Information Visualization (2006) 5, 237–249

www.palgrave-journals.com/ivs

Walking the path: a new journey to explore and discover through visual analytics

Pak Chung Wong[1]
Stuart J Rose[1]
George Chin Jr.[1]
Deborah A Frincke[1]
Richard May[1]
Christian Posse[1]
Antonio Sanfilippo[1]
Jim Thomas[1]

[1] *Pacific Northwest National Laboratory, P.O. Box 999, K7-28, Richland, WA 99352, U.S.A.*

Correspondence:
Pak Chung Wong, P.O. Box 999, K7-28, Richland, WA 99352,U.S.A.
Tel: +1 509 372 4764;
Fax: +1 509 375 3641;
E-mail: pak.wong@pnl.gov

Received: 20 February 2006
Revised: 31 July 2006
Accepted: 3 August 2006

Abstract
Under the leadership of the US Department of Homeland Security (DHS), researchers at the Pacific Northwest National Laboratory (PNNL) established a research center focusing on the discipline of visual analytics in 2004. A year later, the center led a multidisciplinary panel representing academia, industry, and government to formally define directions and priorities for future research and development (R&D) for visual analytics tools. The R&D agenda, *Illuminating the Path*, defines the term visual analytics as 'the science of analytical reasoning facilitated by interactive visual interfaces'. This article describes our progress to date in walking that path. We briefly describe the background of the subject, present major professional activities and accomplishments of its community, and highlight some of the ongoing R&D efforts being carried out by researchers at PNNL to fulfill the requirements and missions of a new discipline that promises to change the way we deal with today's information.
Information Visualization (2006) **5**, 237–249.
doi:10.1057/palgrave.ivs.9500133

Keywords: Visual analytics; data exploration; knowledge discovery; information visualization; analytical reasoning; interactive visual interfaces; visual representation

Introduction

People understand their world through integrating direct observations with conceptual models. Visual analytics offers a natural continuation of mankind's history of discovery and insight through reconciliation of theoretical models with direct observation of evidence. Discovery and insight become possible when observations do not fully conform to a prevailing theory intended to explain those observations.

Scientific revolutions in astronomy and cosmology developed out of efforts to align theoretical models with observations of the positions of planetary bodies on the celestial sphere. The realization that the earth moves around the sun was made possible through carefully measured and recorded observations about their positions over time. Observations of the planet Venus' apparent retrograde motion conflicted with a geo-centric model of the solar system and offered Galileo[1] a convincing case to further refine the heliocentric model.[2]

In the 19th century, prevailing theories on how inheritance passed through generations involved the notion of *blending* in which the parents' traits intermingled in order to define characteristics for their children. A natural outcome to blending inheritance would be a population of individuals all possessing the same traits, averaged across their ancestors due to the lack of a means to retain unique traits (such as red hair) over generations, an outcome not supported by all the facts. The Austrian monk Gregor Mendel[3] conducted plant breeding experiments in the late 1800s and considered that distinct inheritance units contributed from each parent better explained his recorded observations of inherited traits through generations of pea pods. The discovery of genes developed from direct observation of inherited traits through generations of peas.[4]

Owing to the limited resources of most scientific endeavors, the subject of study is not always available for direct observation. More often, a visual representation is generated from one or more sets of observations that are typically refined and abstracted to their essential components. This visual representation, to the extent that it is faithful to observations of the subject, serves as a useful record as it reproduces for the observer the essential characteristics of the subject.

Visual representations are essential aids to human cognitive tasks and are valued to the extent that they provide stable and external reference points upon which dynamic activities and thought processes may be calibrated and upon which models and theories can be tested and confirmed. The active use and manipulation of visual representations makes many complex and intensive cognitive tasks feasible. A visual representation is able to convey relationships among many elements in parallel and provides an individual with directly observable memory.

Furthermore, information that is not directly observable may be abstracted to a visual representation for reference. Tufte[5] outlines early examples of statistical graphics in which quantitative data are plotted in a time-series. A visual representation graphically conveys the pertinent relationships among multiple sets of information.

Rapid development and improvement of information technology (IT) in the 20th century provided computational resources that excelled in performing specific quantitative functions to solve equations and calculate metrics. In order to arrive at a useful answer, a scientist or analyst needed a well-formed question from which an instruction set could be derived as input for a computational function. Primarily a result of technical constraints, computational capabilities for the better part of the 20th century had limited capabilities for allowing a person to interact with information pertinent to a problem in a manner that supported intensive online analysis. To obtain useful results, a user had to know which questions to pose. Problems had to be framed in very precise terms. As computation was a limited resource, reasoning about a specific problem set needed to occur before and after applying a computational function.

While computational power has increased exponentially, the ability to interact with useful information has only increased incrementally. In recent decades, the exponential increase in computing power has allowed many more questions to be posed and more complex problems to be addressed. Information is now massive, disparate, and disorganized. The dimensionality of data has also increased, requiring greater effort to identify and comprehend relationships relevant to a particular analytic task. Simple metrics miss significant detail; the classic example from Anscombe[6] shows that metrics are insufficient for understanding complex relationships.[7]

Throughout the last few decades, researchers from different disciplines and backgrounds have gradually and independently developed novel technologies to address these complicated visual representation issues. Classic references from Tukey,[8] to Tufte,[5] to more recently Card *et al.*[9] have all contributed to formalize the ever-evolving studies demanded by industry, government, and society at large. This article presents the latest requirements and challenges developed along that evolution, and the leading community efforts[10] to define the future of the next-generation analytical reasoning technologies through visual means.

National Visualization and Analytics Center™ and visual analytics

Operated by the Pacific Northwest National Laboratory (PNNL) for the US Department of Homeland Security (DHS), the National Visualization and Analytics Center™ (NVAC™)[11] is a national and international resource providing strategic leadership and coordination for visual analytics technology and tools. NVAC supports DHS's mission to secure our homeland and protect the American people by giving analysts and emergency responders technology and capabilities to:

- Detect, prevent, and reduce the threat of terrorist attacks;
- Identify and assess threats and vulnerabilities to our homeland;
- Recover and minimize damage from disasters.

In the summer of 2005, NVAC published a 5-year research and development (R&D) agenda[12] for visual analytics to facilitate advanced analytical insight. Visual analytics as described in the R&D agenda 'is the science of analytical reasoning facilitated by interactive visual interfaces'. The R&D agenda presents recommendations to advance the state of the art in the following major visual analytics focus areas:

- Analytical reasoning techniques;
- Visual representations and interaction techniques;
- Data representations and transformations;
- Production, presentation, and dissemination.

The goal of NVAC is to facilitate the creation of new technologies that help address the challenges of DHS. To accomplish this, we need to (1) foster unique partnerships among national laboratories, university research centers, scholars, and other government agencies as part of an ongoing commitment to collaboration in the discipline of visual analytics and to reduce the time transitioning software to wide analytical use by improving the access of analysts and researchers to new technologies and real challenges and (2) enable an enduring talent base to advance visual analytics research. This research will benefit not only DHS but also a variety of scientific and business areas, such as more effective telecommunications and more reliable power grid infrastructure.

NVAC also leads a series of community outreach efforts to promote the new research initiatives among different academic and industrial communities. In the last 2 years, NVAC scientists have guest-edited a special issue of *IEEE Computer Graphics and Applications on Visual Analytics*,[13] and an upcoming special issue of *IEEE Transactions on Visualization and Computer Graphics*[14] on the same topic. In February 2005, NVAC unveiled its roadmap at the American Association for the Advancement of Science (AAAS) annual meeting[15] and later presented it at the 2005 International Conference on Intelligence Analysis[16] and InfoVis 2005.[17] In 2006, NVAC scientists will play a major role in organizing the first IEEE Symposium on Visual Analytics Science and Technology (VAST)[18] in Baltimore, MD. Cutting-edge visual analytics technologies developed by industry, academia, and NVAC will be showcased at the symposium.

While *Illuminating the Path* lays out the R&D agenda for visual analytics, this article describes our progress to date in walking that path.

Recent R&D work at PNNL

Much of the visual analytics research at PNNL has focused on developing and applying new interactive visual interfaces that may bring new insights and greater clarity to scientists and analysts. A primary goal is to visually elucidate patterns and features in the data that have scientific or analytical relevance and meaning. The following subsections discuss our research progress to date. All demonstrated examples used in the discussion are based solely on publicly accessible data as indicated. All analytical results reported in the paper have been presented in public forums as cited. Readers are encouraged to follow the references and read the corresponding papers for full details.

Visual representations of knowledge

An essential quality of visual analytics is to represent and apply a person's working concepts, patterns, and theories – drawing from his or her cognitive model of the science, analysis, and data. Thus, the patterns and insights that evolve originate from their specific experiences and knowledge. Within an analytic process, analysts dynamically operate between these two perspectives as they formulate concepts and hypotheses and then interpret massive amounts of data to identify new or verify existing knowledge, which then serves to validate or evolve the initial concepts and hypotheses.

Scientific researchers and information analysts often conceptualize problems and theories in graphical form. Using tools from pen and paper to computer-based drawing packages, researchers construct diagrams and pictures to convey complex scientific or analytical concepts and processes (see Figure 1). And despite differences in symbols, shapes, colors, fonts, and drawing styles, researchers are generally able to comprehend and interpret graphs and diagrams generated by others within the same field or discipline. Visual representations provide a language that researchers may apply to document their scientific or analytical knowledge and to communicate this knowledge to others. Furthermore, studies have shown that the cognitive processing of complex subject matter is improved when the structure of ideas and concepts are made explicit[19] and that visual representations such as diagrams and semantic graphs may serve as effective 'scaffolds' for cognitive processing.[20]

PNNL develops methods of representation that allow analysts to visually integrate what they know with what the data is telling them. At PNNL, we are deploying semantic graphs within applied computing systems as central user representations of scientific or analytical knowledge. These systems provide concept-based user interfaces that allow analysts to visually define and capture conceptual models of their scientific or analytical problems, hypotheses, theories, and processes. Once defined, the conceptual models then become a visual interaction framework for accessing and applying computational resources and capabilities. As described in the systems below, captured conceptual models may be logical in representing how individual concepts tie together to form higher theories, analytical in conveying intermediate or final analysis results, or temporal in describing experimental or investigative processes in which concepts are physically and computationally explored. Through such systems, interactions and computations are always carried out in the context of visual knowledge representations, and thus, researchers are continually reminded of the scientific or analytical concepts that enable, support, and constrain their investigations.

In the development of the regional climate modeling problem-solving environment (RCM-PSE),[21] scientific workflow graphs were constructed early on during requirements analysis. Regional climate modelers drew scientific workflow graphs on paper and whiteboards to describe the steps of their experiments and the computational resources they applied along the way.[22]

Finding that workflow graphs were both intuitive and easy to construct for regional climate modelers, we incorporated the scientific workflow graph as a central representation of the RCM-PSE, where modelers would electronically draw workflow graphs and then use them to manage and semi-automate their computational experiments (see Figure 2). With this concrete computer representation of scientific workflow, modelers were able to view, control, and manipulate their own work and research processes.

As shown in Figure 3, the Visual Modeling Environment for Biology (VMEB)[23] allows biologists and bioinformaticists to graphically construct concepts, hypotheses, and theories using 2D drawing tools. Unlike typical drawing software, however, VMEB allows biologists to attach

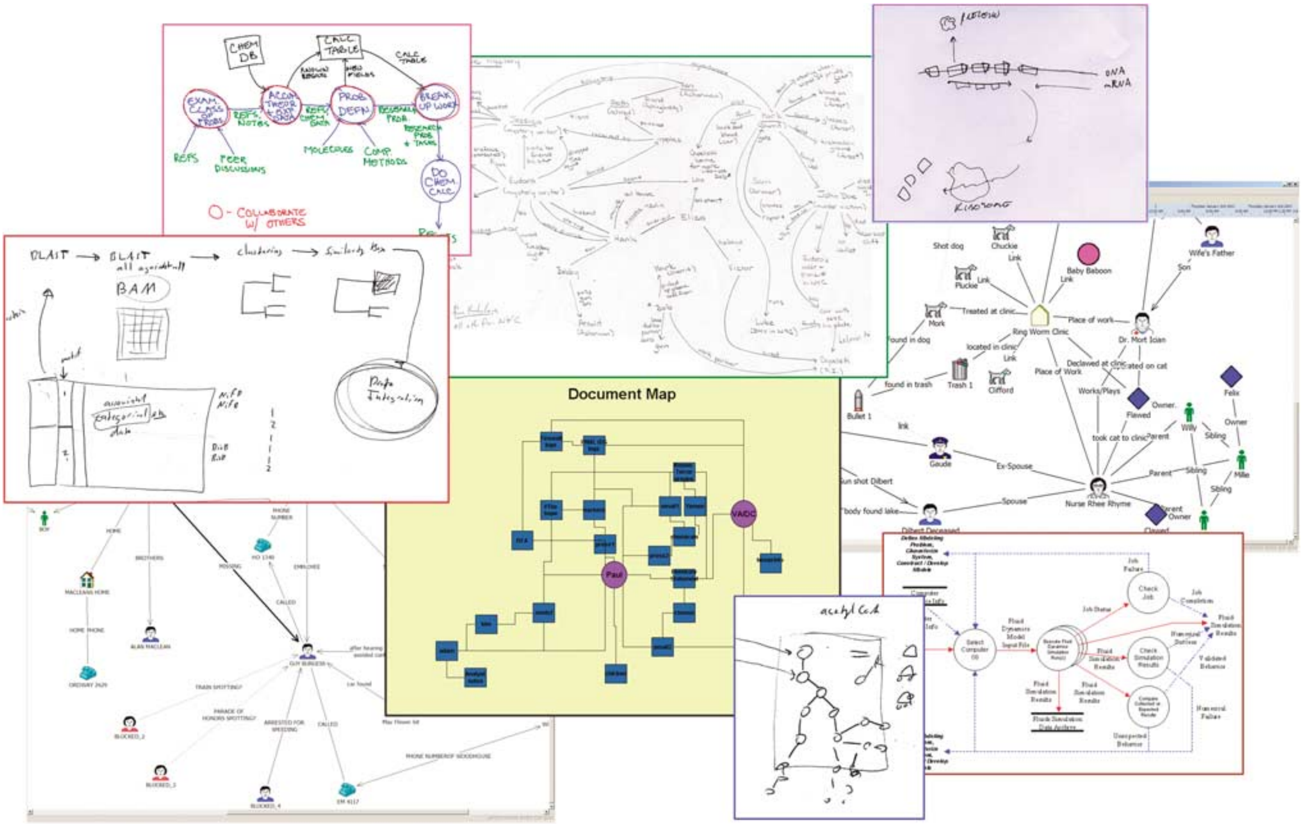

Figure 1 Examples of researcher-constructed diagrams conveying scientific and analytical concepts and processes.

semantics to individual and collections of graphics objects. For example, a yellow-filled box might indicate a protein; similarly, a graphical collection of four proteins linked together in a particular pattern with an arc may identify a chemical reaction. Semantics are captured from the biologist as a set of rules that infer biological meaning from the graphical and spatial relationships of objects. For each concept, the biologist may also define standard properties associated with that concept. Once the rules base has been populated, VMEB may automatically attach biological meaning to graphical objects as they are constructed by the biologist in the drawing area by traversing and applying the rules in the rules base.

Link analysis is a popular form of analysis where analysts graphically draw out and link different kinds of information such as events, peoples, places, locations, facts, weapons, money, and contraband. The objective of the Scenario and Knowledge Framework for Analytical Modeling (SKFAM)[24] system is to capture and represent these link analysis diagrams as scenarios that may be computationally compared to one another. Graph-based scenarios of past or current cases may be generated and placed in a reference case library. As a new situation arises, it also may be represented as a scenario and compared against the scenario case library to identify other similar or relevant cases. As shown in Figure 4, the graph of the evolving scenario could then be mapped upon each of the matching scenarios from the case library to identify potential missing nodes and links or to predict possible results and outcomes based on prior cases.

In comparing two scenarios, SKFAM performs subgraph pattern matching on the underlying graph representations. Rather than arbitrary subgraph patterns, however, SKFAM searches for graphical patterns that have specific meaning and importance to analysts. Specific subgraphs patterns may distinguish critical analysis themes, such as

- a hierarchical communication structure among a number of person nodes might identify a command structure,
- financial transactions across people and institution nodes where the amount of money changing hands successively decreases might indicate a money laundering scheme,
- nodes with many edges (high degree) or singular or weak links that tie two or more portions of the graph together (bridges) might identify potential points of disruption.

For the most part, scientific or analytical knowledge is generally maintained and managed in a transient form in the analyst's mind. An important objective of our R&D efforts is to migrate scientific or analytical knowledge from

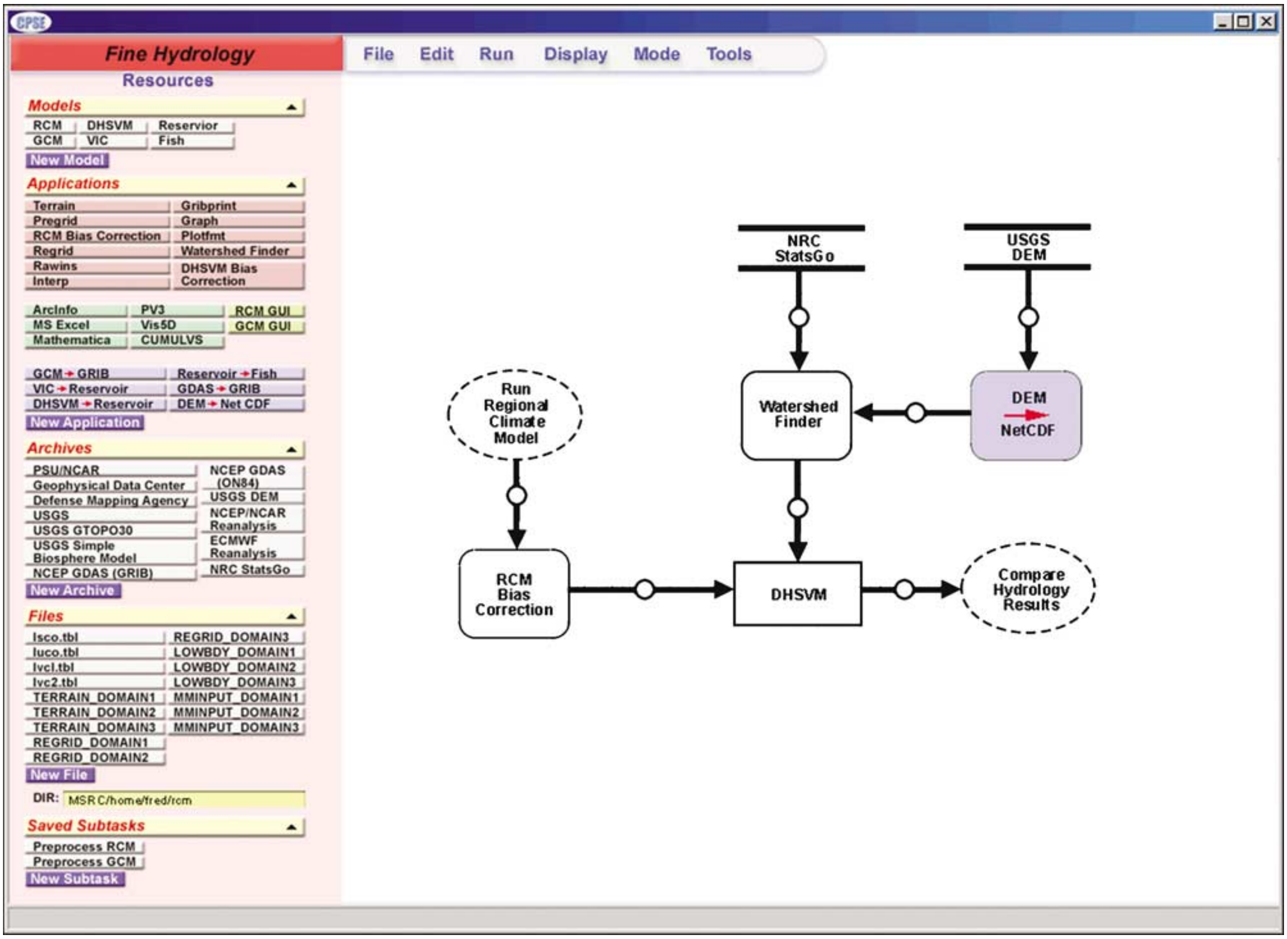

Figure 2 Workflow in regional climate modeling problem-solving environment.

its ethereal form to one that is concrete, usable, shareable, and computable. The systems and interfaces described above establish and promote visual, graph-based knowledge representations, and interaction paradigms that are rooted in cognitive and perceptual principles, support analytic reasoning on complex and dynamic problems, encapsulate and synthesize large amounts of data and information, and facilitate sharing and dissemination of concepts, analyses, and theories through the use of fundamental and intuitive visual representations.

Network extraction from unstructured sources

Manually creating diagrams is a laborious process. In comparing subgraph patterns (diagrams), it is also useful to automatically derive subgraph patterns from free text. At PNNL, we have developed a methodology for automatically generating contextual social network maps from free text. In such maps, named entities are clustered according to the events or contextual background in which they were mentioned in the documents, while links represent specific interactions between the entities. The methodology combines robust natural language processing for extracting named entities and their interactions and statistical analysis for uncovering contextual background information. Furthermore, ontologies can be supplemented to restrict the generation process to a particular domain of interest.

Our approach attempts to alleviate some of the issues traditionally encountered in network analysis and visualization. First, the generation process is entirely automated in order to sift through large amounts of unstructured text. However, this process allows for manual curation to correct potential errors and enrich the various signatures with critical information not contained in the set of documents being analyzed. Second, most network visualizations favor counter-intuitive representations of the social network structure by organizing the graph layout according to criteria such as minimizing link overlaps. This often leads to a representation where highly connected entities that are close end up far apart from each other. Furthermore, node location has no meaning *per se* and makes map comparison and communication challenging. Finally, automated methods for social networks rely either on natural language processing alone or statistical analysis

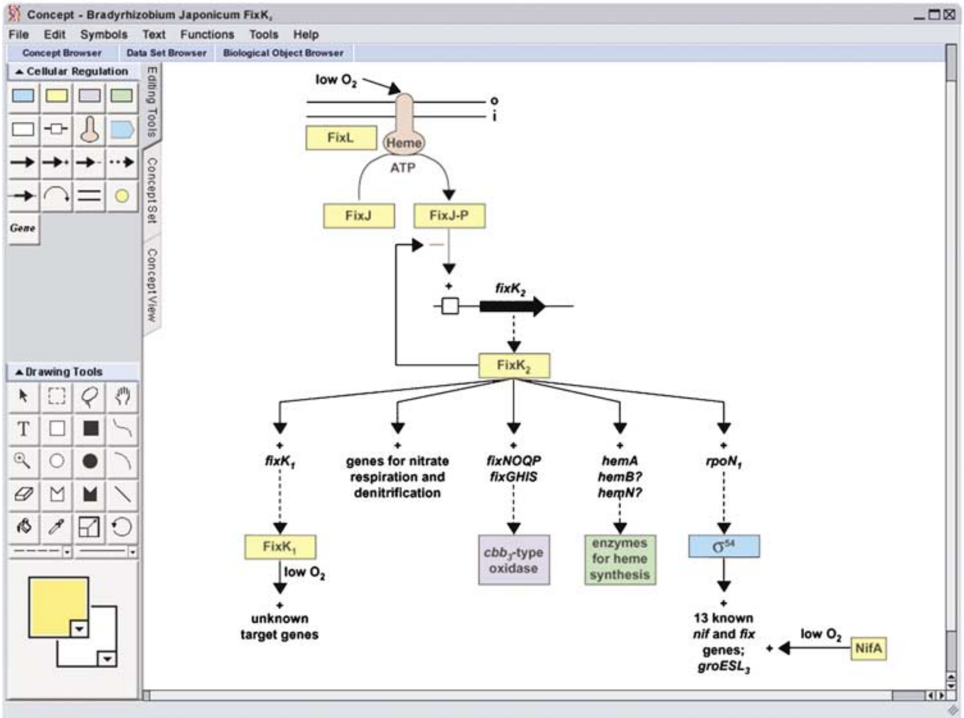

Figure 3 Drawing biology diagrams with Visual Modeling Environment.

alone. In the first case, latent background information is not captured and the larger picture is difficult to grasp. In the latter case, interpretation and understanding require significant post-analysis.

Our approach consists of the following steps executing in a single scan of the input:

(1) *Named entities recognition* (NER) and *co-reference resolution* algorithms are applied to incoming text.
(2) Whenever a named entity (person, organization, location, etc.) is recognized, two information extraction processes occur.

   (a) A *text window* around the named entity is selected and added to the *pseudo-document* associated with the named entity. The text window is tunable and may consist of the sentence in which the named entity is found, this sentence plus a set of surrounding sentences, or a sentence-free window that consists of a chunk of text determined by the number of terms before and after the named entity to be included in the window. If the named entity is encountered for the first time, then the text window initializes its pseudo-document.
   (b) An *event pattern* involving the named entity is added to the *set of event patterns* already extracted for that entity. The pattern is obtained via a syntactic parser and consists of a triplet of the form (NE_i, AT+, NE_j^*) where NE_i denotes the current named entity; NE_j denotes another named entity; and AT denotes an attribute such as a lemmatized verb or the member of a selected set of properties (e.g. sister, father, sarin, bomb, weapon, Indonesia, London). The star denotes zero or more consecutive occurrences, while the plus sign denotes at least one or more consecutive occurrences. For example, valid triplets are (Amrozi, meet, Abu Bakar Bashir), (Amrozi, arrive in, Malaysia), (Murai, make, sarin). If the named entity is encountered for the first time, the set of patterns is initialized with the pattern just extracted.

   Each extracted named entity is accompanied with a *functional signature* made of two components: a pseudo-document and a set of event patterns. The pseudo-document provides the context in which the entity acts, while events detail actions of the entities.

(3) The social network map is obtained with the pseudo-document signatures. First, each pseudo-document is represented with its bag-of-words vector representation. Second, a powerful dimension-reduction technique, projection pursuit, is applied to the vectors to provide a clustered view of the entities. Entities that appear together in the map have been involved in the same contextual background, while entities that are far apart were engaged in different circumstances.
(4) In order to understand and interpret the closeness of entities in the social network map, links are added

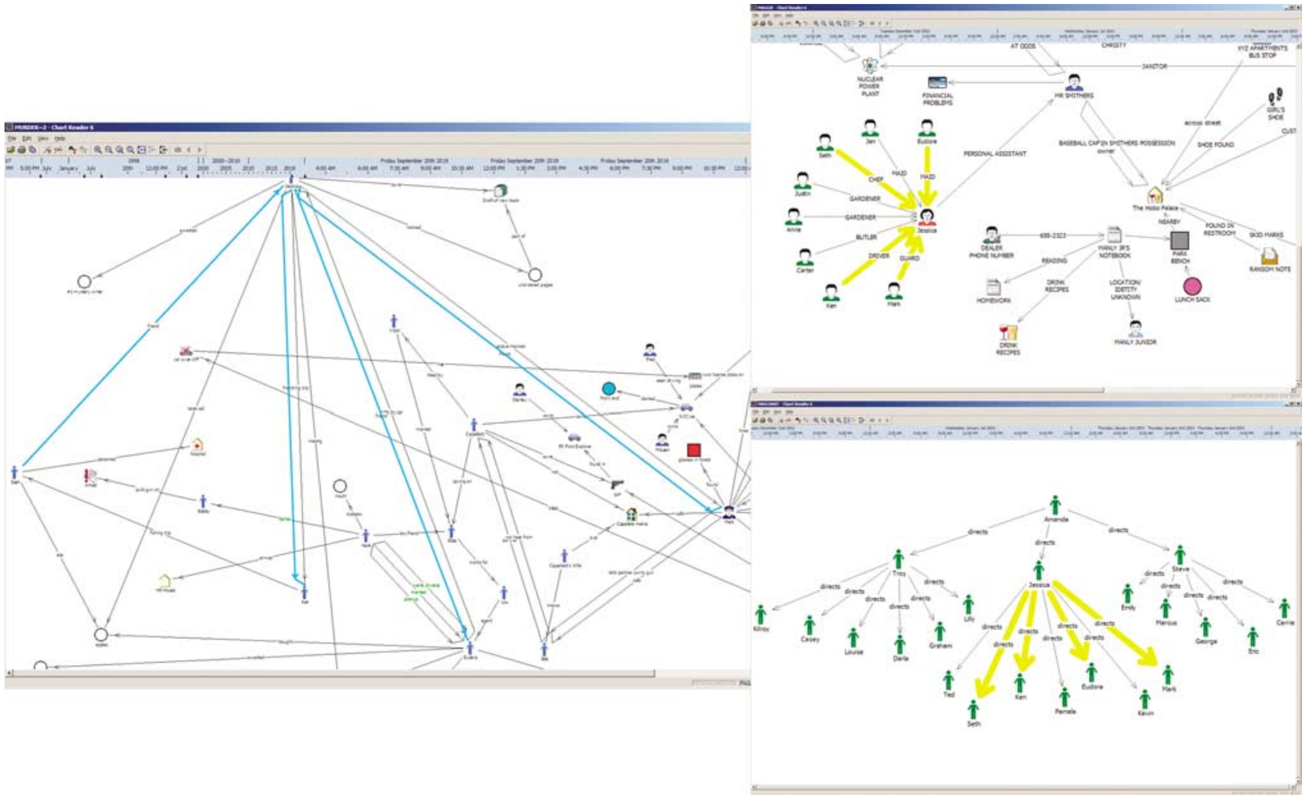

Figure 4 Query subgraph (highlighted in blue in left window) matches two scenarios from the case library (matching subgraphs highlighted in yellow in right windows).

between entities that are sufficiently close to each other. This is carried out as follows. For each entity, a local neighborhood is chosen (think of a circle surrounding the entity in the map). For each entity captured in the local neighborhood, a search is performed in the patterns associated with the entity for patterns involving both entities. If such a pattern is found, a link is displayed in the map, with the attribute AT displayed next to the link.

The contextual social network map is interactive, and for each named entity, users can access the set of event patterns and the pseudo-documents to gain more insight.

We illustrate this approach (see Figure 5) with a set of 14 news stories that were selected between October 17, 2002, and November 29, 2002, for their focus on terrorist activities in Southeast Asia at the turn of the century. Ninety-one named entities were extracted from those files, which led to the contextual social map displayed in Figure 5. The map reveals four clusters. The right cluster deals mostly with political entities in Indonesia; the bottom cluster involves many law enforcement entities. The remaining two clusters concern terrorist activities. The first cluster focuses exclusively on one particular incident, the Manila bombing on December 30, 2000. The second cluster concentrates on the relationship between Al-Qaeda and the Jemaah Islamiyah. The cluster interpretation was obtained by investigating the entities that comprise them and reading their pseudo-documents. The closeness of the last clusters reflects their strong relationship. For example, the Manila bombing was orchestrated by Jemaah Islamiyah, who also used the alias Hambali. It is also known that Amrozi and Al-Ghozi met, and so on.

Focusing on the Manila Bombing partition in Figure 5, the roles and relationships among the characters were identified and annotated (in blue) in Figure 6. In 2003, Hambali and Bafana were indicated on the basis of confessions and evidence that they had conspired in the plot that resulted in the 2000 bomb blast. Hambali is now imprisoned in Jordan in US custody. The two allegedly met Al-Ghozi a week before the blast. Al-Ghozi was arrested a month later in Manila but escaped in July 2003.[25]

Obviously, the above analysis could be refined. For example, we did not apply a filtering to the list of extracted named entities in the social network map. Hence, content providers such as Kyodo News or The Sydney Morning also appear in the map as such sources are often quoted in these documents. It would be a trivial task to add such a refinement that could be either user-defined or data-driven. Our approach could also be enhanced by integrating ontological knowledge representation of entity roles and purpose.[26]

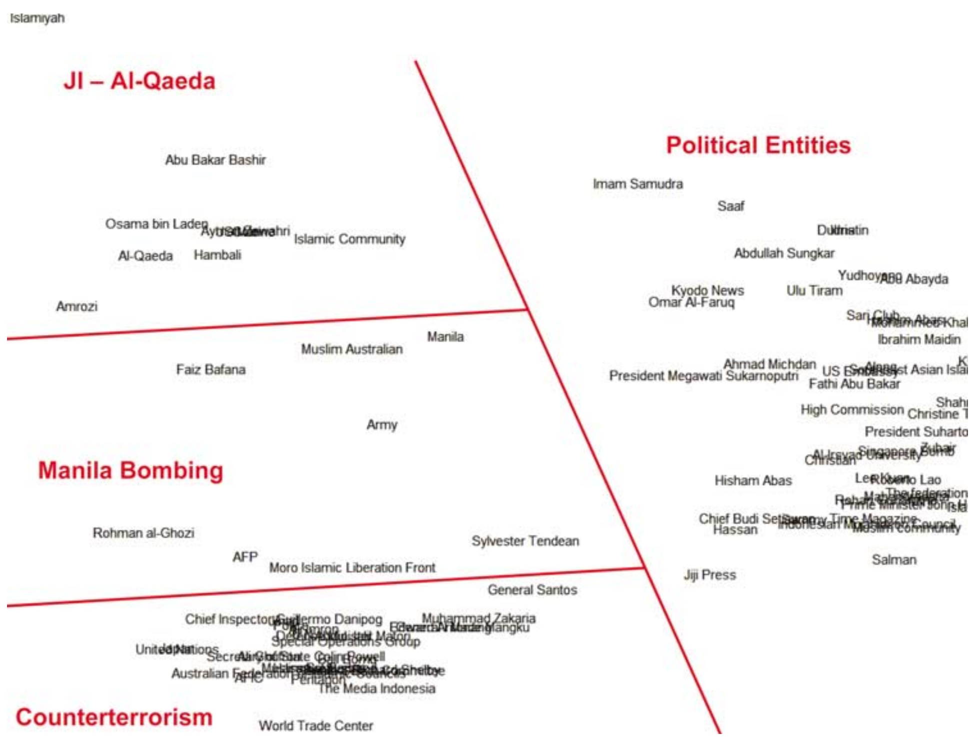

Figure 5 Contextual social network map showing terrorist activities in Southeast Asia at the turn of the century.

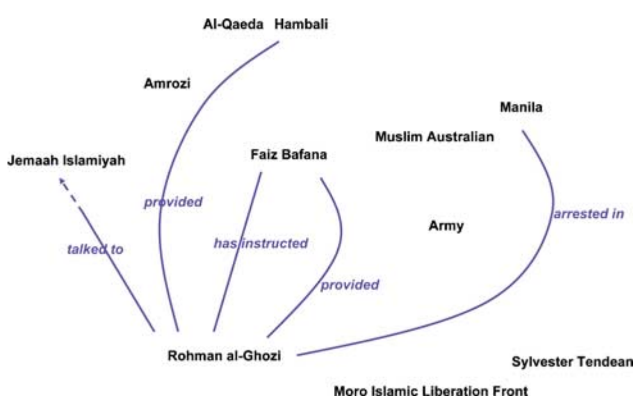

Figure 6 Manila bombing graph, with links between entities.

Parallel analysis of competing scenarios

The ability to support simultaneous creation and validation of competing scenarios is a major challenge for today's analysis systems. Parallel Analysis of Competing Scenarios (PACS) is needed to avoid premature commitment to a single expected outcome with consequent neglect of relevant evidence relative to other plausible outcomes.[27]

PACS is a difficult analytic technique to perform without computational aid. As a result of memory limitations on human cognition,[28] most people are unable to retain several hypotheses and relevant supporting evidence in working memory. Moreover, supporting evidence for competing scenario analysis needs to be distilled from potentially huge data repositories. Such a task would require extravagant expenditure of human resources without the help of machine-aided information extraction and analysis processes (and would not result in timely actions).

Current analysis tools provide some of the functionality needed to support PACS. For example, link analysis tools (e.g. Analyst's Notebook)[29] allow users to build scenarios manually (or automatically from structured data) and display graphical views of the timeline and network information encoded. Chappell *et al*.[30] describe a system that allows users to extract facts from document sets automatically, provides timeline and network visualization of the information extracted, and presents an environment that supports scenario construction. Fikes *et al*.[31] discuss how this system can be extended to handle alternative scenarios through (1) the use of 'contexts' in which each hypothesis is developed independently and (2) the determination of relationships across contexts to detect incompatibilities and common features among alternative scenarios.

However, no analytic tool is available yet that enables the automatic retrieval of facts supporting, contradicting, or originating a given event[32] to sustain analysis techniques such as PACS. In addition, large event network displays tend to be unmanageable unless effective ways of condensing and compacting nodes that encode compatible information are available. Graph summarization techniques that are semantically driven are therefore

Figure 7 PNNL's PACS environment, showing Document Space, Evidence Marshaling Space, and Hypothesis Space.

necessary to distill the most significant patterns from large event networks. Such functionality is not yet available in link analysis tools or similar toolkits.

The work carried out at PNNL addresses this gap through the development of a visual interactive environment for event analytics that integrates text analysis, discourse analysis, and visualization capabilities to support the creation and validation of competing scenarios.

Using insights from recent advances in the theory and practice of analysis[33,34] and related government-funded R&D work,[35] we have designed an analytical environment intended to support the analytic workflow underlying PACS.[36] The conceptual design is realized through three 'spaces' – each corresponding to a step in the analytical cycle – represented on individual monitors (or distinct screen partitions), as shown in Figure 7.

The first screen 'Document Space' gives the analyst the opportunity to interact with a document corpus of interest and replicates the data collection process in the analytical workflow. Using IN-SPIRE^{TM} ,[37] documents are clustered based on their content. A two-dimensional Galaxy view of all the documents is created, with points representing each document and clusters showing how those documents are related. The analyst may interact with the document collection at this level, zooming in and out to get a better appreciation of what documents are included and their main themes. Additional mechanisms can be used to select documents, including full-text search and 'query by example'. When ready, the analyst may select a subset of the documents for further analysis in the Evidence Marshaling Space.

The purpose of the second (middle) screen, the Evidence Marshaling Space, is to allow analysts to work in a highly flexible environment, where they can (1) look at the entities, relationships, and events expressed within a selected set of documents, and (2) manipulate them, move them around, and view them from alternative angles. Using this window, users define the level of granularity desired for investigating the information. Entities and events are represented in the form of a link-node graph, as shown in Figure 8. At this stage, the user can abstract away from individual events by generalizing over subsets of events using the inheritance relationships and class-instance links encoded in the event ontology. Each event class (e.g. *express*, *act* in Figure 8) subsumes a number of events that occur as instances or subclasses of the event class in the ontology. Users can choose to cluster events by class or to display only event classes to have a bird's-eye view of the entire Evidence Marshaling Space, and then narrow their investigation by focusing on the entities.

Relationships between individual events are elucidated through the use of rhetorical relations. We use rhetorical relation classes (e.g. *contrast*) to generalize over sets of relationships between events (e.g. *but*, *however*, *yet*). Figure 9 provides an abstract representation of how the entity, event, and rhetorical relation layers are interleaved in terms of class inheritance and class-to-instance relationships.

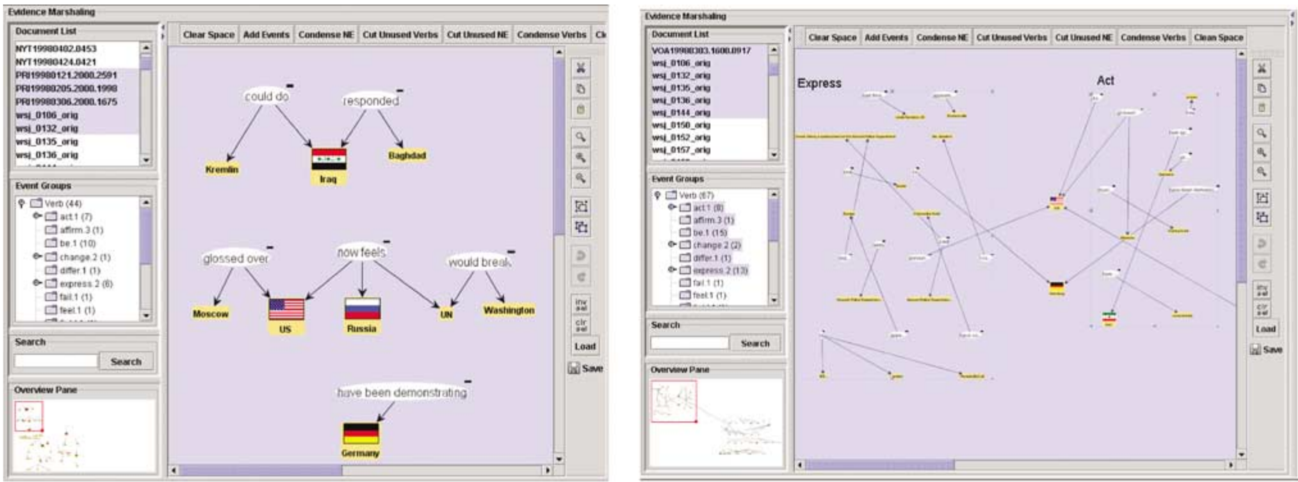

Figure 8 The Evidence Marshaling space of the PACS environment.

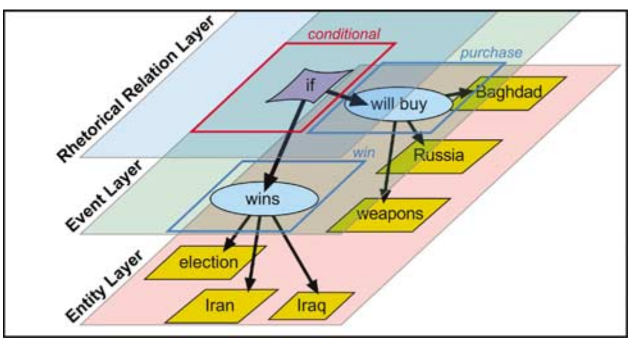

Figure 9 Abstract representation of how the entity, event, and rhetorical relation layers are interleaved.

The final screen, the Hypothesis Space, gives the analysts a natural interface to structure their analytical product, by building up their arguments through the support of evidence (e.g. documents and semantic graphs) marshaled in the previous two screens. Items from the other screens can be dragged and dropped from other screens to make these linkages. It provides an environment in which analysts can construct and verify plausible scenarios. Users can investigate an information path, retreat, and follow a new path, while constantly updating and reorganizing their hypothesis graph to create a final appreciation of the analytical results. The validation of the scenarios constructed is performed using an indications-based estimation of outcomes from observed evidence, following the Prospective Analysis approach described by Turner *et al*.[38]

Dynamic analysis in group environments

The R&D agenda[11] has clearly suggested the importance of deploying analytical capabilities outside the traditional, individual workstation. Modern environments require fast-paced and dynamic access to information. Visualization tools have been optimized primarily for conventional desktop displays and interaction devices, which mostly support individual interaction. Transitioning from individual spaces to group spaces requires support for new workflows and interactions. Display technologies in group spaces also span the range of limited personal digital assistants (PDAs) to ultra-high-resolution full-wall and theater-sized displays in which many people can interact.

Technology-enabled mobility has value to the degree that it increases productivity, enhances ability for individual expression, provides a sense of safety and security, enables ubiquitous exploration and discovery, and supports aspirations for communal belonging.[39] There is a wide variety of information needs dependent on the type of user and the task to be performed. This requires a careful scoping of the problem space being addressed and a realization of whether and how the results apply to other situations.[40] Interfaces need to support a range of techniques to accomplish the same task in less favorable contexts (e.g. while walking, in dark environments). Selection of the appropriate interaction technique depends on the needed level of control and situational capability of the user to use the interface.[40]

For the last several years, PNNL has been exploring use of large-screen, group interaction workspaces to complement our research in visual analytics. Part of that effort has focused on how to support new and innovative interactions with large-screen systems. We have developed the Human Information Workspace (HI-Space) which uses video tracking to recognize multiple simultaneous hands and objects as inputs for a wide variety of applications using large displays (see Figure 10).

The HI-Space system consists of two separated displays, one on a table surface and the other on a large wall screen. Tracking is performed by using an overhead

Figure 10 HI-Space users interacting with 3D geometry data generated from 48 confocal microscope image slices at 512×512 resolution.

camera centered over the table surface. This creates an interaction volume over the table, tracking everything that is placed on or reaching over the table. The video-based tracking of multiple hands and objects is used in conjunction with traditional mouse and keyboard inputs to provide a robust interaction research environment. The separate vertical wall display is set up for both mono and active stereo projection. The wall, however, is primarily only a display surface with limited interaction support.

One of our ongoing projects is to develop applications for human-centered environments that explore biological data and to evaluate the effectiveness of these applications and environments for real-world tasks. A 'human-centered' environment uses both interactions and visualizations that take into account human performance and capabilities more fully than current desktop interfaces. Human-centered environments can support wide-field-of-view, stereoscopic display, and multimodal interaction including gesture-based interactions. Such environments have the potential to present visual information that is easier to understand with natural and efficient controls for interactively exploring and analyzing complicated time-varying data. This work is being carried out in partnership with Brown University and the University College of London. Our goal is to better understand the strengths and weaknesses of human-centered environments and to develop an initial set of guidelines for future research.

The range of information quantity and complexity that can be displayed on a PDA is dramatically different from that for a large, multi-screen environment. However, there are promising characteristics of mobile devices and how people use them. For the past two years, PNNL has deployed a mobile-based application to provide real-time information for attendees of the Supercomputing Conferences[41] in 2004 and 2005 with an application called Infostar.

The goal of Infostar was to provide Supercomputing attendees real-time data on all aspects of the conference. This included information not available through other means, such as exhibitor presentation schedules, and capabilities such as full-text queries of all conference information that only existed on Infostar. Characteristics of the application environment required that Infostar be easy to load and run on as many mobile devices as possible. There was no feasible way to distribute software or train users. For these reasons, Infostar was developed with common protocols for a majority of the application. The visual explorer was implemented using Macromedia's Flash™. The web pages were designed to provide several ways to access any piece of information. Users could search for events by type (papers, posters, panels, exhibitors, Birds of a Feather sessions, and more), a keyword query, or date and receive a list of relevant events. Each event listed could then be expanded to see details or the event's location on a map. From any event listing, further lists could be generated. Given an event, a list could be generated based on several common attributes like showing all events that took place on the same day, in the same room, or of the same type. This provided the capability to rapidly access a class of events from different starting points, minimizing the need to move through numerous web pages.

To further help users access and understand the relationship between various events at the conference, we used the IN-SPIRE[42] engine to analyze relationships between the events. This meant that in addition to the traditional keyword or date searches, attendees could look for related events regardless of the event type, time, or track it was assigned. Access to this functionality was provided by a text list and by a visual exploration interface.

The visual exploration tool provided a simplified view of the IN-SPIRE Galaxy visualization. As the mobile devices have limited screen real estate, representing documents individually in a single view was not reasonable. The technique provided by Infostar displayed groups of similar documents, based on the IN-SPIRE clustering, as circles with key terms to represent the topics of each group along with the number of documents in each group displayed in each circle. The number of circles could be dynamically selected by users depending on the device they were using. For example, if the device had a small, low-resolution screen and it was being read while walking, users could set the display to a small number of clusters. The use of circles was chosen to maximize the size, which allowed easier selection by using a stylus or finger on touch-sensitive screens. Drilling down into the data was accomplished by selecting a circle. This would sub-select only the documents associated with that circle and break them into a new set of circles (clusters). Users repeated this selection process until they had drilled down to a sufficiently small set of documents. This reduced set of documents could then be viewed as a list and displayed.

As we now look to new applications for mobile visual analytics, the challenge is to adapt different types of information, visualizations, and interactions to fit specific domains.

Conclusions

The new discipline of visual analytics addresses many information understanding problems through interactive visual interfaces. It demands an ambitious undertaking by the IT community to examine today's shortcomings and challenges for tomorrow's solutions. The PNNL technology presented in this paper focuses mainly on improving the value of visual analytics for those who engage in complex cognitive tasks in a multidisciplinary collaborative environment. Its ultimate goal is to improve the capabilities for extracting knowledge from, and reasoning about, information, knowledge, and data. This paper covers only a limited number of technologies that promise to revolutionize the exciting new field of visual analytics. However, with the strong support from both government and industrial sectors, we hope this community will rise to the challenges and develop novel ideas and practical tools as outlined by Thomas and Cook.[12]

Acknowledgments

The preparation of the paper was supported by the National Visualization and Analytics Center™ (NVAC™) located at the Pacific Northwest National Laboratory in Richland, WA. Special thanks go to Kris Cook, Lee Ann Dudney, Sharon Eaton, Beth Hetzler, and Amanda White who provided assistance of many forms throughout this work. The Pacific Northwest National Laboratory is managed for the US Department of Energy by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

References

1 MSN Encarta Encyclopedia. Galileo Galilei. http://encarta.msn.com/encyclopedia_761557587/Galileo.html (access 16 July 2006).
2 MSN Encarta Encyclopedia.Heliocentric. http://encarta.msn.com/encyclopedia_761586364/Heliocentric_System.html (access 16 July 2006).
3 MSN Encarta Encyclopedia. Gregor Mendel. http:// encarta.msn.com/encyclopedia_761562358/Gregor_Mendel.html (accessed 16 July 2006).
4 National Human Genome Research Institute. Mendelian Inheritance. http://www.genome.gov/glossary.cfm?key = mendelian %20inheritance (accessed 16 July 2006).
5 Tufte ER. *The Visual Display of Quantitative Information*. Graphics Press: Cheshire, CT, 1983.
6 Anscombe F. *Graphs in Statistical Analysis*, The American Statistician, 195–199, 1973.
7 Exploratory Data Analysis. http://www.itl.nist.gov/div898/ hand book/eda/section1/eda16.htm (accessed 16 July 2006).
8 Tukey JW. *Exploratory Data Analysis*. Addison-Wesley: Reading, MA, 1977.
9 Card SK, Mackinaly J, Shneiderman B. *Reading in Information Visualization: Using Vision to Think*. Morgan Kaufmann: San Francisco, CA, 1999.
10 Wong PC, Thomas JJ. Visual Analytics. *IEEE Computer Graphics & Applications* 2004; **24**: 20–21.
11 National Visualization and Analytics Center (NVAC). http://nvac.pnl.gov (accessed 16 July 2006).
12 Thomas JJ, Cook KA (Eds). *Illuminating the Path*. IEEE Computer Society Press: Los Alamitos, CA, 2005.
13 Wong PC, Thomas JJ (Eds). *IEEE Computer Graphics & Applications* (Special Issue on Visual Analytics) 2004; 24.
14 IEEE Transactions on Visualization & Computer Graphics (TVCG). http://www.computer.org/tvcg (accessed 16 July 2006).
15 American Association for the Advancement of Science annual meeting (AAAS) 2005. http://www.aaas.org/meetings/Annual_Meeting/ (accessed 16 July 2006).
16 2005 International Conference on Intelligence Analysis. May 2005. https://analysis.mitre.org/proceedings/toc.html (access 16 July 2006).
17 IEEE Symposium on Information Visualization (IEEE InfoVis) 2005. http://www.infovis.org/infovis/2005/ (accessed 16 July 2006).
18 IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST) 2006. http://conferences.computer.org/vast/vast2006 (accessed 16 July 2006).
19 Tergan SO. Managing knowledge with computer-based mapping tools. In: *Proceedings of the ED-Media 2003 World Conference on Educational Multimedia, Hypermedia & Telecommunication* (Honolulu, HI), AACE: Norfolk, VA. 2003; 2514–2517.
20 O'Donnell AM, Dansereau DE, Hall RH. Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review* 2002; **14**: 71–86.
21 Chin G Jr., Leung LR, Schuchardt KL, Gracio. New paradigms in collaborative problem solving environments for scientific computing. *Proceeding of the 2002 International Conference of Intelligent User Interfaces* (San Francisco, CA), *ACM Press*: Neq York. 2002; 39–46.
22 Chin G Jr., Schuchardt KL, Myers JD, Gracio DK. Participatory workflow analysis: Unveiling scientific research processes with physical scientists. *Proceedings of Participatory Design Conference* (New York, NY), *CPSR*: Palo Alto, CA. 2000; 30–39.
23 Chin G, Stephan EG, Klicker KR, Corrigan AL, Sofia HJ. Supporting computational visual theories in biology. In: *Proceedings of the 2004 IEEE Symposium on Visual Languages and Human-Centric Computing* (Rome, Italy), IEEE Press: New York. 2004; 69–71.
24 Chin G Jr., Kuchar OA, Whitney PD, Powers ME, Johnson KE. Graph-based comparisons of scenarios in intelligence analysis. *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics* (The Hague, The Netherlands). IEEE Press, 2004; 3175–3180.
25 Cable News Network (CNN). Manila: Key Terror Suspect Killed. http://edition.cnn.com/2003/WORLD/asiapcf/southeast/10/12/ghozi.killed/index.html. October 12, 2003 (accessed 16 July 2006).
26 Fan J, Barker K, Porter B, Clark P. Representing Roles and Purposes. *Proceedings of the 1st International Conference on Knowledge Capture* (Victoria, BC, Canada), 2000; 38–43.
27 Heuer Jr. RJ. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency: Washington, DC, 1999.
28 Miller G. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review* 1956; **63**: 81–89.
29 Analyst's Notebook. http://www.i2inc.com/Products/Analysts_Notebook (accessed 16 July 2006).
30 Chappell AR, Cowell AJ, Thurman DA, Thomson JR. Supporting Mutual Understanding in a Visual Dialogue between Analyst and Computer. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (New Orleans, LA), 2004.
31 Fikes R, Ferrucci D, Thurman D. Knowledge Associates for Novel Intelligence (KANI). In: *Proceedings of the 2005 International Conference on Intelligence Analysis* (McLean, VA), 2005.
32 Marcu D, Echihabi A. An Unsupervised Approach to Recognizing Discourse Relations. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, Philadelphia, PA, 2002.
33 In: Dearth D and Goodden R (Eds). *Strategic Intelligence: Theory and Application*. 2nd ed., Joint Military Intelligence College: Washington, DC. 1995.
34 Krizan L. *Intelligence Essentials for Everyone*. Joint Military Intelligence College: Washington, DC, 1999.
35 Thurman DA, Cowell AJ, Andrew AH, Chappell AR. Human-information Interaction with Knowledge Associates. In:

Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society (Denver, CO), 2003.
36 Sanfilippo A, Baddeley B, Cowell A, Gregory M, Hohimer R, Tratz S. Building a Human Information Discourse Interface to Uncover Scenario Content. *Proceeding to the 2005 International Conference on Intelligence Analysis* (McLean, VA), 2005.
37 INSPIRE. http://in-spire.pnl.gov (accessed 16 July 2006).
38 Turner A, Hetzler E, Cheney B, Williams L, Zabriskie S. Method for Prospective Analysis of Alternative Futures. *Proceedings to the 2005 International Conference on Intelligence Analysis* (McLean, VA), 2005. https://analysis.mitre.org/proceedings/ (accessed 16 July 2006).
39 Stoop E. Mobility: Freedom of Body and Mind. In: Aarts E and Marzano M (Eds). *The New Everyday: Views on Ambient Intelligence*. 010 Publishers: Rotterdam, The Netherlands. 2003.
40 Campbell C, Tarasewich P. Creating Notification Cues for Mobile Response Teams. *Proceedings of the 2004 Decision Sciences Institute 2004 Annual Meeting*, 2004.
41 Supercomputing Conference. http://www.sc-conference.org (accessed 16 July 2006).
42 Analysis Experiences Using Information Visualization, Hetzler E, Turner A. *IEEE Computer Graphics and Applications* 2004; **24**(5): 22–26.