February 2007

VAC Views

A publication of the Visualization and Analytics Centers

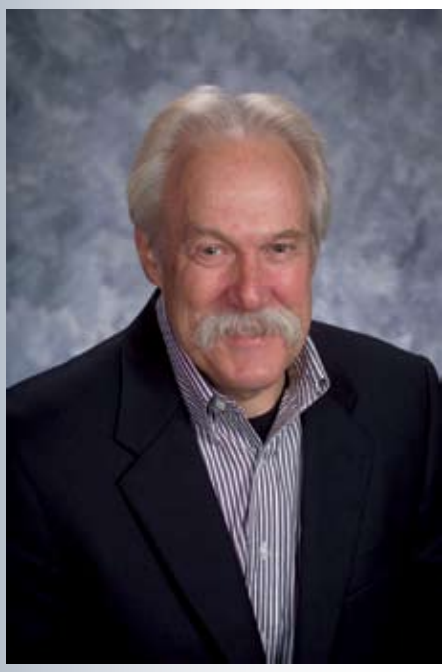
Features:

Disaster Response Life Cycle A true visual analytic environment will aid first responders (page 5)

Threat Stream Generator Synthetic test data helps analytic tool builders perfect their creations (page 10)

Flow Maps Using clues from 19th century cartography to automate flow maps (page 16)

A Message from the Department of Homeland Security



Joe Kielman serves as Science Advisor in the Office of the Under Secretary for Science and Technology for DHS. He is Director of Research Futures for the Command, Control and Interoperability Division and serves as Acting Director for Transition for the Human Factors Division at the Science and Technology Directorate. Before joining DHS, Joe worked for 20 years at the FBI. He has an undergraduate degree in Physics and graduate degrees in Biophysics and did his postdoctoral research in Genetics.

Many years ago, many of us began our involvement with the field of information analytics from the simple yet compelling viewpoint of knowledge management. The metaphors then, as now, were not instructive aids. Rather, they were worrisome specters: *massive information flows, data deluge, overload, connecting the dots* and *mining for nuggets.* Knowledge discovery was the key, and presenting the greatest amount of information to the harried or bewildered user was the challenge.

Visualization, and specifically visual analytics, was then seen as the next best hope to address that challenge. The progress made in the last 15 years has been phenomenal.

The Department of Homeland Security determined early in its history that visualization was a key component in its ability to understand the threats posed by terrorism and the consequences, both immediate and long-term, of natural or manmade disasters. That broad need to ensure the security of our homeland through the application of novel technologies and techniques was the main reason that DHS decided to establish a National Visualization and Analytics Center™.

It has been my privilege to be involved in some of this work and to have played a small part in the establishment of NVAC™ and the development of the associated complex of university centers, other government centers, and industry partnerships. Together, all of these elements constitute an enduring, globally based science and technology as well as a talent base that will enable the U.S. to continue to play a leadership role in the field of visual analytics.

Nonetheless, we cannot be complacent and comfortable with the progress that has been made in visualization science and technology and its outstanding ability to address the problems of information overload and making connections and finding nuggets. Capabilities in these areas must be replaced with new techniques. Intelligence analysis must make way for information or knowledge syntheses, retrospection must be replaced with prospective analysis and understanding of static data must give way to real-time awareness of dynamic, wide-ranging data.

My hope for NVAC, the regional and government centers and the industry consortium members is that they work together to discover and develop new techniques that provide those capabilities. The objective for the VACs is ultimately to seek ways to deliver "designer information"—up-to-date, customized or tailored knowledge delivered just as needed. Visual analytics is the key to this quest.

Joe Kielman
Director, Research Futures
Department of Homeland Security

Current Happenings

2006 was a great year with many exciting events for the VAC community. 2007 promises to be even more active. Here are some of the events that have been brought to our attention.

- **ISI 2007: Intelligence and Security Informatics conference**
  May 23-24, 2007 in New Brunswick, New Jersey, USA
  (http://www.isiconference.org/2007/)

- **IRI-07: IEEE International Conference on Information Reuse and Integration**
  August 13-15, 2007 in Las Vegas, Nevada, USA
  (http://www.sis.pitt.edu/~iri07/)

- **VAST 2007: IEEE Symposium on Visual Analytics Science and Technology 2007**
  October 27 - November 1, 2007 in Sacramento, California, USA
  (http://conferences.computer.org/vast/vast2007/)

In addition to the above conferences, look for these publications' special editions focused on visual analytics:

- *International Journal of Geographical Information Sciences* special issue on **Visualization, Analytics & Spatial Decision Support,** Volume 21, issue 7, will be published online in late July and in print in September bound with issue 8.

- *Information Visualization* (IVS) special issue on **Visual Analytics.** Contributing authors include Joe Kielman, Jim Thomas, Jim Foley, Catherine Plaisant, Stuart Card, Stephen Eick, Daniel Keim and Alan MacEachren (see http://www.palgrave-journals.com/ivs/index.html).

For a more extensive list, visit our website at http://nvac.pnl.gov/events/.

If you know of other events and happenings related to visual analytics, please send a note to:

Richard May
VAC Views Editor and Chief
nvac@pnl.gov

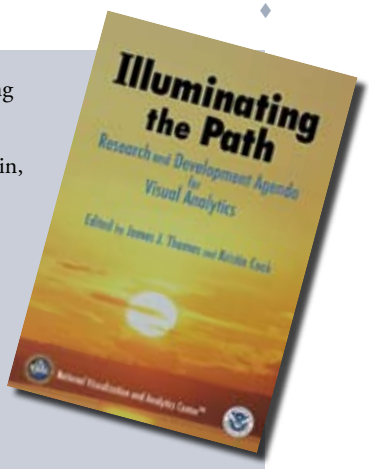
VAC Views is published under the NVAC™ program with funding provided, by the U.S. Department of Homeland Security.

Special thanks to PNNL contributors Carrie Almquist, Sue Chin, Kris Cook, Sharon Eaton, Jill Farris, Jamie Gority, Ian Roberts, David Spiel, Ted Tanasse, Jim Thomas and Barb Wilson.

For more information, visit the National Visualization and Analytics Center™ website at **http://nvac.pnl.gov**

This edition of VAC Views is also available at **http:// nvac.pnl.gov/vacviews**

Illuminating the Path: The Research and Development Agenda for Visual Analytics is available for download at **http:// nvac.pnl.gov/agenda**



Contents

4 Challenges and Viewpoints
4 Mission-Critical Technology Suites
5 Disaster Response Life Cycle

6 VAC News
6 VAC Consortium

8 Research Highlights
8 Aiding Knowledge Discovery
10 Threat Stream Generator
12 Mobile Test Beds
14 Knowledge Integration
16 Flow Maps
18 RimSim

20 Education
20 Digital Library
21 2006 Interns

22 Partnerships

Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

P.O. Box 999, MSIN: K7-10
Richland, WA 99352

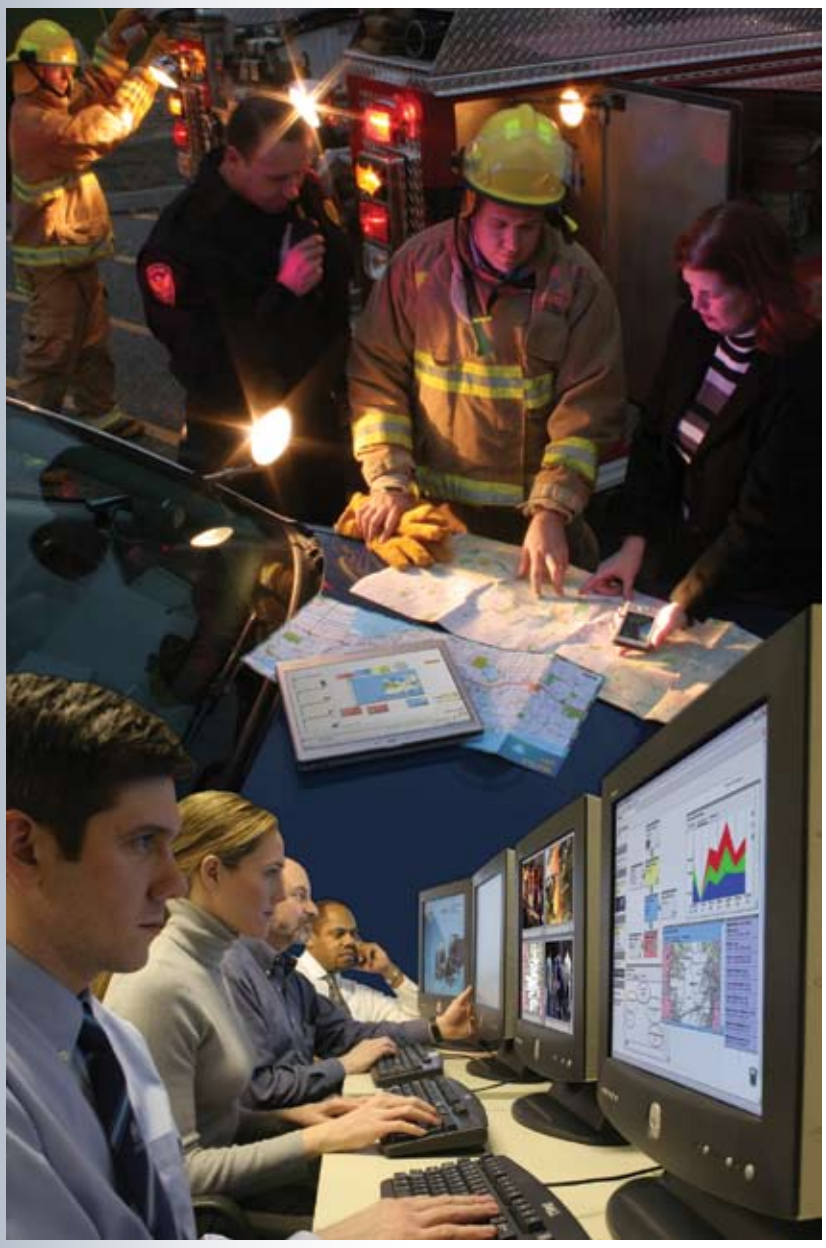
The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the National Visualization and Analytics Center, U.S. Department of Homeland Security, Pacific Northwest National Laboratory, or Department of Energy of any product or service to the exclusion of others that may be suitable.

A New Approach to Mission-Critical Technology Suites

James J. Thomas, National Visualization and Analytics Center

A mission-critical technology suite, in the context of NVAC, is a cluster of technologies from multiple research and development partners, including academia, government and industry. Together these clusters provide a rich solution—that is, a core capability—to address one or more DHS mission needs.

One example of a DHS mission-critical technology suite is a system that collects, categorizes, analyzes and visualizes numerous types of information, including images, sketches, videos and maps. This system would be used by regional investigative units as well as first responders in the field. In addition to aiding analysis and decision making, this system would also include a capability to immediately disseminate information to different audiences, including policy makers and the public. Because such a technology suite requires multiple, often geographically dispersed, R&D partners, a planned, coordinated and integrated approach is needed from development through testing and implementation.



The concept of technology suites is not new. However, the approach most commonly used today is rather haphazard: independent R&D groups produce different technology applications that together may solve part or all of a capability need. These technologies must then be integrated with other technologies that they were never intended to work with. This process is often expensive and the result is usually a complicated product that users find difficult to learn and frustrating to use.

Changing today's approach is one of the goals of NVAC and its partners. Using a planned approach, we wish to develop lightweight technology components that can be rapidly integrated, evaluated and deployed.

How Can You Help?

If you are a homeland security professional, we need you to provide us with:

- your capability needs
- support during test and evaluation
- guidance and feedback about intermediate prototypes.

If you are a technology developer or provider, work with us to change our thinking from developing a single, complete technology platform to developing technology components that can be rapidly integrated into more powerful technology suites.

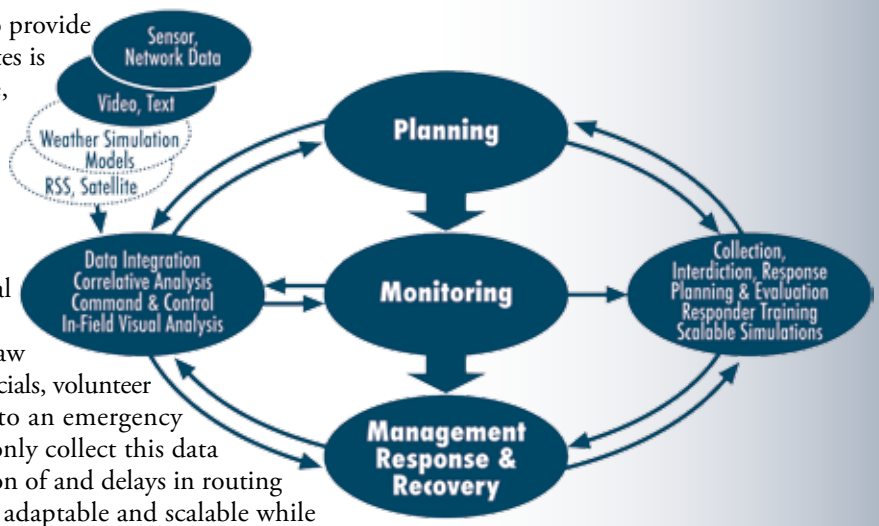
Visual Analytics for the Disaster Response Life Cycle

Dr. David Ebert and Timothy Collins, Purdue University

At the Purdue Regional Visualization and Analytics Center, we are looking to develop visual analytics tools that address the entire disaster response life cycle. Most large-scale emergency operation center systems and tools are designed only to coordinate support and provide basic information for first responders and decision makers during a disaster. The next generation of systems needs to be a true visual analytics environment that supports planning, prediction, detection, mitigation, monitoring, crisis management, response planning and implementation. This comprehensive environment requires integrated visual and traditional systematic analysis of massive data, including improved strategies for exploratory visual analysis of data, hypothesis testing and user-specific presentation of relevant information as a basis for actionable decision making.

We have identified the following challenges to enable actionable decision making:

- **Planning** – for effective planning, accurate simulation models and training systems need to be incorporated, as well as information for event detection, capability assessment and resource allocation. Key components include incorporating an accurate database and surge-capacity modeling of infrastructure and the healthcare system.
- **Monitoring and Detection** – improving systems to provide high detection rates while reducing false positive rates is crucial. Monitoring systems also need to integrate, synthesize and present the relevant information, allowing easy exploration of the evidence and providing the ability to quickly compare and analyze this information.
- **Management** – a basic management component for crisis response is to provide accurate situational awareness. During a crisis, a substantial amount of data is generated from many disparate sources. Law enforcement, fire, medical services, media, public officials, volunteer organizations and many others feed information to an emergency operations center. Current management systems only collect this data and distribute it in one dimension. Misinterpretation of and delays in routing information are common. Future systems must be adaptable and scalable while integrating all information to provide real-time temporal and geospatial awareness.
- **Response & Recovery** – most current systems only extend to the common operation picture/situational awareness capability. For effective response and recovery, resource awareness and resource management capabilities must be included. Integrating analysis and simulation capabilities will allow effective hypothesis testing, scenario analysis and response plan evaluation.



Integration of information and simulations into a visual analytic environment for the disaster response life cycle.

Several key factors crosscut all of the above activities to ensure effectiveness:

- effective data integration
- effective presentation of relevant, necessary information
- adaptation of the visual analytics environment to users and their tasks
- appropriate temporal updating of information
- security and privacy.

Integrating all these components into a comprehensive visual analytics environment is required to enable time-critical decision making and response.

Fall 2006 Consortium Meeting

More than 120 people from industry, government and academia attended the 2-day VAC Consortium Meeting at the Pacific Northwest National Laboratory in October. The theme, *Discovering the Unexpected*, provided a forum for speakers and demonstrators to show how visual analytics research and technologies are enabling first response.

Rod Hilden, chief security officer for the Seaport Division of the Port of Seattle, addressed "Recovery from Maritime Terrorism" in his keynote address. He highlighted the challenges faced by port authorities dealing with large volumes of cargo and complex logistics associated with port facilities.

Hilden is responsible for the security of all Seaport Division assets and facilities including four major container terminals producing two million 20-foot equivalent units (TEUs) annually, two cruise ship terminals handling 750,000 passengers a year, a grain terminal, a major break-bulk cargo facility and five marinas. Today, the combined ports of Seattle and Tacoma comprise one of our nation's top three largest container load centers.

In his speech, Hilden stressed that new security solutions, such as cargo seals and sensor technologies (e.g., radiation portal monitors), are needed to complement current physical technologies. He discussed new security initiatives and highlighted the need for future technological solutions. Hilden emphasized that visual analytics is a critical component for ensuring safer ports.

"What PNNL and NVAC are doing to advance visual analytics is nothing short of extraordinary for applications ranging from national security to business intelligence."

– Peter Kuper, Morgan Stanley

Meeting Highlights

On the first day of the meeting, participants showcased their technologies during a demonstration session. NVAC projects, companies including i2, Intelligent Results, Objectivity, Inc. and Tableau Software, together with the Regional Visualization and Analytic Centers participated. Attendees explored and interacted with the different exhibits and learned about current abilities while networking. The IN-SPIRE™ Assessment Wall was also on display throughout the event.

Member demonstration session.

On the second day, Peter Kuper, an internationally recognized industry analyst from Morgan Stanley, gave a delightful speech, "Seeing the Future; It's Closer than it Looks." Kuper stressed that while homeland security mission objectives drive the research agenda, the resulting products and services could have a valuable role in the evolving marketplace known as business intelligence. Kuper provided a fast-paced, entertaining view of commercial software, quickly drilling into the changes under way, the direction of the business intelligence market and the opportunities available for visual analytics technologies.

The meeting concluded with four panel sessions: First Response, Power Grid, Mobile Analytics and Information Discovery. The First Response panel focused on technological needs and requirements and included representatives from law enforcement, the U.S. Coast Guard, public health and emergency management. The discussion centered on the need for technologies that provide ample, real-time situational awareness that require little or no training and that prove useful for day-to-day operations.

During the Power Grid panel session, academic and industry researchers provided their perspectives on using visualization technology to address power grid reliability and security applications. They discussed emerging technology that will increasingly empower grid operators to enhance their situational awareness of electric power grid infrastructure operations.

The Mobile Analytics panel comprised four presenters who discussed research challenges, usability case studies, the future of cellular infrastructure and mobile security developments.

The discussion period following the presentations was lively with different opportunities and issues pertaining to the future use of mobile appliances being raised.

The Information Discovery panel comprised four representatives from banking, legal and software development. They discussed the need for methodologies and software that allow for quickly searching multiple databases up to petabyte sizes. The focus of the session addressed three topics: We know what we know; We know what we don't know; We don't know what we don't know.

Check out our website for Spring 2007 meeting details!

Member Privileges

Consortium members, including the RVACs and industrial partners, experienced the privileges of membership throughout the event. They were honored at a dinner where they received a plaque for their partnership with NVAC.

David Moore, technical director of the Office of the NSA/CSS Senior Intelligence Authority, facilitated a workshop for members titled "An Introduction to Critical Thinking." The fun, interactive 2-1/2 hour session, based on Moore's latest monograph, *Critical Thinking and Intelligence Analysis*, engaged 75 individuals in small groups. The session goal was to gain a deeper insight into the critical thinking process using exercises to reveal the nature of biases and assumptions and to reflect on intellectual standards for sound reasoning. Participants refined their analytical skills while gaining exposure to the kind of critical thinking training often undertaken by analysts who are the intended users of visualization tools. Moore helped participants to understand the role of critical thinking in exposing alternatives, improving reasoning and communicating effectively.

For more information, visit http://nvac.pnl.gov/consortium.stm.

Point of Contact

Jim Thomas, NVAC Director, Pacific Northwest National Laboratory, nvac.consortium@pnl.gov

"NVAC meetings offer a stimulating infusion of viewpoints on critical data analysis challenges from leading thinkers and solution developers in government, industry and universities. An important step toward moving from ideas to effective action."

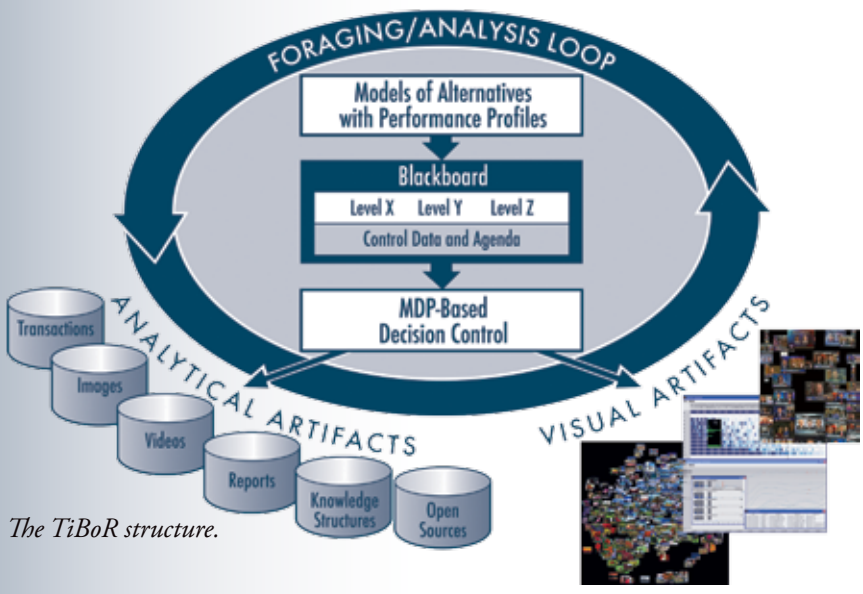
– David Lantrip, Intelligent Results

Consortium members representing: Attenex, Boeing, i2 ChoicePoint, GCS Research, Intelligent Results, Microsoft, Objectivity, Inc., OSIsoft, Tableau Software, University of Massachusetts Lowell.

Visual Analytics Research to Aid Knowledge Discovery

Visual analytics research under way in the SouthEast RVAC will play a key role in helping analysts and industry detect fraud, discover truth and make decisions. SRVAC, a collaboration between the University of North Carolina at Charlotte and Georgia Institute of Technology, is focused on developing visual analytics capabilities to improve and promote analytic discourse. Although SRVAC's primary applications are intelligence analysis and emergency response, the research is applicable to any problems involving complex reasoning, timely decision making and large amounts of data.

TiBoR: A Time-Bounded Reasoning Agent for Information Foraging and Analysis



The TiBoR structure.

Real-world knowledge gathering and investigative tasks are very complex because the problem-solving context is constantly evolving, and the data may be incomplete, unreliable and/or conflicting. SRVAC is developing a mixed-initiative reasoning agent, TiBoR, that will help intelligence analysts choose from and reason about enormous databases of text, imagery, video and webcast.

Recently, the project team added a novel hierarchical blackboard capability to TiBoR. The blackboard is an agent-based tool for following multiple complex-reasoning paths and hypotheses. The hypothesis-building component is grounded on a hierarchically organized task method-knowledge language (TMKL). This approach provides a way to recognize story plots connecting seemingly isolated events. Hypotheses formed provide suggestions for evidence gathering and analysis to TiBoR. Competing hypotheses can be managed in this joint system, and gathered evidence or analyses will cause hypotheses to change in a continuous round of updating and refinement. To integrate the joint system and thereby create a mission-critical technology suite, SRVAC will work with NVAC researchers who are developing a motivation and intent analytical framework to support evidence-based reasoning.

The Picture Becomes Clearer

A significant part of SRVAC's interactive visualization research concerns the exploratory visualization of large, complex, multidimensional data. A major focus is the automated, intelligent analysis and visual exploration of large stores of images or broadcast video. This effort is a true example of visual analytics because it involves a complete integration of visualization and high-end analysis, with each tailored to the other. Repositories of unknown content are automatically analyzed, permitting exploratory visualization, understanding of overall relations among the media and changes over time, and focused selections of images or videos of interest. In the case of video, a true multimedia visual analysis occurs with results fused from image, audio and (where available) closed captions.

While this research has concentrated on the automated identification and analysis of broadcast news, the same visual analysis techniques could be applied to other content types. The ultimate goal is to extend the visual analysis techniques to multimedia content of any sort, especially content from the web.

Visual Analytics Helping to Detect Bank Frauds

SRVAC is collaborating with Bank of America on a completely new visual analytics approach to discovering suspicious activity in bank transactions. Bank analysts' needs and challenges are similar to those of intelligence analysts. Bank analysts often spend a great deal of time looking at suspicious transactions but lack the information and tools to trace the routes of deceit. SRVAC researchers are developing tools that help analysts quickly pinpoint the most suspicious subset of transactions or discover new paths of deceit. These tools can be applied to all types of financial transactions and risk assessment. They will be made available through NVAC.

For more information about the technologies described here and other research being conducted at SRVAC, visit http://www.srvac.uncc.edu.

Team Members

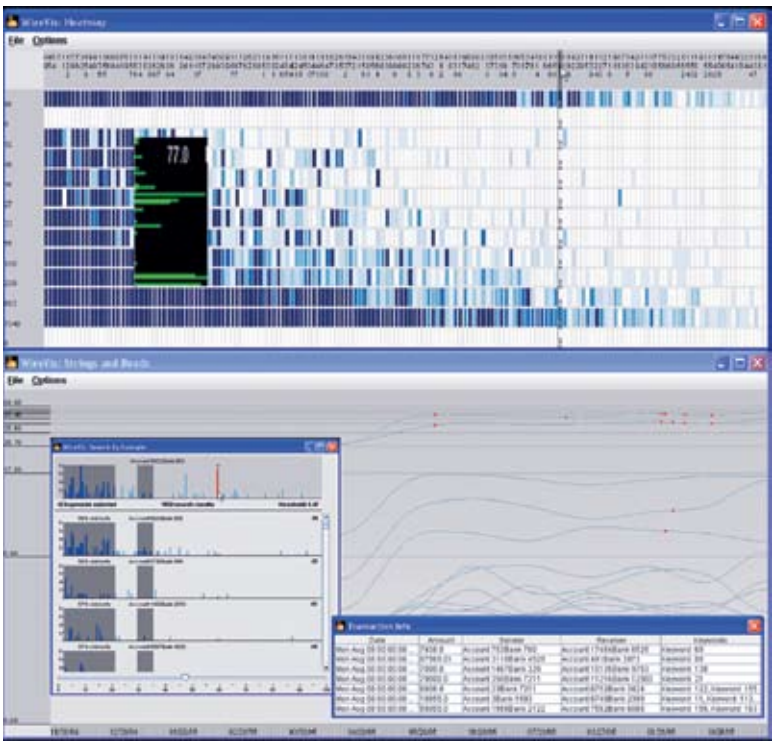
UNC Charlotte: Jing Yang, Robert Kosara, Jianping Fan, William Ribarsky, Anita Raja, William Tolone

Georgia Tech: John Stasko, Jim Foley, Beki Grinter, Ashok Goel

Point of Contact

William Ribarsky, Director, Charlotte Visualization Center, ribarsky@uncc.edu

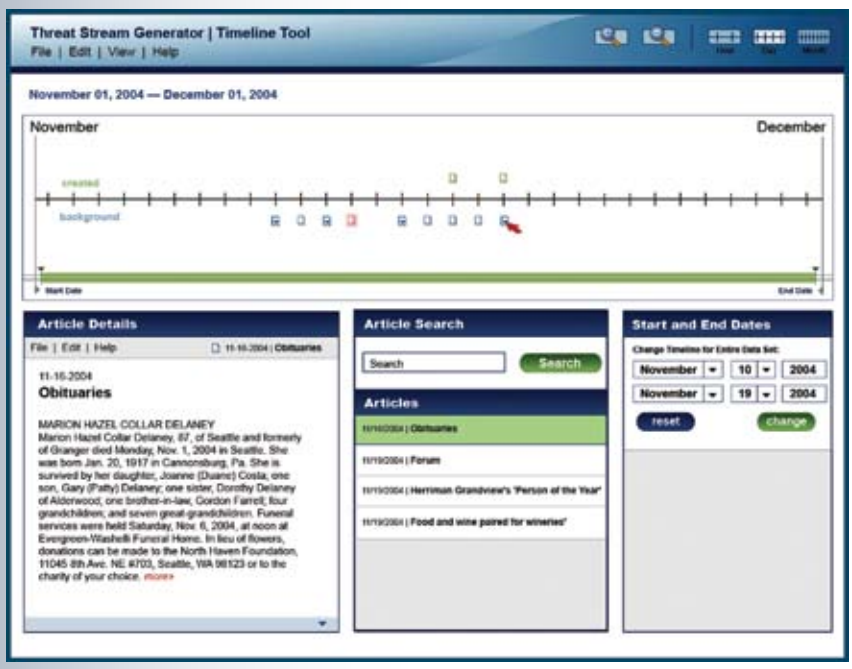
Topics arranged according to "interestingness" for a given time period. Hottest topics (those most reported) appear in the central column. Side columns are used for topics of lesser impact by the interestingness measure.



As the user moves the mouse in the top panel, transaction activity over time and selected keywords are highlighted in the bottom panel. The time period for the bottom panel is one year, and the bright red dots indicate that selected keywords appear in selected transaction clusters. In the lower panel, capability for search by example (left inset) and details for selected transactions (right inset) are also shown.

Threat Stream Generator

Ever wonder if your technology will be able to detect a terrorism threat buried in a multi-source, composite data stream?



The TSG Timeline Tool allows data set creators to manage the time-ordering of data elements in a data set. Elements may be moved individually or in groups to best fit the overall scenario and threat associated with the data.

The NVAC Threat Stream Generator project team has developed a new approach to creating realistic test data sets and has created a series of data sets that are becoming increasingly popular with advanced commercial analytic tool developers, government information analysts and technology developers at universities. TSG data sets are being recognized as the test data sets of choice—with a set of freely available data—allowing users to determine how well their tools perform, because they know what the correct results should be ahead of time.

The distinct advantage to TSG data sets is that known ground truth in the form of a pre-determined threat, such as terrorist activity or a law enforcement concern, is created and translated into data cues. The cues are inserted into the data set with a known expressivity—the number of cues and their subtlety of representation are controlled. "The process is similar to writing a mystery novel," notes project manager Mark Whiting. "The task of finding the threat must intrigue and engage the person analyzing the data, while inducing a suspension of disbelief, since they will know they are working with an invented scenario."

The NVAC TSG project's goals are to understand the characteristics of the data that analysts encounter and to create a tool that will build synthetic data sets mimicking real data to help analytic tool builders test and evaluate their creations.

Putting TSG to the Test

The TSG team recently put their products to the test in an international setting for this year's IEEE symposium on Visual Analytics Science and Technology, held October 31-November 2, 2006 in Baltimore, Maryland. The contest, jointly sponsored by DHS/NVAC, the Office of the Director of National Intelligence/Disruptive Technology Office and National Institute of Standards and Technology, involved industry, academia, government labs and agencies partnered with practicing analysts to evaluate their technologies. The forum consisted of an offline contest, where contestants had five months to prepare their contest submission, and a real-time three-hour analytic challenge at the conference.

Following the offline contest, one participant stated in an e-mail to the team that "the VAST data set is the most realistic one I've seen anywhere." Another contestant wrote, "I have to say that I was very impressed by the quality of the data set…. I am writing to ask you if we could use it for demos of our tools… good data sets are difficult to find but are essential to express the capabilities of visualization software." Several participants expressed similar statements in the contest session at the conference.

In the TSG process, scenarios and threats are invented by a team of intelligence analysts and computer scientists. NVAC software utilities assist the team in identifying key themes and elements of the scenario and to map them into single-format (homogeneous) or multiple-format (heterogeneous) data. These utilities also help the team determine what kind of data set to map the elements into. For example, one cue to the threat might be best expressed in text, while another might be best buried in an Excel spreadsheet of numbers. One utility helps the team manage the sequencing of events in the scenario. If the scenario maintains a

logical ordering, this maintains the credibility of the story. A different team of analysts reviews the data sets after they are created. They play and critique the challenge of finding the threat before the new TSG data set is released to users.

TSG Provides New Opportunities for Test and Evaluation

Finding or creating useful test data sets is difficult for a team focused on creating analytical solutions to intelligence problems. First, real data that might be considered good for testing analytic applications is usually classified; most tool builders will not have the clearances needed to use, or even see, the data. Second, analysts' time is scarce and obtaining the needed characteristics of real data from them to create a test data set is difficult. Finally, generating good test data is challenging. Commercial data generators are focused on large database testing, not information analytics tool testing. Creating a data set based on a threat scenario is a complicated, labor-intensive problem. The TSG effort provides an alternative to organizations that want to avoid these issues. TSG data sets are being used for testing and evaluation of large-scale U.S. government analytic tools. A typical response to the TSG data sets is "this looks like the real thing!"

Several data sets are already available for development groups to test against. The TSG team is ready to help you with your testing requirements with its approach to bringing more realism in testing your innovative solutions before they are placed into operations.

Analysts and tool developers partner to tackle a TSG real-time analytic challenge at the 2006 VAST Conference.

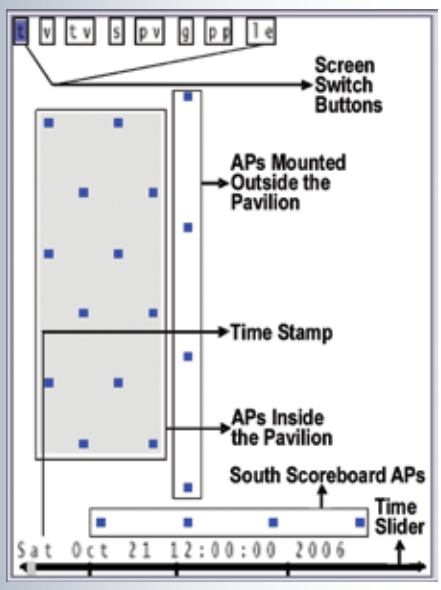
The VAST contest data sets are available at http://www.cs.umd.edu/hcil/VASTcontest06.

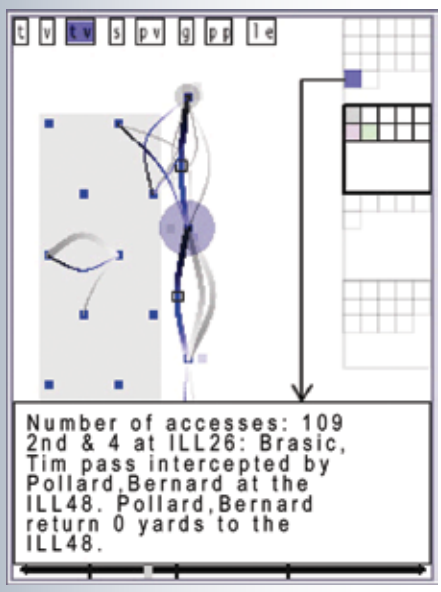
Team Members

Mark A. Whiting, Jereme Haack, Carrie Varley, Wendy Cowley, Doug Love, Stephen Tratz, Pacific Northwest National Laboratory

Point of Contact

Mark A. Whiting, Chief Scientist, Pacific Northwest National Laboratory, mark.a.whiting@pnl.gov





Basic visualization screen setup showing location of wireless access points and the basic user controls (top). The bottom image shows a heavily downloaded video (109 downloads) with text details of the play and distribution of the accesses to that video (circles).

Two Test Beds for Mobile Visual Analytics

Mobile devices can become a vital tool for in-field analytics in numerous applications including emergency response and security. However, while their functionality is increasing, their limitations continue to pose a major challenge in developing effective applications. Getting past these challenges will allow field visual analytics to solve critical mission needs of situational and context awareness.

The Purdue Regional Visualization and Analytics Center is exploring the potential of mobile analytic techniques using two applications as test beds:

- wireless network/sensor mobile analytics in Purdue's Ross-Ade Stadium for the eStadium project
- first responder mobile analytics using the Muscatatuck Urban Training Center.

Both applications effectively display information from streaming data sources to enable in-field decision making. To do so, the PURVAC team first had to overcome the following mobile analytics limitations:

- **Data flow.** Both network bandwidth and device memory are limitations, especially for applications for which real-time streaming data is often one of the inputs.
- **Interface.** Application developers must design for small screens and non-traditional interaction modalities.
- **Display.** Both graphics processing and screen resolution are limited.

Aerial view of Muscatatuc

PURVAC has developed client-server architectures that work within these restraints for both the eStadium and MUTC applications. A powerful, well-connected server receives all of the streaming data and integrates the data into a unified representation. Basic information analysis, extraction and abstraction are performed on the server to determine which data and at what rate it should be sent to the mobile client. The client processes all user interactions and creates the appropriate visual display for the users and their tasks, allowing in-field information analysis and exploration.

eStadium Test Bed

The eStadium project allows football fans to get play-by-play (text and graphical), player and stadium information; game statistics; and streaming video replays via their PDAs or smartphones.

The eStadium team uses PURVAC's mobile visual analytics tools to monitor both the wireless network infrastructure performance and the load on wireless access points. By visualizing the video server log, access log and access point information in real time, network engineers can determine if access point and network settings need to be adjusted. Because streaming video creates the most demand on the system, PURVAC has created tools to 1) analyze streaming video requests per access point during the game and 2) perform a comparative analysis of the download frequency by location and play type.

PURVAC plans to expand the use of this analytical tool to aid in crowd guidance and security operations.

Muscatatuck Urban Training Center Mobile Analytics Test Bed

MUTC was created in July 2005 to train emergency response and homeland defense professionals. MUTC's unique 1000-acre environment allows virtually unlimited exercise scenarios for single units, multiple agencies and joint forces to train in both contemporary urban and rural settings. MUTC is becoming a full-immersion contemporary urban operating environment where organizations can hone their skills and test the systems and concepts under various weather conditions and terrains (field, forest, reservoir, urban). The emergency response and training capability forms our second mobile analytic test bed: creating mobile analytics tools for first responders to use during training exercises at MUTC.

In this project, PURVAC needs to tailor the display capability to the type of responder and his or her role and provide a succinct, quickly understood display of relevant information extracted from all of the information recorded during the exercise. In the case of a SWAT team responding to a gunman in a school, the first and most critical need of the team leader is the most accurate situational awareness possible:

- Where are team members located?
- Where are the locations of responding personnel?
- Where are the secure, neutral and hot zones of the incident?
- What locations provide opportunity or threat information?

ck Urban Traning Center.

The ability to provide information back to the emergency operation center, such as indicating rooms cleared or information contradictory to current situational assessment, is also vital. As previously indicated, relevant information is specialty-dependent. A firefighter responding to a fire at the same building would need some of the same information as the SWAT team but also task-specific information, such as fire spread, and potential reactive measures appropriate for a rapidly changing environment. First-generation mobile analytics could include PDAs and smartphones. Next-generation displays may be incorporated into heads-up displays.

More information about eStadium can be found at: http://estadium.purdue.edu/php_site/web/index.php.

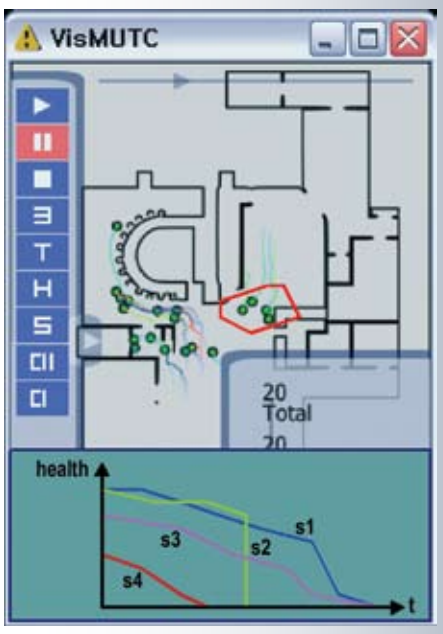
More information about the Muscatatuck Urban Training Center can be found at: http://www.mutc.org.

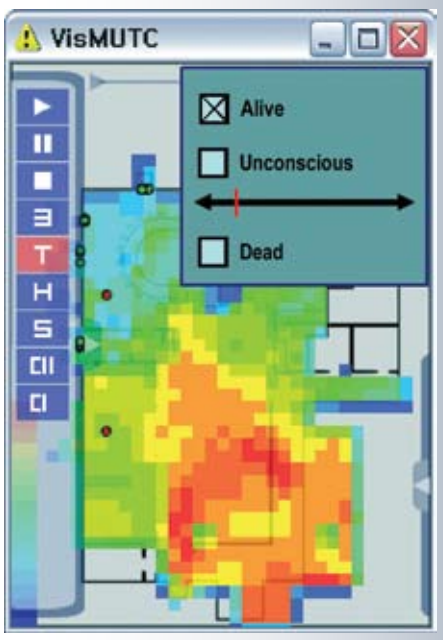
Team Members

David S. Ebert, Timothy Collins, Sungye Kim, Yun Jang, Avin Pattath, Purdue University

Point of Contact

David S. Ebert, Director, Purdue University Regional Visualization and Analytics Center, ebertd@purdue.edu



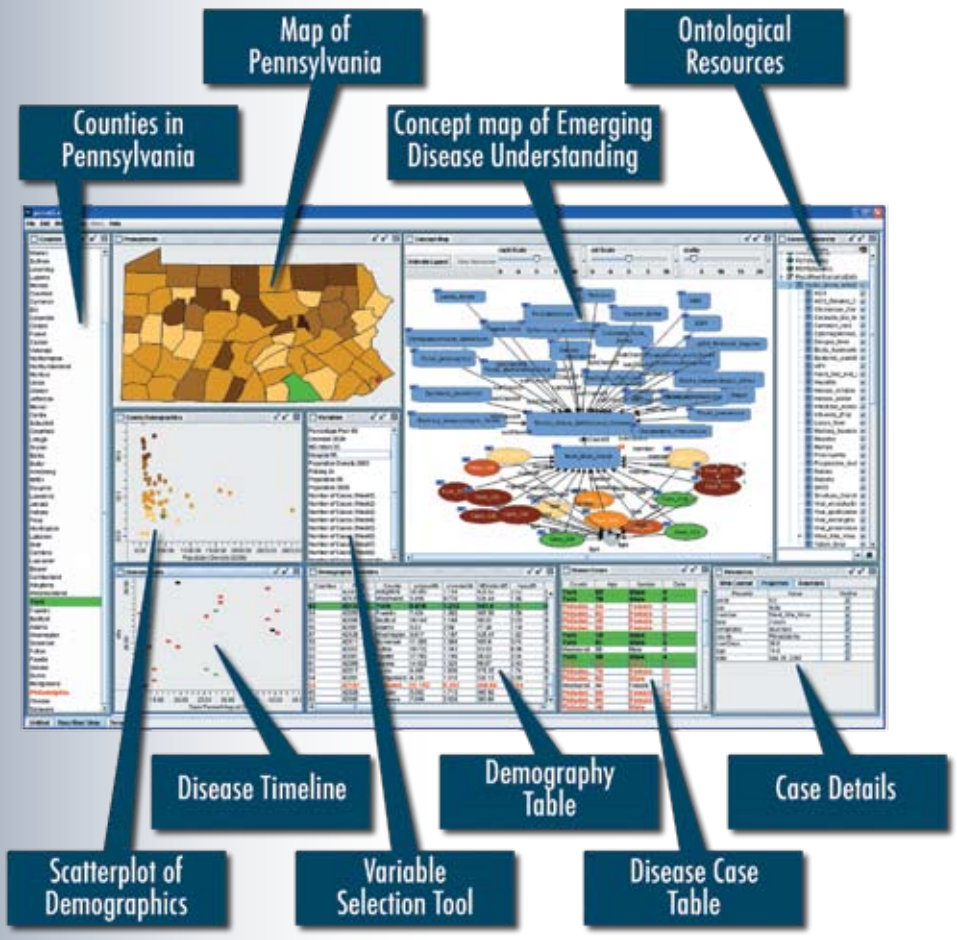


Two planned displays from our prototype emergency response mobile analytics system using simulation data of the evacuation during "The Station" nightclub fire. The prototype displays have been augmented with our planned additional visual analytic capability overlays. Top: location and paths taken by individuals during the evacuation and overlay showing the health of a selected subset of individuals. Bottom: location of individuals in the building and the extent of the fire spread using a visualization of temperature simulation data.

Integrating Knowledge and Data: A Practical Demonstration Using a Disease Scenario

A key challenge for analysts charged with identifying and assessing threats to public health, infrastructure or border security is the task of connecting knowledge to data. Knowledge and data do not exist in isolation. Concepts and relationships can impose useful structure on complex data, leading to the detection of the unexpected; likewise, data can modify or give rise to new concepts and relationships. However, current systems enforce an artificial separation, with knowledge in one system (ontologies, concept maps, knowledge bases) and data in another (databases, information visualization, exploratory data analysis).

The North-East Visualization and Analytics Center team at Penn State University has been developing a conceptual implementation framework to connect the knowledge held in ontologies and concept maps to the data from which knowledge is derived and to which knowledge-derived structure is imposed (an ontology is a description of the concepts and relationships within a domain, represented in formal logic). For example, a concept in an ontology can connect to instances in a geographical map, or several key data observations in data can be exemplars for defining a new (emerging) concept. NEVAC has implemented its ideas in a prototype application that links two of its existing tool sets: *ConceptVista*—for exploring concept maps and ontologies and *Improvise*—for highly coordinated information visualization. These ideas are demonstrated below through a scenario focused on a hypothetical vector-borne disease outbreak.



An overview of the combined knowledge and data application, with the various tools and displays available annotated. Disease instances (cases) are logged, mapped and browsed spatially, temporally and in terms of their attributes. The knowledge and data displays are coordinated so that user actions in one display will typically lead to related actions in linked displays.

Demonstration Scenario: Responding to a Public Health Threat

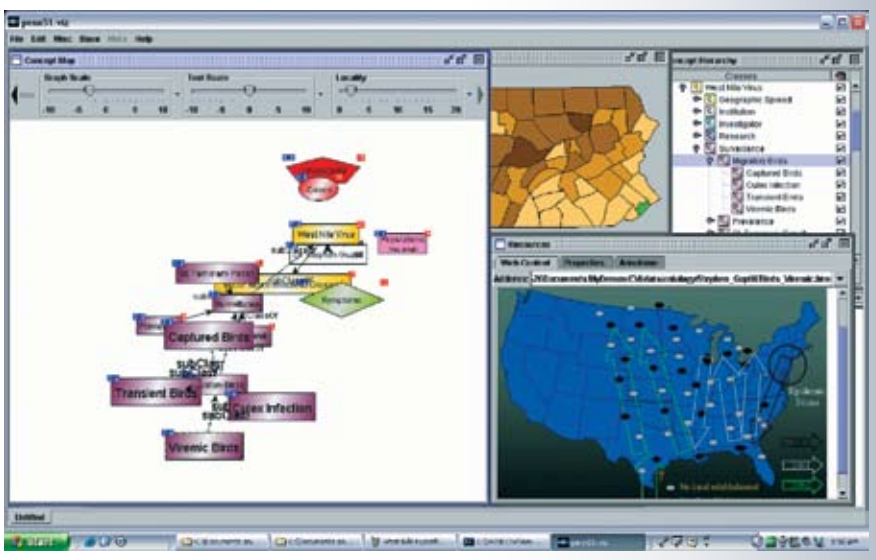
In the scenario, a disease outbreak is recognized. As it plays out, hypotheses about possible causes are evaluated, new knowledge is added to the ontologies and useful resources and information are gathered from the internet by a user-steered, semantic search driven by the ontologies. Search is currently supported in Google, Amazon and the Digital Library for Earth Science Education; information in Citeseer and Wikipedia can also be accessed.

As the scenario evolves, human disease cases are logged and connected to a geographical map (of Pennsylvania), with symptoms and related case information held in a table. The analysts begin by exploring a likely interpretation—that the outbreak is a resurgence of West Nile Virus. However, they notice that some reports do not tally with expectations for West Nile Virus. Using the semantic search capabilities of *ConceptVista*, they quickly locate useful Centers for Disease Control and Prevention reports and additional ontological resources that describe the relationship between the disease and migratory birds. These

reports are then linked into the existing conceptual knowledge base.

Analysts focus on a symptom that appears in the case reports but not in the existing disease ontology—a "tingling" sensation in major joints. This symptom is evidence counter to the initial hypothesis of a West Nile Virus outbreak. A new "tingling" concept is added to a local version of the disease ontology as a symptom to which the various cases can be connected. Lack of initial connections suggests that the disease ontology may be incomplete, and likely causes for the outbreak are yet to be identified. To address this, analysts consider alternative hypotheses—other diseases that might better fit the observed data. A semantic search for vector-borne diseases, with search terms harvested from the ontology and the added search term of "tingling," uncovers a candidate disease, Ross River Virus, and leads to authoritative descriptions of likely symptoms from Health Authorities in Australia. The next step is to check for a match in the anticipated infection patterns of the disease in Pennsylvania by examining spatial and temporal aspects of spread.

The steps outlined above generate a plausible hypothesis that an unexpected disease has entered the U.S. Analysts will then focus on the tasks of collecting more data and assessing competing hypotheses. If Ross River Virus is confirmed, then subsequent work will turn to determining how the disease entered the country (whether by accident or design) and to anticipating impacts and responding quickly.



Exploring the disease outbreaks in terms of migratory bird behavior.

Summary

Integrating knowledge with data supports holistic analysis, wherein conceptual findings can be directly represented and linked to appropriate data and where data is evidence used to select concepts appropriate to the situation. The entire analysis outcome outlined above—including maps, modified ontologies and the resources found—can be captured for sharing with other analysts or for later review.

A shortcoming of currently deployed knowledge management and data analysis systems is the lack of an integrated data model that encompasses knowledge, data and the myriad of possible connections between the two. Such an integrated data model is vitally needed to underpin visual analytics applied to realistically complex, real-world scenarios.

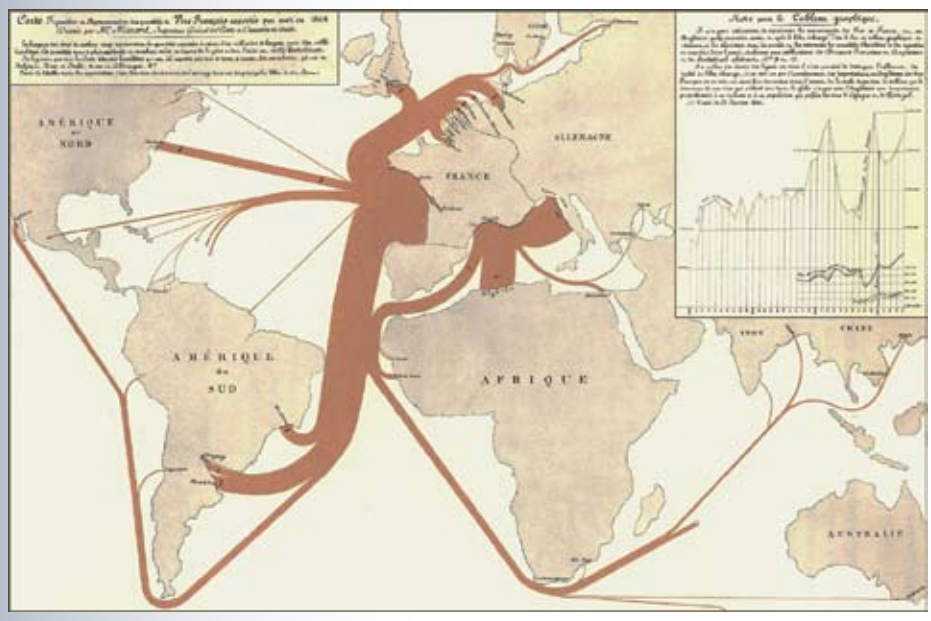
Team Members

Mark Gahegan, Ritesh Agrawal, Tawan Banchuen, Junyan Luo, Steve Weaver, NEVAC and GeoVISTA Center, Penn State University

Point of Contact

Mark Gahegan, Professor of Geography, The Pennsylvania State University, mng1@psu.edu & Alan MacEachren, Director, NEVAC, maceachren@psu.edu

Automating Flow Map Layout

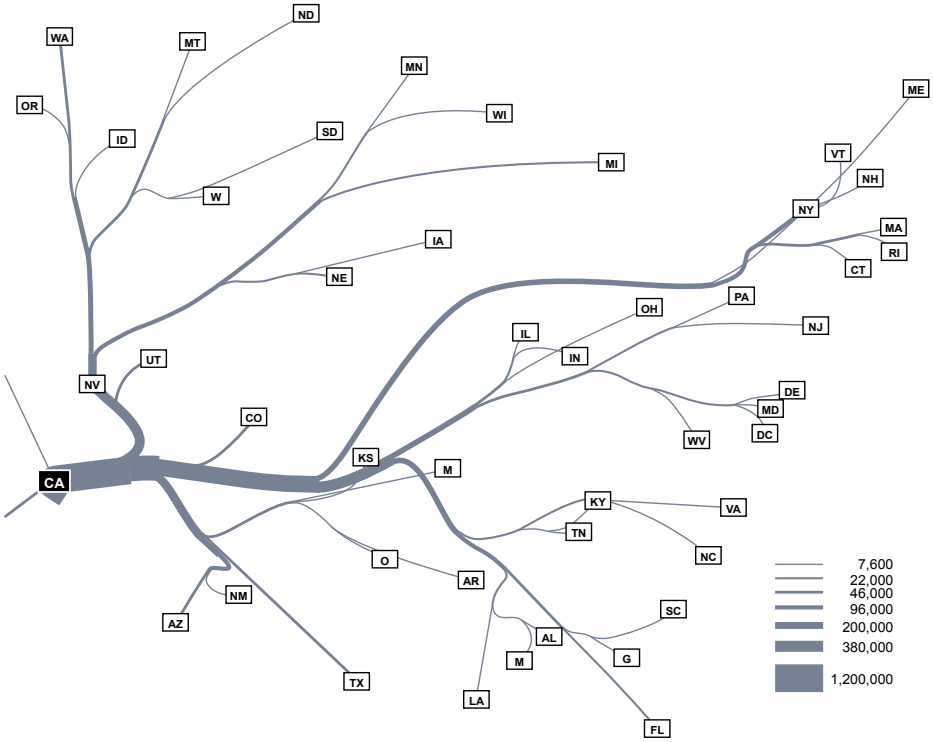


French Wine Exports (Minard 1864). The width of each line represents the volume of wine exported from France. The lines decrease in proportion at each branch point. To keep the presentation clear, Minard distorted many geographic features such as the Strait of Gibraltar.

Analysts frequently need to understand how goods and, more recently, information move through complex webs of connections. Visualizing such traffic and the network's topology is challenging because a display of numerous connections with lines results in visual clutter that makes it difficult, if not impossible, for the analyst to interpret. Researchers have tried to adapt visualization techniques such as node maps and matrix displays to this problem and have even created novel 3D representations to reduce visual clutter, but representing a large number of individual connections is still likely to produce cluttered maps.

Cartographers in the 19th century addressed this problem by creating flow maps. Flow maps (Slocum 1999) depict the movement of objects among locations by overlaying a geographic map with lines of decreasing widths representing the number of objects as they moved from a source towards their destinations. Constructing a flow map is like moving upstream from the mouth of a river system, such as the Mississippi, to all the tributaries. The river's width and then the width of each tributary gradually decrease where the river branches. Minard's 1864 flow map of wine exports from France is a beautiful example of an early flow map.

However, even today's flow maps are usually drawn by hand. Because of this, although they are widely employed to present data to audiences, flow maps have not been used by analysts when they are interactively exploring data, which is the heart of the analytic process.



California migration 1995-2000 (U.S. Census). This map shows the distribution of people leaving California for other states. The automatic layout feature of the system has slightly moved the positions of some state labels from their absolute geographic location.

Approach

At the Stanford Regional Visualization and Analytics Center, the goal was to develop automated techniques that generate flow maps in a way that captures the features of hand-drawn ones. Analysis of good, hand-drawn maps revealed three common characteristics: 1) intelligent distortion of positions to create more usable space; 2) merging of edges that share destinations; and 3) intelligent edge routing.

Minard's map of wine exports illustrates each of these characteristics. First, the Strait of Gibraltar has been widened and Great Britain has been shifted away from France to make room for passing flow lines. However, these distortions preserve the relative positions of the land masses. Second, edges going to the various regions of the world are merged. Finally, the edge crossings are minimized. Cartographers identify additional characteristics about visual appearance: 1) flow lines should be dominant and easily distinguished from other map symbols; 2) the thickness of lines should be in direct proportion to the data values; 3) if edges cross, smaller flow lines should be placed on top; and 4) the map legend must be clear and provide key values for line widths.

The Stanford RVAC system (Phan 2005) achieves intelligent distortion of position by using a layout algorithm that maintains the relative ordering of left-right and up-down among objects. Edges are merged by using hierarchical clustering to find nodes that are close together. The edge drawing and routing also use hierarchical clustering. The legend for the map is automatically generated but is placed manually. Finally, the system can layer several maps together to enable the analyst to see the result of multiple data queries.

Results

The Stanford RVAC first applied the system to migration data sets from the U.S. Census mapped to a Mercator projection of the United States. They later adapted it to show global ecological data provided by the Footprint Network on Google™ Earth.

Because analysts often work in conceptual spaces, not just geographic ones, the Stanford RVAC has experimented with using the algorithms in a human rights context to show flows of documents in a dictatorship and to give analysts an alternative view of intrusions on a computer network.

References

Phan D, L Xiao, R Yeh, P Hanrahan and T Winograd. 2005. "Flow Map Layout." In *InfoVis '05 Proceedings of the 11th IEEE Symposium on Information Visualization,* pp 219-224. October 23-25, 2005, Minneapolis, Minnesota. IEEE, Piscataway, New Jersey.

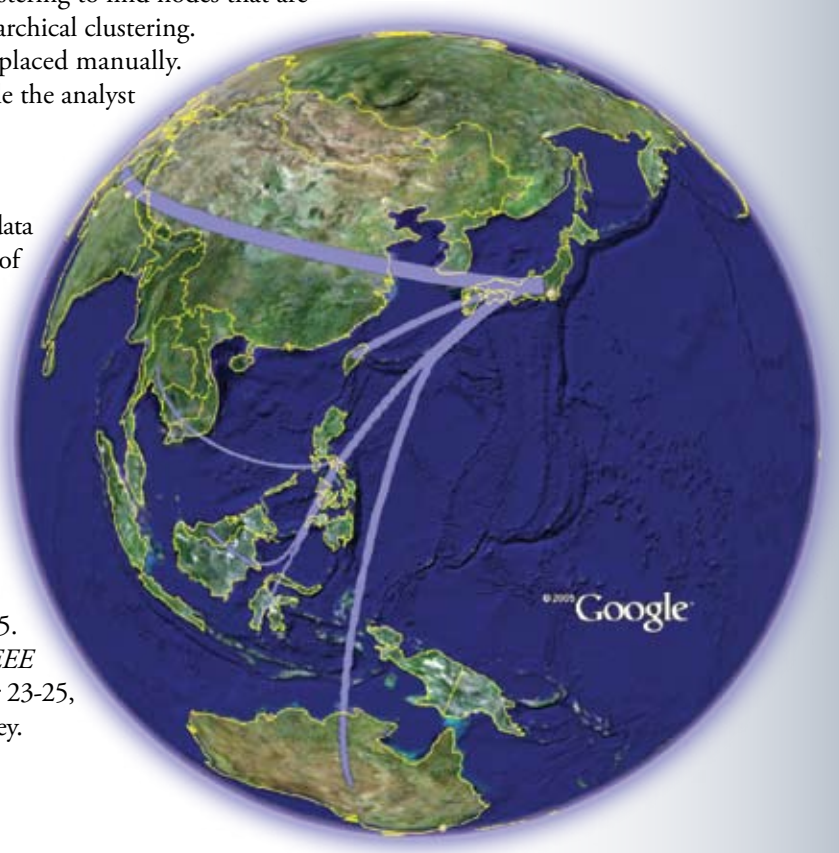
Slocum T. 1999. *Thematic Cartography and Visualization.* Prentice-Hall, Upper Saddle River, New Jersey.

Team Members

Doantam Phan, Ling Xiao, Ron Yeh, Pat Hanrahan, Terry Winograd, Stanford University

Point of Contact

Dontam Phan, Research Assistant, Stanford University, dphan@stanford.edu

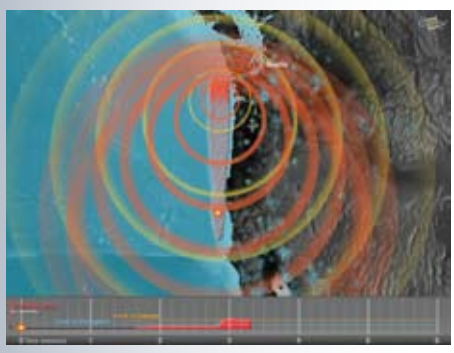


Ecological Resource Migration (Footprint Network 2005). *Flow maps can be very powerful when superimposed with geographic substrata. This flow map shows the countries that Japan depends on for natural resources. The large flow that disappears to the west is America.*

RimSim: Technology Solutions for Pacific Rim Disasters

When a first responder arrives at a chaotic scene of human trauma and dangerous physical obstacles, adrenaline and instinct rise. Perception and cognition are pushed to the limit. A lack of contextual information, critical to situation awareness, adds to the confusion. Optimal long-term recovery actions take second priority to short-term triage.

A March 2006 earthquake scenario workshop held in Seattle revealed that personnel responsible for different aspects of response often find themselves competing for the same resources—for example, fire personnel need water from the water utility to combat fires while health personnel need clean water for drinking. In addition, short-term and long-term goals often conflict and need to be balanced against each other.



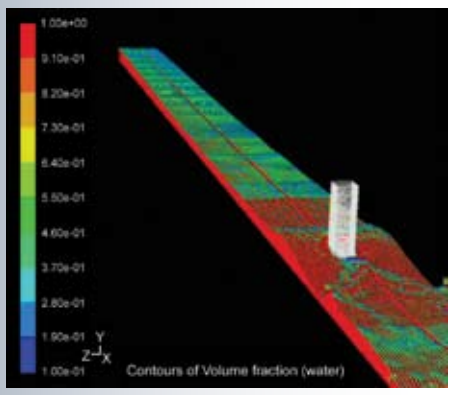
Visualization of earthquake three minutes after start.

RimSim Tackles Emergency Response Concerns

How can technology help first responders prepare for their roles and continue to aid them as they arrive at the scene? The Pacific Area Regional Visualization and Analytics Center, led by the University of Washington, is working to address these issues through a project named RimSim.

The RimSim project, a collaboration among PARVAC and others in the Pacific Rim, plans to capture regional response and recovery plans in a software system for use during multi-jurisdictional training exercises. This multi-user software framework will evolve toward simulating catastrophe scenarios (e.g., common Pacific Rim threats such as earthquakes, tsunamis, volcano eruptions and disease outbreaks) where response personnel can role-play their actions and evaluate their response.

Within this software framework, a simulator will constantly examine how participants interact with each other and provide visual analysis aids during role play. The simulator will provide a scientific visualization of the event and reveal damage to health and property as it might unfold. Afterwards, potential opportunities for improvement will be identified and discussed. Follow-up simulation sessions can reinforce lessons learned by playing a similar scenario or the same scenario again.



Visualization of tsunami.

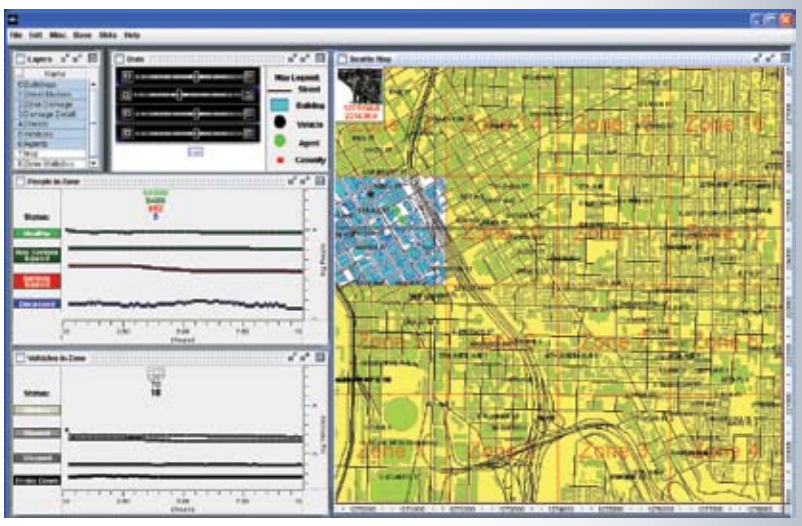
Approach

After performing interview and tabletop techniques to identify opportunities for simulation, the RimSim project will provide technology for modeling distributed cognition explicitly and identifying socio-cultural barriers to overcome. Building upon Edwin Hutchins' models of distributed cognition (1995), which have been used to analyze and evaluate the flow of data representations within minds and across shared visual artifacts in real-world cooperative work settings, the RimSim project team will build the crisis response simulator described above that helps response teams learn to explicitly manage their cognition and offload important real-time data to shared visual artifacts.

Software applications like SimCity™ let players (who usually play alone) plan communities and respond to catastrophes with an idealized knowledge of what is happening. RimSim, though similar in some ways, will attempt to simulate a more realistic state where communication links may fail, damage and health status are often unknown or incorrectly reported, and fires or water main leaks may be revealed over long time periods. Simulated interfaces from emergency operation center communications will be simplified to suggest appropriate actions to take when absorbing the full extent of details is not possible or practical.

The RimSim project's technical focus will be on managing and presenting explicit distributed cognition through a competent multi-user technological infrastructure. Multi-user facilities, including inter-object messaging, player-object messaging, synchronous communication channels, and asynchronous discussion features will need to be fine-crafted to simulate existing or feasible facilities.

With those facilities in place, other technologies from the visual analytics community can provide rich experiences for simulation players: Map display technologies can continuously visualize the state of the community; agent simulation technologies can simulate people, vehicles and other system model outputs (e.g., water, air, earth); databases can maintain the state of other key variables such as medical supplies and supplies used by response personnel; a game engine can execute rules regarding the interaction of simulation variables and can execute simulation flow including any optimization algorithms that players request be run.



RimSim prototype interface showing an earthquake.

Technology Integration

PARVAC looks to NVAC and other RVACs to provide solutions for use within the simulation. If a technology is available to community planners or is used during response and recovery drills, the RimSim project team will consider how best to integrate it into the simulator. As examples, the strength of the North-East Visualization and Analytics Center's geospatial and emerging temporal presentation will allow PARVAC to provide powerful map display components when needed, and the strength of the Purdue Visualization and Analytics Center's agent modeling abilities will allow PARVAC to better represent thousands or millions of people in a simulation. PARVAC's own advanced three-dimensional interfaces, augmented reality technologies and optimization modeling efforts can provide simulator components.

If you are interested in learning more about Distributed Cognition, two good starting points are:

Harris, Steven. "Distributed Cognition." Available at: http://www.slis.indiana.edu/faculty/yrogers/dist_cog

Rogers, Yvonne. 1997. "A Brief Introduction to Distributed Cognition." Available at: http://www.informatics.sussex.ac.uk/users/yvonner/papers/dcog/dcog94.pdf.

References

Hutchins, Edwin. 1995. *Cognition in the Wild.* MIT Press, Cambridge, MA.

Team Members

Thomas Furness III, Suzanne Weghorst, Elisabeth Cuddihy, Bruce D. Campbell, Daniel Belcher, Anirudhan Vijayakanthan, University of Washington

Point of Contact

Bruce D. Campbell, Research Scientist, University of Washington, brucedc@u.washington.edu

New Digital Library Promotes Visual Analytics Learning

Educating the next generation of visual analytics scientists is critical to helping prevent terrorism and solving problems in energy and the environment. The newly established Visual Analytics Digital Library, developed by the SouthEast RVAC at Georgia Tech and UNC-Charlotte, is key to supporting this education effort. The SRVAC initially created the library to assist instructors building academic curricula of courses ranging from individual undergraduate or graduate classes on visual analytics to shorter professional development and continuing education workshops.

VADL is a repository of visual analytics-related materials, including articles, course notes, lecture slides, video lectures, homework exercises, and more. Materials are submitted by visual analytics researchers and educators who are developing visual analytics courses. By accumulating all these materials at one site, VADL provides a convenient, easy-to-access collection that will benefit a variety of academic institutions and researchers. VADL is also intended for students and others who are simply seeking to learn more about the field.

VADL is not a loosely scattered collection of different documents. Instead, it is a digital library with a developing taxonomy of visual analytics that serves as the organizational framework for the materials. The taxonomy consists of a small number of top-level categories, including Analytic Reasoning and Processes; Visualizing Data, Information and Knowledge; Production, Presentation, and Dissemination, with each providing a few subcategories. The taxonomy is expected to evolve and expand over time.

Materials within VADL are located in an appropriate category or categories to help users who are browsing by topic. Users also can use a text-based search capability that provides an innovative tree-map visualization of retrieved documents within the information space. They also can search by material type such as PowerPoint slides or documents. Our goal is to provide a flexible set of options to help people with differing styles or objectives to explore the library.

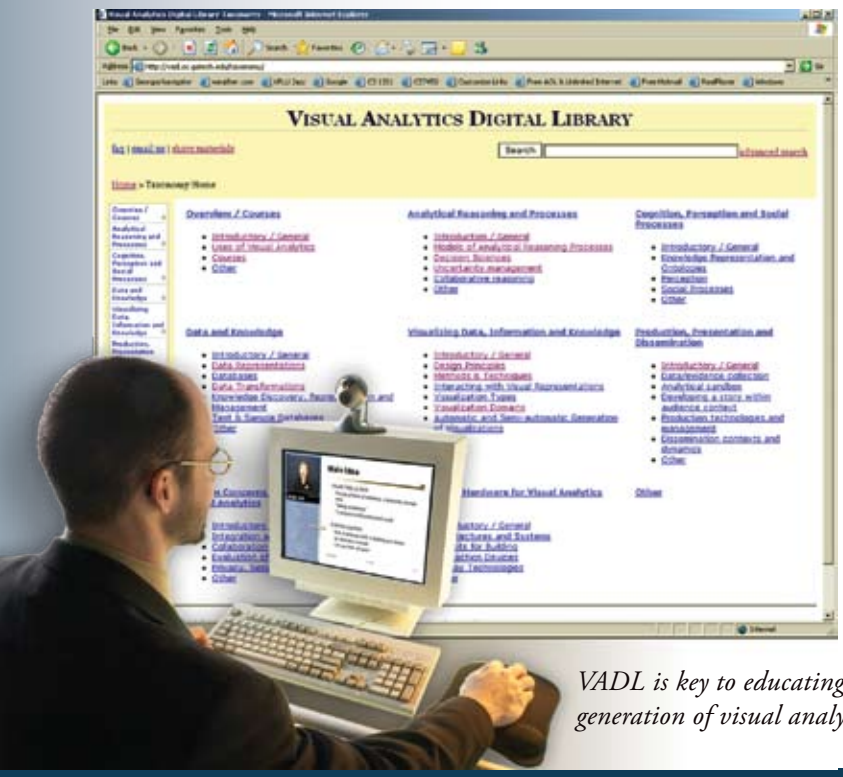
Currently in the early stages, the library contains a small but diverse collection of materials. The collection includes a presentation on semantic video classification and automatic annotation, a web lecture on visual communication in computer graphics and art, a presentation on network security data visualization and an article about the sense-making process.

All materials within VADL are user-contributed. Authors who have written papers on visual analytics-related topics or instructors who have lecture slides, homework exercises, or exam questions are encouraged to contribute their materials to the library. The home page includes a Share Materials link that activates a submission form. The goal is to make this site the place to go for accessing visual analytics materials.

VADL is located at http://vadl.cc.gatech.edu.

Point of Contact

John Stasko, Professor, Georgia Tech, stasko@cc.gatech.edu



VADL is key to educating the next generation of visual analytics scientists.

2006 Interns

NVAC's outstanding group of 2006 interns brought with them experience in a wide range of technical areas including cognition, visualization, geography, mathematics, computer animation, psychology and computer graphics. Each intern had an NVAC project lead as a mentor and worked on a challenging task that enabled him or her to provide a real contribution to the project.

Some of the 2006 interns and their assigned projects include:

Heather Eslinger, George Fox University, undergraduate, worked on adding detailed 3D building and vehicle models to 2D maps. This work will help provide valuable navigational and orientation aids for remote operational situational awareness.

Lyndsey Franklin, Washington State University, undergraduate, worked on a scenario analysis system adding a variety of new functions and capabilities. These efforts enhanced support for the analysis of competing hypotheses technique.

Avin Pattath, Purdue, Ph.D. candidate, worked on temporal visual analytics techniques for mobile appliances adapted from visualizations developed for the desktop. This work plays a vital role in our ongoing efforts to push analytical capabilities to field users.

Kanu Singhal, Georgia Institute of Technology, Ph.D. candidate, worked on techniques to control force-directed graph layouts according to semantic properties. This is part of our continuing research on visual analytic uses of structured graphs.

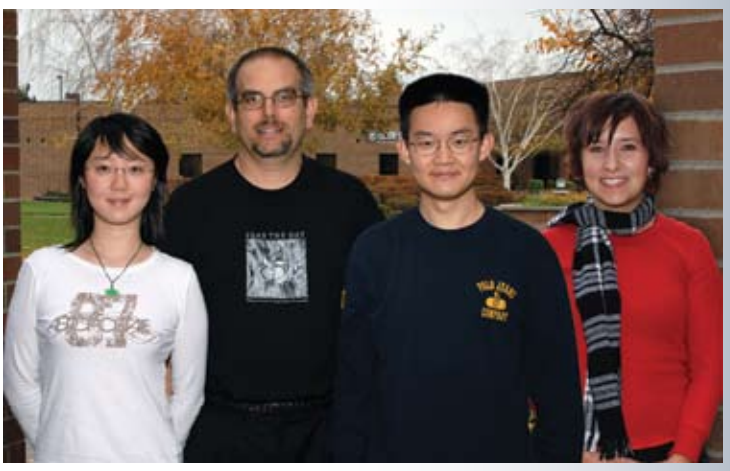
Yang Wang, Purdue, Ph.D. candidate, developed techniques to improve the speed and reliability of context-driven knowledge base access and defined and implemented approaches for rapid entity extraction that can be embedded into ongoing research.

"The internship at NVAC gave me the chance to contribute to a project which would solve a true real-world problem and to collaborate with the scientists getting the research done. The research being done is focused on producing generally applicable techniques for handling information from a variety of problem sources, each requiring unique expertise. It's not enough to be specialized in a single field because there is too much to be gained from the collaboration of many."
– Lyndsey Franklin

Point of Contact

Richard May, Chief Scientist,
Pacific Northwest National Laboratory, richard.may@pnl.gov

Summer 2006
Kanu Singhal, Georgia Institute of Technology
Brennan McQuerry, Whitworth College
Lyndsey Franklin, Washington State University
Jamison Conley, Penn State University
Avin Pattath, Purdue University



Fall 2006
Yang Wang, Purdue University
Bruce Campbell, University of Washington
Ling Xiao, Stanford University
Heather Eslinger, George Fox University

A Partnership with Academia, Industry and Government

Consortium Members – Provider Level

BOEING
inxight
Objectivity
Intelligent Results
Microsoft
GCS research
i2 A ChoicePoint® Company
TABLEAU SOFTWARE
Raytheon

Program Representatives

John Hoyt
Department of Homeland Security

Joseph K. Kielman
Department of Homeland Security

James J. Thomas
National Visualization and Analytics Center

Resources

The following publications are related to visual analytics. This list is by no means inclusive, nor is it meant as an endorsement of those mentioned here versus those that are not.

Committee on Basic and Applied Research Priorities in Geospatial Science for the National Geospatial-Intelligence Agency, Mapping Science Committee, National Research Council. 2006. ***Priorities for GEOINT Research at the National Geospatial-Intelligence Agency.*** The National Academies Press, Washington, D.C.
This report defines 12 problems in geospatial science that NGA must resolve to meet future needs.

Heuer R. 1999. ***Psychology of Intelligence Analysis.*** U.S. Government Printing Office, Washington, D.C. Available online at http://www.cia.gov/csi/books/19104/
Discusses how subconscious cognitive processes can limit our reasoning abilities and how we can try to understand and negate these effects.

Johnston R. 2005. ***Analytic Culture in the U.S. Intelligence Community.*** Central Intelligence Agency, Washington, D.C.
An inside look, from an anthropological perspective, at the analytical culture in the intelligence community.

Jones M. 1995. ***The Thinker's Toolkit: Fourteen Skills for Making Smarter Decisions in Business and in Life.*** Crown Business, New York.
Provides practical methods for simplifying problems and making faster, better decisions.

Lowenthal M. 2005. ***Intelligence: From Secrets to Policy.*** 3rd edition. CQ Press, Washington, D.C.
Describes how the intelligence community's history, structure, processes, and functions affect policy. This 3rd edition analyzes pre- and post-9/11 issues and events.

National Commission on Terrorist Attacks. 2004. ***The 9/11 Commission Report.*** WW Norton & Company, New York.
The report is a comprehensive examination of the events leading up to the 9/11 terrorist attacks. The commission traces the roots of al-Qaeda's strategies, describes how the 19 hijackers entered the United States and boarded airplanes, and details the missed opportunities of law enforcement officials.

Schwartz P. 1991. ***The Art of the Long View.*** Currency Doubleday, New York.
Outlines a scenario approach for planning. Schwartz describes new techniques, originally developed within Royal/Dutch Shell, based on many of his firsthand scenario exercises with the world's leading institutions and companies.

Thomas JJ, KA Cook (Eds). 2005. ***Illuminating the Path: The Research and Development Agenda for Visual Analytics.*** IEEE Computer Society Press, Los Alamitos, CA.
Describes the field of visual analytics and outlines the visual analytics science and technology research agenda.

Become a Member Today!

Join as either Providers of products or Consumers of solutions.

"The Consortium is a true collaborative environment."

David Lane, Objectivity, Inc.

Key Benefits of Membership:

- Paid registration to Consortium meetings
- Participation in annual technical workshops
- Opportunities to engage DHS/other federal agency staff
- Interactions with technology providers and consumers
- Access to published reports of software evaluation and test datasets
- Awareness of key DHS programs and procurement opportunities.

A complete listing of benefits by type and level of membership can be found on the VAC Consortium website at:

http://nvac.pnl.gov/consortium.stm

"Being a member of NVAC provides us the ability to be in the position of leading and pushing the field forward."

George Robertson, Microsoft

Printed on recycled paper

PNNL-SA-53589