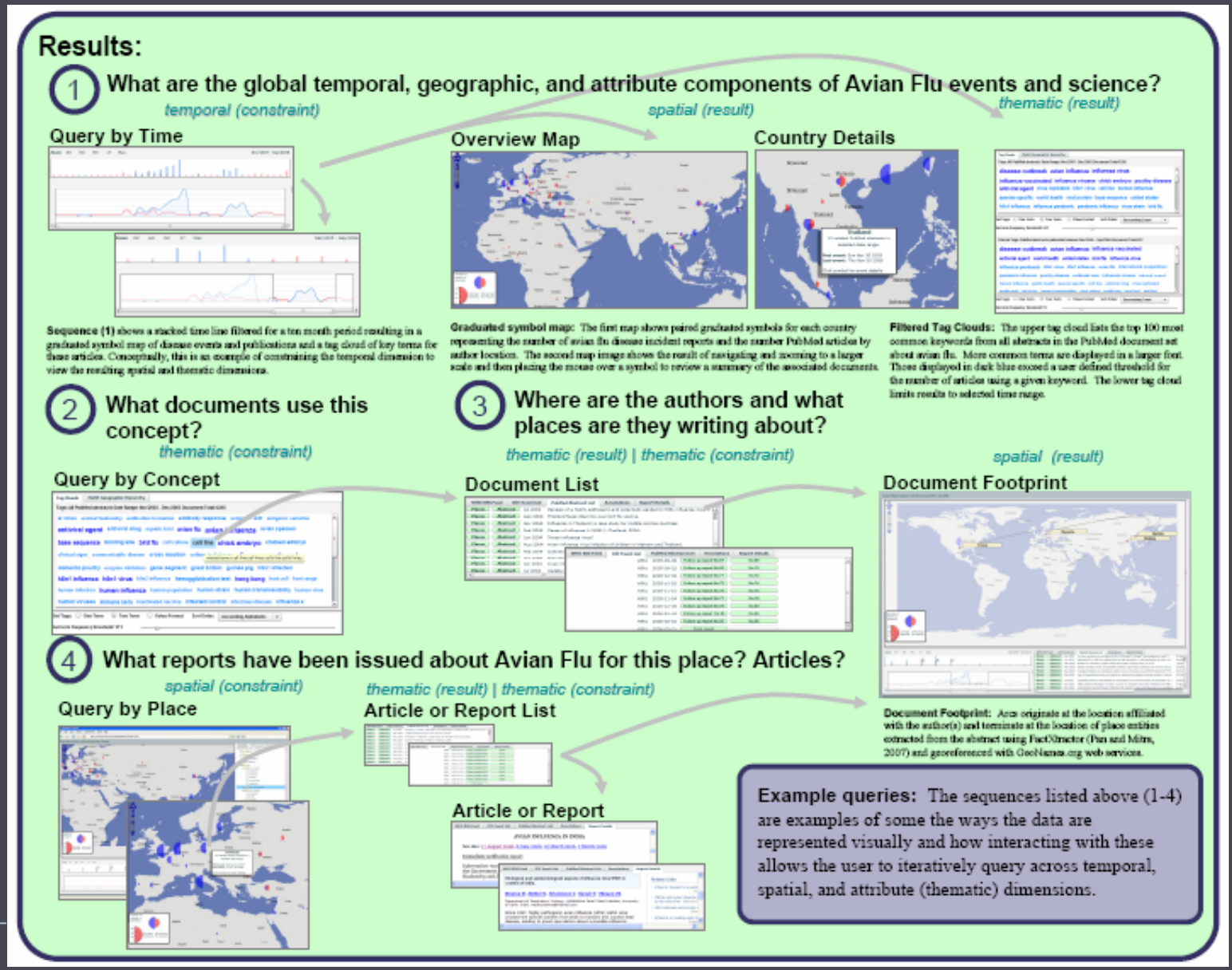


Health GeoJunction: Coordinated views to support space-time-concept exploration

Temporal Filtering

Concept Filtering

Location Filtering



Work in progress

- Structured annotation
- Visual overview and filtering
- Managing histories of user interaction

Visualizing, structuring, and managing geographic annotation

- Annotating a dynamic, interactive geoanalytic interface
  - What does the user need to do?
  - How should he or she be able to do these things?
  - How will he or she know these actions are possible?
- Making sense of annotations: searching and summarizing
  - How can relevant annotations be discovered and made available? (directed searches and serendipitous discoveries)
  - How can an overview of a subset of annotations be incorporated while still providing a usable interface?
- Given these needs, how do you structure and manage annotation contributions?
  - Argumentation model
  - Related work in library sciences

Visualizing Annotation

- Geographic context of individual contributions
  - Point, line, area geographic references
  - References to multiple places
  - References at multiple geographic scales
  - References to vague geographic regions
  - Incorporating localization for places
- Capturing issues and concepts within a given geographic context
  - Argumentation maps (Rinner, 2001)
  - Argumentation model: four types of contributions

    (1) Questions

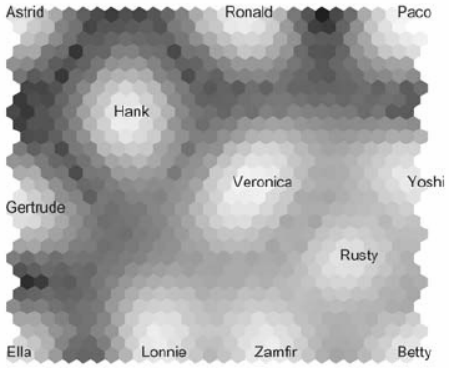
    (2) Solution/Idea

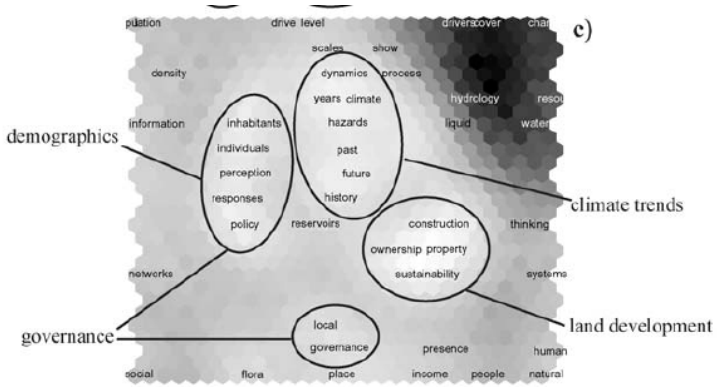
    (3) Evidence or values in support

    (4) Evidence or values in opposition

Visualizing Annotation (continued)

- Visual summaries (computational approaches for grouping and reducing the number of elements)
  - Self Organizing Map (SOM) to visualize clusters of concepts or individuals by similarity of contributions (Pike and Gahegan, 2003)
  - Reducing the number of elements to display
    - Filtering by relevance (using Lucene indexing technology for customized ‘more like this’ searches)
  - Hinting at elements as the user navigates to an area or happens upon similar search criteria





Summary

- Health GeoJunction is a webportal that provides a visual analytic interface to web services that allow the exploration of entities extracted from a collection of documents

- Health GeoJunction core features include:
  - Text processing of document collections to populate a data store accessible through OGC compliant webservices
  - Geographically focused search through coordinated views that support space-time-concept queries
  - Overview+detail document collection view through paired tag clouds
  - Query-by-example provides related documents ordered by relevance

- An initial implementation of georeferenced annotations are proposed as an approach for capturing, managing, and visualizing user insight and facilitating collaboration within a visual analytic environment

Acknowledgements

This work is supported by the National Visualization and Analytics Center, a U.S. Department of Homeland Security program operated by the Pacific Northwest National Laboratory (PNNL). PNNL is a U.S. Department of Energy Office of Science laboratory.

Scott Pezanowski provided actionscript library for OpenLayers functionality in Flex.

Thank you for you time.

Comments and questions are welcome.

For more information:

Contact: stryker@psu.edu

Visit: http://www.geovista.psu.edu/

References

Endsley, M. R. (1995). "Toward a Theory of Situation Awareness in Dynamic Systems." Human Factors 37(1): 32-64.

MacEachren, A.M., 2005. Moving Geovisualization toward Support for Group Work. In: J. Dykes, A.M. MacEachren and M.-J. Kraak (Editors), Exploring Geovisualization. Elsevier Ltd., Kidlington, Oxford, pp. 445-461.

Pan, Chi-Chun, Mitra, Prasenjit. , October 30, 2007. 'FemaRepViz: Automatic Extraction and Geo-Temporal Visualization of FEMA National Situation Updates', IEEE Symposium on Visual Analytics Science and Technology (VAST 2007), Sacramento, CA.

Pike, W. and M. Gahegan (2003). Constructing Semantically Scalable Cognitive Spaces. COSIT 2003: Conference on Spatial Information Theory, Springer Verlag.

Rinner, C. (2001). "Argumentation maps: GIS-based discussion support for on-line planning." Environment and Planning B: Planning and Design 28: 847-863.

Smallman-Raynor, M. and A. D. Cliff (2008). "The Geographical Spread of Avian Influenza A (H5N1): Panzootic Transmission (December 2003ƒ??May 2006), Pandemic Potential, and Implications." Annals of the Association of American Geographers 98(3): 553 - 582.

contact: stryker@psu.edu