Using Text Analysis to Reduce Information Overload in Pandemic Influenza Planning

James E. Powell, Presenting Author

Los Alamos National Laboratory

Linn Marks Collins, Jorge H. Roman, Mark L. B. Martinez, Ketan K. Mane, Xiang Yao, A. Shelly Spearing, Miriam E. Blake

Los Alamos National Laboratory

Geoffrey Hoare, Rhonda White

Florida Department of Health

The proliferation of plans can result in debilitating information overload in public health and medical emergencies. In the case of pandemic influenza, the US Department of Health and Human Services (HHS) and its Centers for Disease Control and Prevention (CDC) have pan flu plans for coordinating the 50 states and each of the 50 states has its own pan flu plan. Plans need to be analyzed, compared, and revised so that they are in alignment with one another. Human analysis of plans is time-consuming and difficult, so text analysis software tools are needed that can help humans (a) compare plans to find gaps or discrepancies and (b) locate relevant sections of plans and display links to them. This research-in-progress describes two text analysis tools being developed at the Los Alamos National Laboratory as part of E-SOS (Emergency Situation Overview and Synthesis): the Theme Awareness Tool (THEMAT) and Content Awareness Tool (CAT). Both tools were used to analyze pan flu plans from the White House, the US Department of Health and Human Services, the Centers for Disease Control, and the 50 states.

Using Text Analysis to Reduce Information Overload in Pan Flu Planning and Response

Proceedings of the 6th International ISCRAM Conference – Gothenburg, Sweden, May 2009
J. Landgren and S. Jul, eds.

Geoffrey Hoare, Rhonda White, Jack Pittman
Florida Department of Health

**Linn Marks Collins, Jorge H. Roman, James E. Powell,
Mark L. B. Martinez, Ketan K. Mane, Xiang Yao,
A. Shelly Spearing, Miriam E. Blake**
Los Alamos National Laboratory

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF HEALTH

Los Alamos
NATIONAL LABORATORY
EST.1943

Problem – Information Overload

- **The proliferation of plans, expert opinion and changing situational status data can result in debilitating information overload in public health and medical emergencies, such as hurricanes or an influenza pandemic.**

- **Pan Flu Example**
  - Florida has a set of state-level Pan Flu response plans. Also, each of its 67 counties has a related set of Pan Flu plans.
  - Similarly the federal Department of Health and Human Services (HHS) and its Centers for Disease Control and Prevention (CDC) have Pan Flu plans intended to coordinate the 50 states' responses – with each state using their own plans.
  - Each one of these hundreds of plans and related implementing materials is often more hundreds of pages long and gets frequently updated.
  - Plans need to be coordinated and subsequently updated.

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF
HEALTH

Los Alamos
NATIONAL LABORATORY
EST. 1943

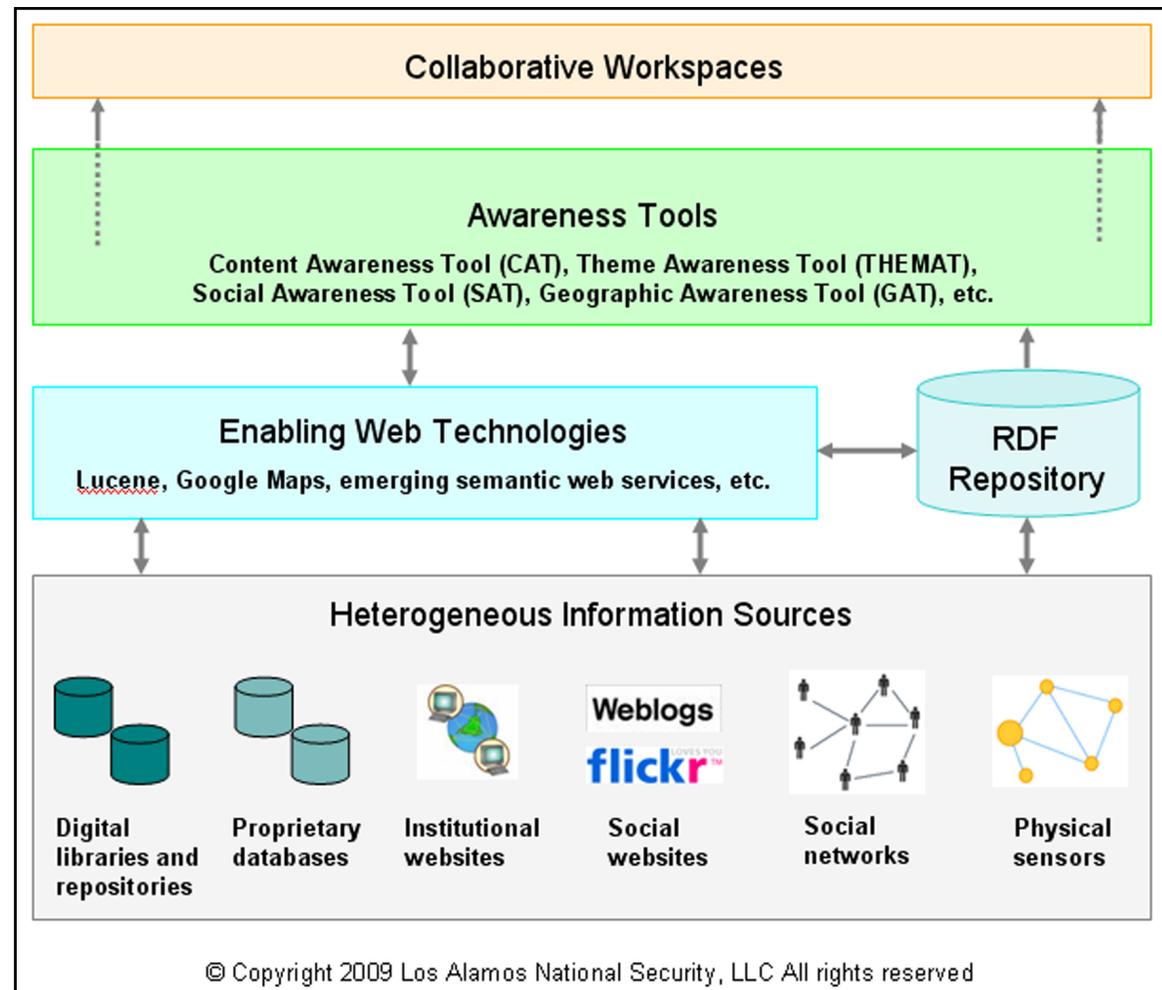# Solutions – Content Analysis and Document Linking

- **Two different tools developed at the Los Alamos National Laboratory are being used to analyze content provided by the Florida Department of Health in order to:**
  - Compare plans to find gaps or discrepancies.
    - It is possible to partially automate this task using content analysis tools that extract key concepts and themes from each plan document.
    - This pattern can then be displayed more intuitively to facilitate comparison.
    - A key question is whether end users can understand the results without the help of an expert analyst and, if not, what additional functionality would make it possible for them to do so.
  - Facilitate locating related information while writing plans, situation reports or public information releases.
    - This capability accesses previously indexed text and displays it to users, based on what they are writing.
    - A key question is how to best organize and index the information for specific writing tasks.

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF HEALTH

Los Alamos
NATIONAL LABORATORY
EST.1943

E-SOS Tools

- **The Emergency Situation Overview and Synthesis (E-SOS) suite of tools is being developed at LANL.**
  - The project started in 2007
  - It builds on several years of prior work by team members
- **E-SOS employs a number of tools.**
  - Collaborative workspaces
  - Awareness tools
  - Semantic web technologies
  - Digital library technologies
- **This project is providing feedback on two of the awareness tools:**
  - THEMAT (Theme Awareness Tool)
  - CAT (Content Awareness Tool)

E-SOS System Overview

- Collaborative workspaces where users can report and discuss information

- "Awareness tools" which display information that's relevant to what users are currently reporting and discussing

- Technologies that synthesize information from heterogeneous sources

Collaborative Workspaces

Awareness Tools
Content Awareness Tool (CAT), Theme Awareness Tool (THEMAT),
Social Awareness Tool (SAT), Geographic Awareness Tool (GAT), etc.

Enabling Web Technologies
Lucene, Google Maps, emerging semantic web services, etc.

RDF Repository

Heterogeneous Information Sources

Digital libraries and repositories

Proprietary databases

Institutional websites

Weblogs
flickr

Social websites

Social networks

Physical sensors

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF
HEALTH

Los Alamos
NATIONAL LABORATORY
EST. 1943

Pan Flu Text Analysis Evaluation

- **Florida Department of Health planners, incident command personnel and public information officers are evaluating the usefulness of information generated by THEMAT and CAT in simulations of text-intensive sub-tasks of three key tasks:**
  - Planning – Revising the interrelated set of state-level Pan Flu Plans based on changes to Federal plans and other States' best practices.
  - Public Information and Risk Communication – Tracking the changing content of RSS feeds from media outlets, blogs and other fast moving information sources to enhance the PIO function in a disaster, including managing communication with the media, risk communication with the public and rumor control.
  - Maintaining Situational Awareness – Managing the large number of information sources needed by the Situation Unit at the state Emergency Operations Center's Health and Medical Emergency Support Function (ESF-8).

Content Used in Pilot Project

- **Pan Flu plans for states:**
  - Florida, New York, California, Minnesota, Rhode Island, Arizona
- **Pan flu plans for cities:**
  - Miami, New York
- **Pan Flu plans for federal agencies:**
  - Health and Human Services (HHS) coordination of the all federal agencies
  - HHS's guidance to states
  - The Centers for Disease Control and Prevention (CDC) Operations Plan (OPLAN)
- **The White House's National Strategy for Pandemic Influenza Implementation Plan**

Steps Using THEMAT

- **Extract themes from each document. *Called knowledge signatures (kSigs).***

- **Create taxonomies for each document or set of documents.**

- **Compare taxonomies to identify unique, overlapping or missing content.**

- **Compute the network analytic relationships among themes.**

- **Generate Theme Network (tNet) visualizations for each document or set of documents.**

THEMAT Results

- Sets of knowledge signatures (kSigs) generated for each document

- kSigs indexed and highlighted for all documents (slide 10)

- Similar themes in Pan Flu plans identified (slide 11)

- Unique themes in Pan Flu plans identified (slide 11)

- Theme networks (tNets) of each plan created to help users focus on key concepts and their relationships (slide 12)

- Comparison of gaps and similar and divergent uses of concepts in different documents

- Trends over time for multiple versions of plans identified

THEMAT Results – Knowledge Signatures (kSigs) for CDC Influenza OPLAN

kSigs for a single document are listed in the left column and highlighted in the right column.

An end user can use this list to jump to the places in the document where the term is used.

The union of kSigs for a collection of documents results in a taxonomy for the set.

THEMAT Results – Similar and Unique Themes for Florida and California Pan Flu Plans

kSigs for collections of documents (i.e. taxonomies) are displayed side-by-side so that similarities and differences can be readily discerned.

THEMAT Results – Theme Network (tNet) Displays for Florida and California Plans
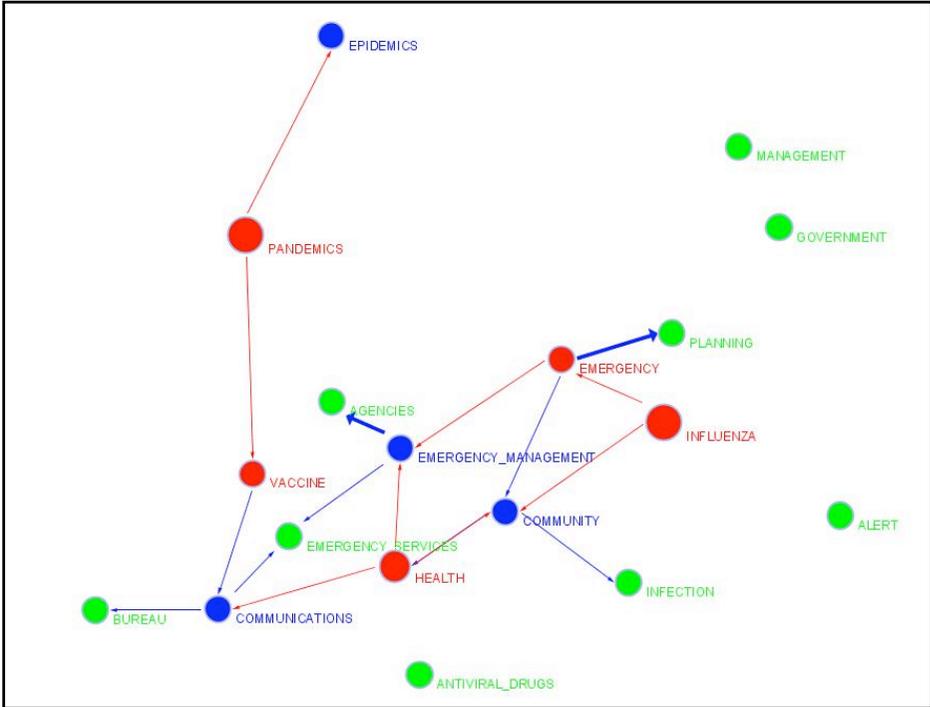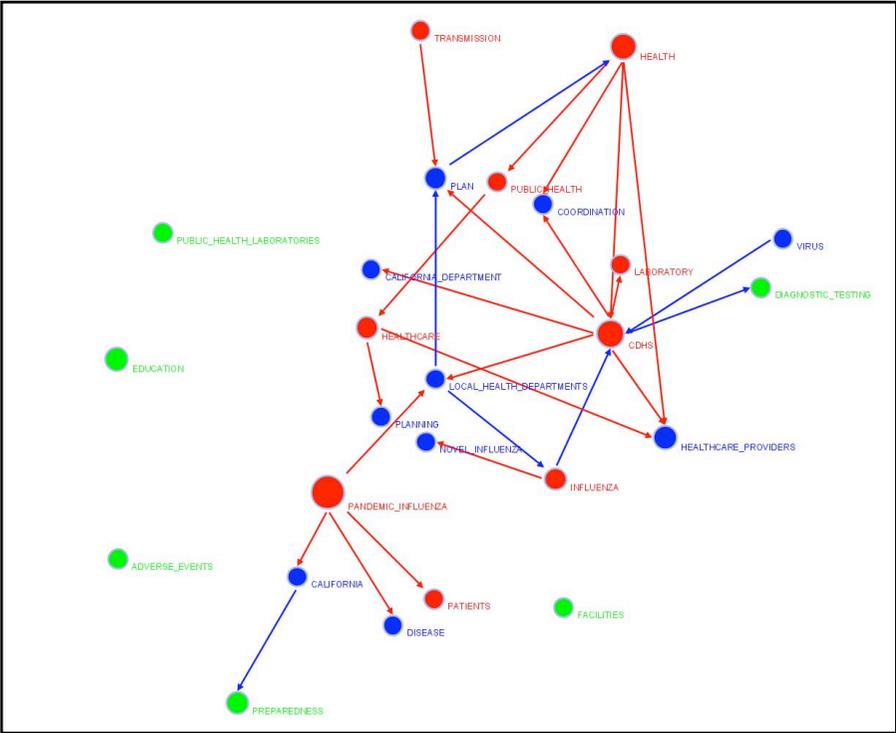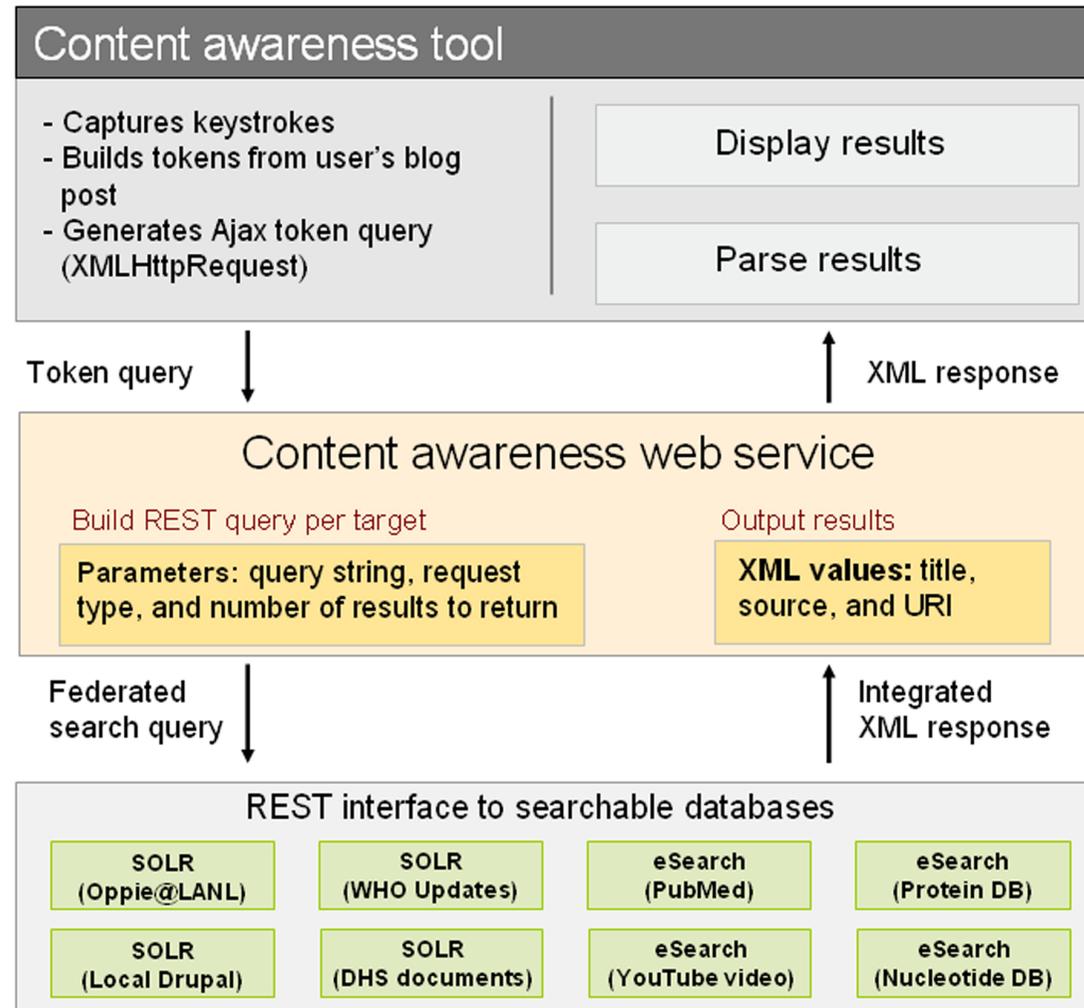
Conclusions about Theme Analysis with THEMAT

- **The THEMAT can be used for some tasks without an expert human-in-the-loop:**
  - Generating and highlighting a set of knowledge signatures (kSigs) for a document helps end users navigate through that document.
  - Displaying a side-by-side comparison of taxonomies for documents helps end users identify unique, overlapping or missing content.
  - Creating theme networks (tNets) allows end users to view and compare the higher-order themes.

- **Fully utilizing the THEMAT's capabilities requires an expert analyst who can use the tool to detect:**
  - Confusion of terminology.
  - Confusion of scale of operations and chain of command needed between federal and state.
  - Lack of actionable instructions such as clear time phased deployment of resources based on measurable trigger events. For example, "Case based treatment of victims with anti-virals will cease when…" "Operations of alternate medical treatment sites will be 'federalized' when…"

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF
HEALTH

Los Alamos
NATIONAL LABORATORY
EST. 1943

Content Awareness Tool (CAT) Test Steps

CAT Components

- Step 1: Create a repository of content at LANL.

- Step 2: Configure the CAT to include this repository as one of the targets of the federated search.

- Step 3: See what links to relevant content the CAT retrieves as users compose text such as situation reports required for incident action plans. (slide 15)

CAT Results – E-SOS Screen Shot

Results of a federated search of (1) one local repository of pandemic flu plans and (2) four websites with pandemic flu content

Conclusions about Decision Making Aided by CAT

- **Creating a repository of content allows the entire repository to become the target of a search.**

- **Creating a collection of smaller, finer-grained repositories is necessary to optimize the federated search capabilities of the CAT.**

- **Under these conditions, the CAT facilitates:**
  - Quickly locating content that is related to the topic the user is writing about.
  - Following links to that content.
  - Creating a list of references for the topic.

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF
HEALTH

Los Alamos
NATIONAL LABORATORY
EST.1943

Conclusions Based on Text Analysis of Pan Flu Plans

[Conclusions will be updated and reviewed as testing proceeds.]

- **The technology is useful.**
  - It makes it possible to more carefully review all the relevant material by targeting key concepts.
  - It yields a set of changes needed to integrate plans.
  - It will be even more useful for information which is less structured and more fluid, such as situational information and public misinformation – rumors.

- **Some of the results are intuitive and can be understood by untrained end users without the help of an expert analyst.**

- **Other results are not intuitive and cannot be understood by untrained end users without the help of an expert analyst or additional functionality.**

Future Work

- **We want to look at less structured information for the public information officers and situation unit staff to analyze.**

- **Situation Unit Data Sources –**
  - "Planning Unit" email box
  - Situation Reports from federal, state and local response partners
  - EM Constellation centralized mission ordering and tracking system
  - Medical provider organizations' information

- **Public Information Unit Data Sources –**
  - Last three sources above
  - Media news releases
  - Blogs and other unauthorized information
  - Reports from local PIOs

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF
HEALTH

Los Alamos
NATIONAL LABORATORY
EST.1943

Future Potential

- **Based on the realization that:**
  - All disasters become primarily health and medical events after the initial impact and
  - No amount of preparation can handle some catastrophic scenarios

- **Discussions and plans of catastrophic incident management have moved from how the response system <u>will support "victims"</u> to how <u>"survivors" will take care of themselves</u>.**

- **So it will be important to:**
  - Use a large number of less structured information sources to understand the emergent organization which is developing

  - Be able to support netcentric management of the response system until the planned for command and control, incident command structure can be re-established.

ISCRAM 2009
Boundary spanning initiatives and new perspectives

FLORIDA DEPARTMENT OF
HEALTH

Los Alamos
NATIONAL LABORATORY
EST.1943