

# Traumatology

<http://tmt.sagepub.com>

---

**A Digital Library for Recovery, Research, and Learning From April 16, 2007, at Virginia Tech**

Edward A. Fox, Christopher Andrews, Weiguo Fan, Jian Jiao, Ananya Kassahun, Szu-Chia Lu, Yifei Ma, Chris North, Naren Ramakrishnan, Angela Scarpa, Bruce H. Friedman, Steven D. Sheetz, Donald Shoemaker, Venkat Srinivasan, Seungwon Yang and Laura Boutwell

*Traumatology* 2008; 14; 64 originally published online May 16, 2008;

DOI: 10.1177/1534765608315632

The online version of this article can be found at:  
<http://tmt.sagepub.com/cgi/content/abstract/14/1/64>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Traumatology* can be found at:**

**Email Alerts:** <http://tmt.sagepub.com/cgi/alerts>

**Subscriptions:** <http://tmt.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** <http://tmt.sagepub.com/cgi/content/refs/14/1/64>

# A Digital Library for Recovery, Research, and Learning From April 16, 2007, at Virginia Tech

Edward A. Fox, Christopher Andrews, Weiguo Fan, Jian Jiao, Ananya Kassahun, Szu-Chia Lu, Yifei Ma, Chris North, Naren Ramakrishnan, Angela Scarpa, Bruce H. Friedman, Steven D. Sheetz, Donald Shoemaker, Venkat Srinivasan, Seungwon Yang, and Laura Boutwell

The authors are developing a digital library to help during the long-term recovery from the mass shooting on April 16, 2007 at Virginia Tech. Content comes from uploaded texts, images, videos, and other files, as well as pages crawled from the Web and information collected, with permission, from those working with Web 2.0 sites such as Facebook and Flickr. The authors are applying data/text mining, social network analysis, and information visualization methods to facilitate systems science, providing key added-value services, especially to social and behavioral

scientists seeking faster and easier ways to analyze, model, understand, and test hypotheses. The article also considers how technologies influence communications. The authors aim to support the university, local community, region, nation, and the world, as it seeks to learn from a tragic event.

**Keywords:** digital library; Virginia Tech; Web; Web 2.0; technation

This article is about our vision, early work, and plans to help with the recovery from the mass shooting on April 16, 2007, at Virginia Tech—to support research and learning using a digital-library-based approach. We also hope that this will lead to faster and more effective methods for applying the social and behavioral sciences to address tragic events and their aftermath, much as advanced techniques have been applied to the natural sciences in e-science initiatives (Atkins et al., 2003).

The backdrop of our work is the set of April 16–related experiences of the 16 people who have contributed to this article. We all decided to express our care through hard work, as a result of feeling a constructive sense-making impulse to collect related

information, organize it, make it accessible, and preserve it for the future. In the interest of brevity, we provide just two representative comments on our personal experiences.

1. Sidebar 1 covers some of the experiences of the first author, a Professor of Computer Science and Director of the Digital Library Research Laboratory.
2. Sidebar 2 was written by the lead social scientist in our group, Professor Shoemaker, from Sociology, and Director of the Center for the Study of Violence in Society.

The context of our work is a world in which radical changes have occurred as a result of the convergence of many innovations and flattening trends, including globalization; advanced information technologies; openness (Raymond, 1999; Van de Sompel & Lagoze, 2000); collaboration, competition, and workflows; new economies; and sociopolitical

---

From Virginia Tech, Blacksburg, Virginia.

Address correspondence to: Edward A. Fox, PhD, Virginia Tech, 2050 Torgersent Hall, M/C 0106, Blacksburg, VA 24061; e-mail: [fox@vt.edu](mailto:fox@vt.edu).

My e-mail folder related to April 16 has more than 2,000 messages. Before noon on April 16, a colleague in Rio had inquired to see if Gail McMillan (Director of Digital Library and Archives, in the University Libraries) and I were okay. Other concerned e-mails, from family, friends, and colleagues, came by April 17 from all parts of the United States, as well as from Australia, Brazil, Canada, Germany, Greece, India, Israel, Italy, Japan, Mexico, Netherlands, New Zealand, Nigeria, Singapore, Spain, Thailand, and the United Kingdom. Particularly poignant was a cell phone call from a cousin in Colorado, who worked for the sheriff in a town near Columbine High School.

That connection with another tragic event fueled my resolve to apply my research and teaching skills to help with the recovery from April 16. I had been frustrated on September 11, 2001, when I was attending the leading international conference on information retrieval, and, in discussions with people in various branches of the government, including the intelligence community, offered to assist—to no avail. But on April 16, 2007, two program directors at the National Science Foundation (NSF) sent me e-mail. Fortunately, on Wednesday, April 18, a number of groups at NSF were able to meet with me, in their offices in Arlington, Virginia, describing some of the support given for research related to prior tragedies. There was encouragement and interest at Virginia Tech in such research, expressed by the Head of the Department of Computer Science, the Vice President for Information Technology, and a number of colleagues. The first meeting of our research team related to this project was at 10 a.m., April 23, shortly after the balloons were launched over the Virginia Tech Drillfield in memory of the victims. We decided to work to develop a digital library related to the April 16 tragedy, and, since then, have broadened that interest to work on digital library support for the broad field of crisis, tragedy, and recovery. This article covers much of our thinking and some of the initial research activities that relate.

In addition to research, teaching activities have provided opportunities to further understanding about the April 16 tragedy. On April 23, our graduate class on information retrieval began to brainstorm on how to apply what was being learned that semester to help manage the information related to April 16. An undergraduate student, Ananya Kassahun, one of the coauthors of this article, agreed to prepare a report in connection with a class; one of her close friends had been injured. A group of three students, in the fall 2007 course CS4624 (multimedia, hypertext, and information access), undertook a term project to help with archiving from sites such as Flickr; see their report at <http://pubs.dlib.vt.edu:9090/191/>.

On July 11, 2007, I heard back from the Director of the University Honors Program that I could teach a course in spring 2008 (with help from coauthor Steven Sheetz); the title of Honors 3004 (see <http://collab.dlib.vt.edu/runwiki/wiki.pl?HonorsApril16>) is “A Digital Library Related to 4/16/07 at Virginia Tech.” Through that course, 14 from among the University’s brightest students are engaged in learning and service activities related to the tragedy of April 16. They will add reports and multimedia content to the digital library. They will include their interviews with a long list of visitors to the class—people from Virginia Tech and the local community who have been involved in the tragedy and recovery, and who will share their experiences, perspectives, and plans.

Thus, in connection with research, teaching, and service activities, a growing team is working to collaborate with people all around Virginia Tech and beyond, to help with the collection of all types of related digital content, as well as a broad range of digital library efforts to assist with organization, access, and preservation. It is my earnest hope that our Web site at <http://www.dl-vt-416.org> will help people to learn from April 16, and will be an aid to recovery.

---

#### Sidebar 1. Personal Reflections by Edward Fox

realignments (Friedman, 2006). One response to these changes is the move, in projects such as ours, to develop cyberinfrastructure (NSF, 2007). Another response is to help prepare professionals, as well as the broad base of students in countries such as the United States of America, for Living In the KnowlEdge Society (LIKES; Fox, 2007a, 2007b). In particular, for April 16, we are building a digital library that combines content and flexible services to support the needs of a targeted user community (Arms, 2000; Chen & Fox, 1996; Fox, 1993; Fox, Akscyn, Furuta, & Leggett, 1995; Fox &

Marchionini, 1998; Fox & Sornil, 1999, 2000; Fox, Suleman, Madalli, & Cassel, 2004; Fox & Urs, 2002; Lesk, 1997, 2005). Thus, we advance beyond the usual highly sequenced and preplanned approach to research and education (see Figure 1).

We believe that this is the right type of support for research and learning related to the events and aftermath of April 16. Traditionally, as can be seen in Figure 1a, research deals with hypotheses and related models or theories, which then lead to data collection, analysis, hypothesis testing, and subsequent publishing. The cycle can be extended, such as when data

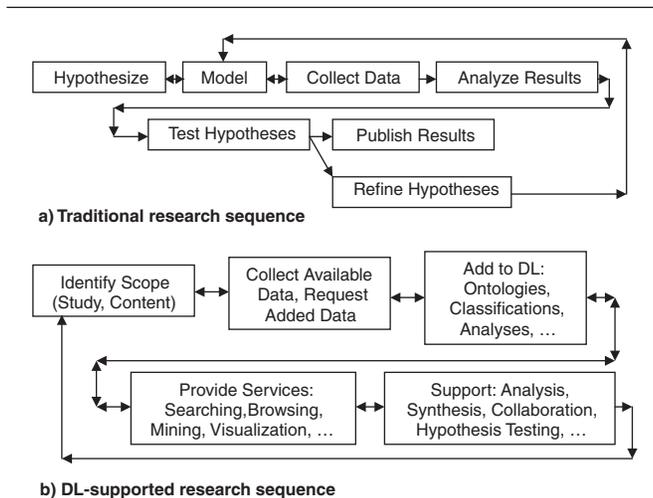
## The Event

The morning of April 16, 2007, began as any other morning on a college campus. However, by 8 a.m., the regular routine of a busy day began to change in an ominous direction. I heard sirens passing underneath my office window. Around 9 a.m., I learned of a shooting at a dormitory on the campus by an e-mail sent out by the university. Later, it was confirmed that two students were killed in that shooting and that the police were searching for the gunman or gunmen. Around 9:30 a.m., I again heard sirens and noticed a large number of police and rescue cars congregating below my office window and police officers roaming around with drawn weapons. Then I heard a loudspeaker announce that another shooting had occurred and that the shooter (or shooters) was still at large. We were urged to remain in our offices and to stay away from windows. At this time, the only information I was able to receive on what was happening came from university announcements, including e-mails (which were not regular), the loudspeaker, and the Internet. Cell phones were jammed, as was the university telephone system. Via another public announcement, we were instructed to remain in our offices and lock our doors. More and more police and rescue units were parking near my building, and I assumed that the shootings were occurring in my office building, although I had not heard any gunshots. I went to the Internet and learned that several students and professors had been shot and killed on the campus, but the location was unclear. We were still assuming that at least one shooter was at large and we were still being instructed to stay in our locked offices. Later that morning, I received a phone call from my daughter's school. They were telling her that shootings were occurring on campus; I confirmed that I was okay. By noon, we were told to leave our offices and to go to another building or to leave the campus. When I got outside, I saw hundreds of police cars and rescue vehicles around my building. I got to my car and had to go to the other side of campus to pick up my wife. The roadways were blocked, but an officer let me through. When I arrived at my wife's building, I could not get in. Eventually, their doors were opened, and I got to watch local television programs concerning what had happened. That was when I first learned that at least 20 students and professors had been shot to death. In addition, one gunman had been identified as the shooter and one of the supposed gunmen earlier identified was instead a victim. I stayed in the cafeteria where my wife works and watched as more and more details of the shooting emerged. Cell phones were still jammed, but now and then people could get through to loved ones.

## What Have We Learned?

Many things could be learned from such a tragedy. For one, some of the news reports of the events were factually incorrect, and this point became more evident as early news reports continued to appear. In the aftermath of this tragedy, it became clear that this was worldwide news. Although classes had been cancelled for that week, many people kept returning to the campus to resume some semblance of a regular work schedule. The university hosted an assembly a couple of days after the shootings and the results demonstrated a bonding among those connected with Virginia Tech that I had not felt before in more than 32 years at the school. People throughout the area were displaying Virginia Tech emblems wherever they went. The sense of community and identification with Virginia Tech assumed a larger, global context in the days and weeks following the tragedy. The rest of that week, my coworkers and I received e-mails and phone calls from colleagues in the United States and abroad, sometimes from anonymous colleagues wanting to share a moment of care and concern. Images of concern and care came to us through national and international news accounts, as well as posters and banners from colleges and public schools throughout the country, with names of unknown well wishers written on them. Even unknown attendees at professional conferences would approach us with concern and regret once they knew we were from Virginia Tech. From the tones and expressions relayed in these messages and images, it became evident that we were being considered as victims, even if we were not personally injured in the attacks. Also, it became clear that others throughout the world, those related to the field of academia or not, somehow identified with our shock and pain as we tried to cope with what had happened. Expressions of sympathy and regret seemed genuine. Ironically, although our sense of safety and community at Virginia Tech was initially shaken by the shootings, a sense of togetherness emerged in its place, not only among those within the Virginia Tech nation but also in a larger, global context, even if this communal development may have been short-lived.

With the passage of time, the expressions of sympathy and concern are less numerous and attention to the event has lessened. However, we are reminded by counselors and psychologists that we can expect delayed effects for students, faculty, and administrators, secretaries, and all connected with the university throughout the first year following the tragedy, especially around the first anniversary of the events. In this respect, we are treading in unfamiliar territory and there is a danger that the impacts we are told to expect may, in part, be effected by our own, perhaps subconscious, expectations. Thus, the rest of the story has yet to be witnessed, and only time will tell the full impacts of the tragic events of April 16, 2007, for all of us present at the time and for those who have been touched by the tragedy, directly or indirectly.



**Figure 1.** Standard research sequence of activities (a) and digital-library-supported version (b).

suggest revisions to the original hypotheses or improvements to models or theories. However, when an interdisciplinary community seeks to study different aspects of an important phenomenon, such as the tragedy on April 16 and its aftermath, the approach shown in Figure 1b could be more efficient and effective. Based on some definition of scope, with regard to data and other types of content, as well as to studies of interest, a broad collection effort can proceed. Much as a university or town supports a local library, or a nation addresses archiving of its culture and heritage, in the digital era, a community can be supported by a comprehensive digital library. Furthermore, whereas brick-and-mortar libraries typically provide only a small set of services (e.g., searching, browsing, reference-desk support), digital libraries can provide an open-ended set of services, much like diverse companies operating on the Web helping millions who are interested in Web pages. As can be seen in Figure 1b, digital library services can provide add-on content, an extensible set of services, and direct aids for analysis, synthesis, collaboration, hypothesis testing, and so on.

One added benefit of the approach illustrated in Figure 1b is that it is feasible to add to the digital library all results of studies, including curricular materials, educational resources, temporary data sets, partial analyses, visualizations of findings, and varied publications (reports, theses, conference papers, journal articles, etc.). Thus, building rapidly on the work of others is facilitated, and collaboration can lead to more comprehensive insights. Also, once a large collection is available, which can support teams of interdisciplinary scholars, new studies can be launched

without having to wait for additional (and expensive) data collection activities. Requests for institutional review board review can be batched to cover a range of studies, further reducing related delays in follow-on research schedules.

These improvements, however, must be carefully arranged, or else they may be constrained because of legal and social concerns, which often lead to limited access, even for those with helpful intentions, to sensitive or personal information. Privacy must be respected, and anonymity must be preserved when desired. In our litigious society, complaints and suits may arise in the aftermath of tragedy, any related data may be subpoenaed, and those providing data may be called as expert witnesses. Thus, even if a research design has separated identifying information from survey responses, a subpoena may lead to identification of individuals. Groups providing data may be viewed as classes, as for class action suits. Individuals in special communities, for example, those injured during the tragedy or administrators directly involved in the event, may be off limits with regard to data collection for long periods of time, typically 3 years or more.

Nevertheless, to the extent possible, this project aims to provide a range of advanced technologies to serve the community of those interested in April 16. In addition, we seek to study technology as a factor in the activities and processes related to April 16, as is discussed in the next section.

## Approach

In this section, we consider how information visualization, data and text mining, and social network analysis, supported by and integrated through digital libraries, can be applied to help regarding April 16, 2007. But, as in any research, we must first understand (at least partially) the needs and requirements.

Broadly, we aim to support the global community with regard to their diverse interests regarding April 16. We will provide access to millions of pieces of news. We will help connect with the hundreds of thousands of individuals who have used blogs or other mechanisms to express their concern. We will support access to tens of thousands of photos and thousands of videos. We aim to support students writing reports, at all levels. Ultimately, there will be scholarly theses and dissertations related to the event. In spring 2008, several of the authors are running an honors class for some of Virginia Tech's

brightest students, leading to class reports that will be accessible online by May 2008.

We also will support the community as it remembers, analyzes, mourns, grieves, learns, and provides assistance for those involved. People will view the tragedy from a variety of personal perspectives, as well as from viewpoints related to disciplines such as psychology and sociology. To avoid bias, and to accommodate those needs, we aim to collect as much information as is possible and to provide browsing and other services that reflect an open-ended set of categorizations, as well as alternative taxonomic and ontological organizations. Because there are so many angles to consider, in the following, to illustrate our plans, we focus on what may be an unusual aspect of the Virginia Tech scene—namely, the widespread deployment of advanced computing, communication, and information technologies.

### **Research Agenda for the Social and Behavioral Sciences Related to Technology**

We plan to develop, through discussions and focus groups, a research agenda regarding the application of such technologies during and after a crisis. Indeed, we must consider more broadly the relationship between technology and crisis, not only as an aid but also as a cause, correlate, and part of the context of crisis. One aim of the research agenda will be to address how using technology during and after a crisis can serve to reduce anxiety and manage distress. Accordingly, we are involved in survey activities to collect data about the use of technology (during and immediately after the crisis), feelings of distress, and coping strategies.

Technology has become embedded into the lives of people across the global society. It is now commonplace for communication to occur through multiple technologies, for example, cell phones, personal digital assistants, and computers, resulting in multimedia traces of the communications being saved as voice, text, video, and instant messages. Similarly, the Internet and the emergence of virtual communities, for example, Facebook (<http://www.facebook.com>) and MySpace (<http://www.myspace.com>), have enabled communication on a scale previously unimaginable. However, we do not know how these technologies were used during and after crises to reduce or mitigate distress. Indeed, we do not know whether the use of technology reduces distress; it is possible (perhaps likely) that problems encountered with normally

dependable technologies that fail during crises exacerbate rather than mitigate stress for those involved. Thus, a series of questions results from the current nexus of technology and tragedy: How do the established theories of dealing with distress apply in the context of the current state of technology? What capabilities should technology have to provide support during crises? How can technology be enhanced to aid in reducing stress immediately after crises? How are social networks used through technology to mitigate stress during and after crises? Is the use of the social network through technology different for individuals experiencing high stress versus those with lower stress? Does the emergence of unique virtual communities necessitate evolution of the theories related to distress? Do the communication patterns of distressed individuals (such as the gunman on April 16) provide insights that suggest ways to identify those in need of counseling or other interventions?

An important sociological issue to be addressed is the social network used to relay and receive information concerning a traumatic event, such as what occurred on the Virginia Tech campus on April 16, 2007. Specifically, where do students, faculty, staff, and others go for information about what happened? Whom do they contact first for such information and to whom do they first relay the information they have received? Whose opinion do they trust? Do people turn to relatives, friends, and authority figures known to them, such as professors, or to the media, or some other source of information? In turn, to whom do people relay the information they have received? In the aftermath of a traumatic occurrence, do these contacts change? Also, are these social networks contextualized by social characteristics, such as age, gender, or race and ethnicity? Of particular interest would be how technology is used to receive and convey information and reactions to traumatic events and the sociological contextualization of the use of technology. For example, one hypothesis could be that college students use technology more often to receive and convey information than do professors or adults in general. The purpose of our discussions and focus groups is to generate such hypotheses and effective methodologies for testing them. Table 1 illustrates the early phase of this effort by summarizing some of the research interests of psychology researchers who participated in the first focus group.

In particular, we are seeking to encourage interaction between social and behavioral scientists and experts in technological fields to produce new and

**Table 1.** High-Level Summary of Psychology-Related Research Interests From First Focus Group

Umbrella Category	General Category
Adjustment to tragedies	Change in support groups and systems
	Coping
	Distress
	Forgiveness
	Health effects
	Resilience
	Role maintenance
	Vulnerability to problems and stress
Social support	Change in support groups and systems
	Helping behaviors
	Use of technology
	Use of postdisaster services
	World view maintenance

innovative ways of developing and testing hypotheses concerning how people react and respond to traumatic violent episodes, both in the short term and over the long term.

It has been estimated that 10% to 40% of individuals who experience trauma will subsequently develop posttraumatic stress disorder (PTSD; Michaels et al., 1999). In light of these figures, PTSD is considered one of the most common anxiety disorders experienced among young adults (Breslau, Davis, Andreski, & Peterson, 1991). *The Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000) defines a traumatic event as “an extreme traumatic stressor involving direct personal experience of an event that involves actual or threatened death or serious injury, or other threat to one’s physical integrity” that elicits “intense fear, helplessness, or horror” (p. 463). Specific traumatic experiences can include natural or technological disasters, violent crimes, abusive relationships, and sexual assaults. Undoubtedly, the Virginia Tech shootings fall centrally within the realm of trauma. Nonetheless, not all who experience a trauma will develop PTSD; many individuals show resilience in the face of tragedy. Conceptual models of PTSD explain its emergence in terms of interactions among the trauma experience (e.g., proximity to the event, degree of life threat, loss/disruption), survivors’ characteristics (e.g., ethnicity, preexisting psychopathology, prior trauma experiences), cognitive processing (e.g., positive or negative coping), and the recovery environment (e.g.,

perceived social support; Green, Wilson, & Lindy, 1985; LaGreca, Silverman, Vernberg, & Prinstein, 1996). The digital library, and specifically the surveys of students, faculty, staff, and community, will allow us to test aspects of this model in the immediate aftermath of this tragedy as well as in a longitudinal manner over the following years. Three hypotheses (of many) are of specific interest in this regard. First, does the psychosocial model predict which individuals will or will not develop PTSD as well as other stress-related reactions (such as depression, substance abuse, and physical health problems)? Second, which aspects of coping (e.g., spiritual, cognitive, active/passive, emotion focused) and social support (e.g., friends, family, classmates, teachers) are most influential in buffering survivors from these negative outcomes? Third, to better understand protective factors, how do coping strategies change and how are social support networks formed and used?

Another potential area of exploration is the impact on physical health of acute and chronic psychological stress from traumatic events such as on April 16. Both short- and long-term negative health effects have been previously documented in response to the stress of traumatic community events (Lovallo, 2005). For example, increased incidence of sudden cardiovascular events such as myocardial infarction and stroke have been noted in acute response to earthquakes (Leor, Poole, & Kloner, 1996), missile attacks (Kark, Goldman, & Epstein, 1995), and even nonthreatening events such as an upset loss by a major sports team (Carroll, Ebrahim, Tilling, Macleod, & Smith, 2002). It is also possible that the lingering effects of trauma, such as the April 16 shootings, may negatively affect health by way of chronic increases in negative emotional states, such as anxiety, anger, and depression, which are known to have negative relationships to physical health (Everson-Rose & Lewis, 2005). Two common psychological factors across stressors are well-known to have deleterious physical effects: control and predictability. Stress and trauma that is unpredictable, and over which individuals exert little or no control, will engage autonomic nervous and endocrine system responses to threat that can have negative effects on the immune and cardiovascular systems (Porges, 1995; Seligman, Maier, & Solomon, 1971; Thayer & Lane, 2007; Weiss, 1991). The April 16 shootings were both unpredictable and uncontrollable and as such may confer similar negative effects on the health of the community in Southwest Virginia. Possible directions of exploration include how

specific coping styles predict health outcomes, as well as the roles of factors such as social support and religiosity in buffering the negative effects of stress. Findings might not only help document the broad effects that this tragic event has had on our community but also might be generally applied toward optimal preparation for and response to unpredictable stressors, traumas, and crises.

### Special Analysis

Our project aims to address the needs of the public, as well as the research and education goals of various interested communities, by providing important add-on services that will be closely coupled to the digital library of related content. In particular, we focus on three analysis aids: information visualization, data mining, and social network analysis.

Information visualization (North, 2005; Spence, 2007) helps us understand the data and information under consideration and facilitates discussion and presentation of results. Advanced techniques have been developed to help us deal with urgent problems faced by the intelligence community, such as terrorism (Thomas & Cook, 2005), and so can be applied to the events on April 16. One key benefit is gaining additional insight (Saraiya, North, Lam, & Duca, 2006), a very important goal for those trying to understand tragedies.

Visualization can be applied to texts (Wong, Whitney, & Thomas, 1999), geospatial information (Jung, 1999), retrieval results (Hearst, 1995; Nowell, 1997), biological pathways (Saraiya, North, & Duca, 2005), and cross-jurisdictional information (Kaza et al., 2005). They can lead to three-dimensional presentations (Amavizca, Sánchez, & Abascal, 1999), handling of graphs (Tollis, 1996), or work with Internet collections (Viegas, Wattenberg, Ham, Kriss, & McKeon, 2007).

An important challenge in our work is the integration of information visualization with digital libraries (Nowell & Fox, 1995), such as through special protocols and transformations (Shen, Wang, & Fox, 2002; Wang, 2002). Even more broadly, information visualization in digital libraries can be integrated with related searching and browsing services, under the broader rubric of exploration (Shen, Vemuri, Fan, da Silva Torres, & Fox, 2006).

Of particular value, as we engage in discussions with social and behavioral scientists, is to demonstrate the benefits of information visualization. Fortunately, Virginia Tech has a laboratory with a flexible gigapixel-scale display arrangement (Ball & North,

2007), so different approaches can be applied simultaneously to related data and then compared and discussed in a collaborative setting. Later, we can select methods that are shown to be of interest and migrate them to smaller display configurations, making the services available remotely over the Web.

Data and text mining are key analysis techniques, which in turn can feed into information visualization, as above. Data mining can work with large data collections or databases (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). It can lead to discovery of patterns, and shifting from data analysis to development of scientific models (Ramakrishnan & Grama, 1999). Data mining can be helpful when working with crimes (Chen et al., 2004). Text mining (Fan, Wallace, Rich, & Zhang, 2006), a type of data mining, focuses on the large amounts of text available on the Internet, a key resource for what is being collected in connection with April 16.

Social network analysis, that is, analyzing contact patterns among people, has proven to be a valuable research tool in various domains. This approach has been applied most notably to study the spread of epidemics (Eubank et al., 2004) and to devise viral marketing strategies (Domingos & Richardson, 2001). Such analyses have revealed that the study of the underlying properties of a network can yield valuable clues regarding the dynamics of the processes in the network.

In the context of this project, we propose to study the technology network, during and shortly after the campus shootings. We will work to construct networks of interactions between people who communicated, especially using the means shown in Table 2. Some of the most important questions to address include the following.

1. What were the most common modes of communication among different groups of people (faculty, students, staff, etc.)? Which were most effective for each group?
2. Who are the most influential, central, or well-connected individuals/groups (dean's office, department heads, focus groups, Greek societies, etc.), if any? Under circumstances like those on April 16, who would be able to communicate information most effectively?
3. How long does it take for information to spread among different groups of people? Are there any bottlenecks?
4. What were the patterns of technology usage in the aftermath? Which ones were used most, and in what ways (blog entries expressing condolences, memorial videos, etc.)?

**Table 2.** Technology Support: Issues and Mechanisms

Issue	Details
Device	Land line telephone Cell phone Pager Personal digital assistant Laptop Desktop computer In-person meeting
Mechanism	Talk Leave voicemail E-mail Listserv Online instant messaging/chat Send text message Website Web 2.0 service (Facebook, YouTube, etc.) Audio conference Videoconference
Goal	Obtaining information Providing information Understanding another person Being understood
Communicating with	One person A group A community The public
Quality concern	Ease of use Utility/benefit Compatibility with personal activities/ tasks/lifestyle

## Digital Libraries

A context and environment to support information visualization, data and text mining, and social network analysis must integrate these and other services with related content. Thus, our project builds atop work in the field of digital libraries.

The traditional library (see Figure 2a) assumes that authors publish works, which are collected by librarians, and then can be discovered and accessed by readers. This works both with bricks-and-mortar libraries and with digital libraries, but many new interaction patterns are only feasible using digital technologies. For example, as shown in Figure 2b, authors or creators can put their works onto Web sites. With the help of search engines, such as from Google, Microsoft, or Yahoo, those works are crawled and placed into large collections, to be indexed and made searchable for Web surfers.

More recently, another paradigm has arisen, whereby millions of creative people can publish on

the Web, with simpler schemes such as blogs and wikis. Furthermore, these Web users can work in contexts of specialized types of content (e.g., photos in Flickr or videos in YouTube) or community collaboration (e.g., Yahoo groups, Facebook, MySpace). Figure 2c summarizes this Web 2.0 (O'Reilly, 2005) approach, discussed in the next subsection. Here, we focus on how to organize such diverse and distributed content, a key goal of libraries.

Table 3 illustrates a conventional library organization, giving the key layers or perspectives. Users or patrons work through interfaces, either in the physical world, where one walks around bookshelves, or in the virtual world, where one has computer screens (or personal digital assistants or cell phones). These lead to services, such as searching and browsing, which make use of catalogs (of descriptive information, i.e., metadata; Weibel, 1995) or indexes, and eventually lead to desired content.

In our work with digital libraries, we have abstracted and generalized the handling of information, developing the 5S framework, a theory-based formal model and environment for describing, understanding, developing, and innovating (Fox, 1999; Goncalves, 2004; Goncalves & Fox, 2002; Goncalves, Fox, Watson, & Kipp, 2004; Gorton, 2007; Zhu, 2002). Table 4 explains each of the five Ss, which can be formally defined, but at the same time gives an intuitive way to characterize the underlying aspects behind any advanced information system. Thus, digital libraries manage content (streams), supporting a variety of organizational aids (structures), across time and space, presented typically in a two-dimensional form (spaces) for various types of uses (scenarios, implemented as services) and for diverse societies (from the personal to the global).

If we focus on the content portion of a digital library, we refer broadly to digital objects that can be of any media type. Sometimes a group of related digital objects, such as the parts of a dissertation, can be viewed as a composite (Fox & France, 1987) or compound document (Lagoze & Van de Sompel, 2007) or package—a larger type of digital object.

Combining metadata and digital objects, along with broad ranges of services, and addressing the distributed nature of modern information systems, we can summarize the layers of a modern digital library as in Table 5. This gives an architectural view of the type of digital library necessary for aiding work related to the April 16 tragedy, where various types of distinct repositories must work together in a federated fashion

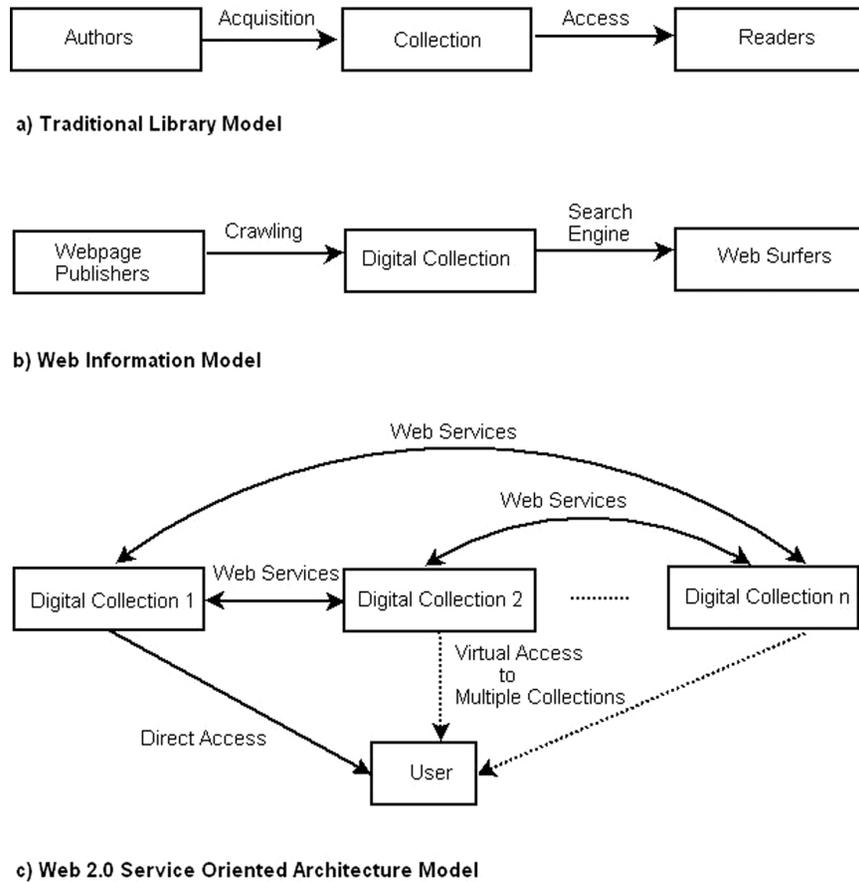


Figure 2. Information models for traditional libraries (a) and using the Web (b) or Web 2.0 (c).

Table 3. Layers of a (Digital) Library

Interfaces
Services
Catalogs
Contents

Table 4. Five Ss

S	Roles (Examples)
Societies	Users, groups, organizations, communities, nations, cultures, world; agents, actors, service managers
Scenarios	Services, workflows, procedures, tasks, activities, programs, routines
Spaces	One-, two-, three-, and multidimensional, vector, measure, probabilistic, topological
Structures	Data structure, database, ontology, taxonomy, graph, org-chart, schema, pattern, grammar
Streams	Video, audio, voice, character, bit, sensor

Table 5. Layers of a Comprehensive Digital Library

Layer	Description/Constituents
1	Interface: browser + Java applets + plugins
2	End-user services: search, browse, download, upload, collaborate, analyze, visualize
3	Content aggregations: repository, catalog, collection, ontology, taxonomy
4	Content types: metadata, digital object, package/composite Content forms: video, audio (speech, music), image (photo, graphic), structured (XML, HTML, SGML, spreadsheet, database), text
5	Logical/accessible content: local content, mapped versions of remote content, surrogates for remote content, pointers to remote content
6	Federation and COLLECTION SERVICES: crawling, harvesting, federated search, retrieving, mapping, fusing, mediating, application programming interface-based accessing
7	Remote sites: Web sites, Web 2.0 sites, federation partners

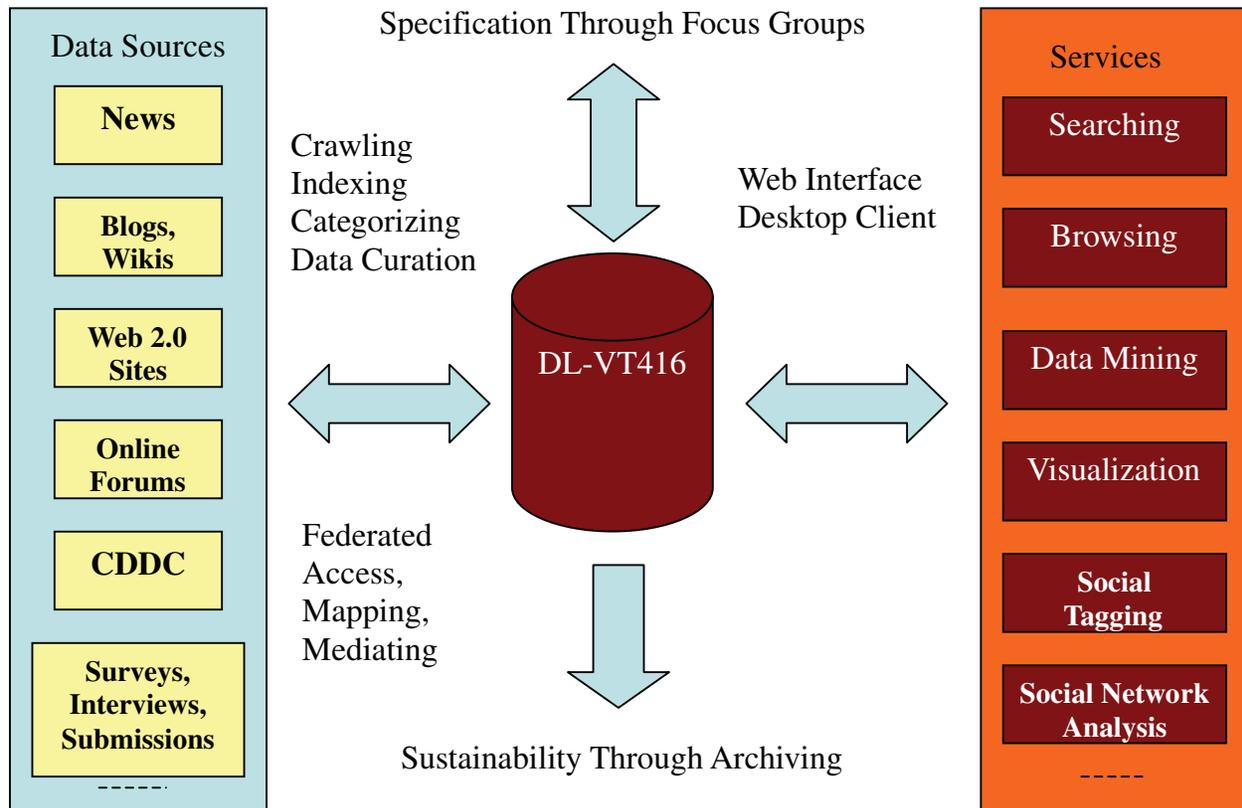


Figure 3. Architectural view of our digital library.

(Candela, Manghi, & Pagano, 2007). In particular, we are building a system as shown in Figure 3 to provide a range of services (right side) over a range of data sources (left side), to suit the requirements of a broad user community (with needs specified through the aid of focus groups—see top center), and to be ultimately sustained by Virginia Tech University Libraries' archival efforts regarding April 16 (see bottom center). The arrows represent the functions and processing to support making all this work together. Of particular interest is the handling of Web 2.0 information, as is discussed in the next subsection.

## Web 2.0

Web 2.0 (O'Reilly, 2005) is a relatively new model of the Web, which collectively refers to the collaborative and value-added services, in addition to the usual content hosting services, that Web sites pervading the Web provide. The collaborative aspect of Web 2.0 refers to the interaction among users visiting the Web sites. Prominent examples are Web sites such as Facebook and MySpace, which provide a platform for users to not just browse the content but also

actively participate in discussions, form focus groups, upload and share content, and so on. Value addition is achieved, for example, by providing a platform for building applications and services on top of the already existing services that the Web sites provide. Google Maps (<http://maps.google.com/>), for example, provides application programming interfaces (APIs), which can be programmatically used to build services on top of the ones already offered by the Web site. This provides considerable value addition to an audience interested in building applications based on Google Maps.

Such value addition, especially by means of providing APIs, has added a new dimension to digital library research. A new paradigm involving a service-oriented architecture has been proposed (Petinot et al., 2004), and successfully used, wherein different digital libraries (possibly heterogeneous, distributed, and geographically separated) can present a single uniform interface to the users and can call on the services of each other (see Figure 2c). The interaction between the digital libraries is achieved by calling the Web services hosted at each others' sites.

**Table 6.** Comparison of Collecting With Using Web 2.0

Scenario	User Effort/Benefit	Volume	Researcher Effort/Effort With API
User upload	High/low	Low	Low/NA
User post to Web 2.0	Low/high	High	High/low

NOTE: API = application programming interface; NA = not applicable.

We direct our efforts toward analyzing resources related to April 16 from prominent Web 2.0 sites, such as YouTube, Facebook, MySpace, Flickr. These probably represent the most popular social networking and content sharing points on the Web today. Table 6 summarizes our user-centered rationale to make use of these sites. The resources hosted at these sites can be accessed either directly from the respective content providers (depending on the copyrights and license restrictions associated with the resource), in which case we will provide the URL to the resource, or locally at our end, in which case these resources will be cached at our site using the APIs provided by the content providers, or using copies provided by content creators. We will use the DSpace platform (MIT, 2003) as our digital library system, running on a powerful (due to generous discounts) Dell server. The resources (or surrogates thereof), as well as the metadata associated with them, will be stored in the digital library and be made available for analysis.

### Supporting Systems-Level Science

A key goal in our integrating special services, digital libraries, and the Web 2.0 infrastructure is to support systems-level science on the social dynamics associated with crisis events. Inspired by current research trends in biology and the life sciences, systems-level science seeks to understand the functioning of very large and complex systems and all the interactions therein in a holistic fashion. This is in contrast to more traditional reductionist science, which narrows the problem down to focus on specific individual variables. Systems-level science is more exploratory in nature, looking for broader trends and patterns of higher complexity; it encourages the development of new hypotheses. In the life sciences, the system of systems-level science refers to the functioning of complex biological organisms. In Virginia Tech's case, the system under study is the complex communities of people and how they respond to crises.

Systems-level science is more challenging but can offer deeper insights into underlying phenomenon. Systems-level science requires the following.

1. Rapid and continuous collection of massive data: The recent growth of systems-level science in biology was supported by the invention of microarray and similar instrumentation that enables simultaneous collection of data about thousands of genes and proteins at the cellular level, thereby offering a complete picture that is both detailed and broad. Similarly, our digital library will be organized to ingest large collections, such as student-submitted collections of e-mail-related data, which are collected and curated over time.
2. Integration of diverse, heterogeneous data: Our digital library must be able to bridge diverse data sources, including vast Internet and Web collections, as well as rich personal sources, such as Facebook pages, surveys, and interviews. We also will investigate new kinds of data that could be captured, such as by volunteers using our two SenseCam devices (Hodges, Williams, & Wood, 2007).
3. Real-time exploratory analysis: To gain deep insight, our digital library must link with analysis tools to support rich exploratory analysis of complex patterns and interrelationships for theory development. Real-time analysis must maintain synchronization with incoming information to promote awareness of breaking hypotheses and quick response. Access, analysis, dissemination, and utilization will be supported by visualizations and data mining tools.

### Community, Technology, and Content

Our community of users includes those interested in the April 16 tragedy, for example, those who through Web 2.0 or other services have contributed to the body of content related to the event and its aftermath. Many of those contributing content made use of technology support mechanisms summarized in Table 2.

### Model

From Table 2, we see that a variety of types of devices are employed for communication and for information access. These allow a number of mechanisms for communication, with diverse goals, serving individuals and/or broader groups. Often, selections among

**Table 7.** Propositions/Assumptions Regarding Factors Related to Stress

Proposition	Description
P1	Event exposure is positively associated with event-related stress
P2	Pre-event stress is positively associated with event-related stress
P3	Coping is inversely associated with event-related stress
P4a	Social support is inversely associated with pre-event stress
P4b	Social support is inversely associated with event-related stress
P5a	Community support is inversely associated with pre-event stress
P5b	Community support is inversely associated with event-related stress
P6	Using technology associated with event-related stress

these are based on quality metrics, relating to utility, ease of use, and contextual factors. Because communication technologies are embedded in modern life, we seek to explore how using communications technologies can be used after tragic events to help in reducing stress.

In a crisis, technology support may have an impact on stress and adaptation to trauma. Table 7 gives a set of propositions we seek to explore in this regard. Figure 4 provides a model of the causality we assume, which needs to be tested, including use of technology. (Recall the earlier section on *Research Agenda for the Social and Behavioral Sciences Related to Technology*.)

### Content Collection

Back in 1996, Virginia Tech led the world by encouraging (graduate) students to upload content, in particular electronic theses and dissertations (Fox et al., 1996); as of January 1, 1997, this was a requirement for all master's theses and doctoral dissertations. In 2003, the Networked Digital Library of Theses and Dissertations formalized the role it had been playing since 1996, incorporating as an international non-profit (Fox, 2003) to further spread the idea around the globe; by November 2007, there were more than 387,000 electronic theses and dissertations available in many languages. Oddly, in spite of the familiarity of uploading to many young people today, many graduate students, faculty, and universities, in parts of the United States as well as in nations with smaller

economies, still have not understood how easy and beneficial such uploading is, and thus lose out with regard to the wider recognition of their research contributions that comes through services that build atop the NDLTD Union Catalog (OCLC, 2004).

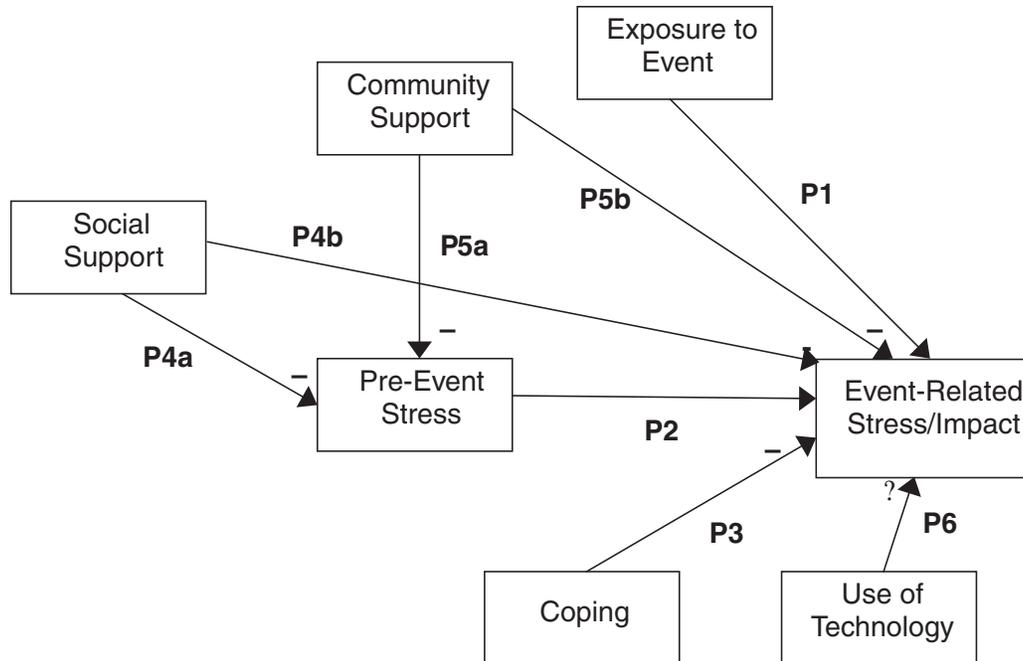
Uploading content related to April 16 is thus a natural approach for Virginia Tech, as shown in Table 8. The campus developed a Web site, with various pages, around the theme of remembering. University Libraries has committed to a permanent archive and has been addressing the thousands of physical artifacts contributed. The Center for Digital Discourse and Culture has launched the April 16 Archive to collect content by crawling the Web (in collaboration with the Internet Archive; a copy of those pages makes up one part of our digital library) and by supporting the uploading of information (CDDC, 2007). Our effort is supplementing those activities, handling other types of content as well, such as videos, audio recordings, and e-mail information.

### Content in Web 2.0

Though to many the actions described in the previous section are novel and advanced, requiring a reach in technology application, to others the people involved are but digital immigrants, not digital natives who are comfortable and who grew up with the Web (Prensky, 2001). Those who are comfortable with current Web services can of course upload content but often prefer communicating directly by working inside applications that support virtual communities, such as Facebook. In Table 8, the counts are low for all rows except for the guest book, which can be viewed as a Web 2.0 application. Thus, as in Table 6, because the effort is less, posting to a Web 2.0 service is much easier, hence to be preferred, relative to uploading (which is called for in applications and sites described in the previous subsection).

Such users vote with their mouse, as can be seen in Table 9. The Web 2.0 applications listed are popular among those interested in the April 16 tragedy, leading to large numbers of entries. By far the largest number is the count of users who have joined related groups, as is detailed in Table 10. This table lists only large groups, that is, with more than 500 in the group, and only with first letter in range A–I. Nevertheless, the table entries account for roughly the estimated count of 400,000 shown in the first row of Table 9.

Clearly, to understand the events of April 16 in the larger perspective, it is important to capture



**Figure 4.** Model of stress from a tragedy and the factors related.

**Table 8.** Local Sites and Counts of Entries

Source	URL	Counts
Main site	<a href="http://www.vt.edu/remember/">http://www.vt.edu/remember/</a>	10 pointers
Memorial site	<a href="http://www.vt.edu/remember/memorial/">http://www.vt.edu/remember/memorial/</a> (pages of condolences)	720
Higher ed support	<a href="http://www.vt.edu/remember/higher_ed_support.php">http://www.vt.edu/remember/higher_ed_support.php</a> (links to pages stating support)	395
Biographies	<a href="http://www.vt.edu/remember/biographies/">http://www.vt.edu/remember/biographies/</a>	32
University relations	<a href="http://www.vt.edu/remember/archive/">http://www.vt.edu/remember/archive/</a> (releases to VT community, only for April 16 to April 22)	30
Photo gallery	<a href="http://www.vt.edu/remember/photos/">http://www.vt.edu/remember/photos/</a>	24
Office of Recovery and Support	<a href="http://www.recovery.vt.edu/">http://www.recovery.vt.edu/</a>	9 pointers
Guest book	<a href="http://www.legacy.com/GB/GuestbookView.aspx?PersonId=87298850">http://www.legacy.com/GB/GuestbookView.aspx?PersonId=87298850</a>	4,282 pages
The April 16 Archive	<a href="http://www.april16archive.org/">http://www.april16archive.org/</a>	720 photos, ~1,600 items
VT Magazine	<a href="http://www.vtmagazine.vt.edu/memorial07/today.html">http://www.vtmagazine.vt.edu/memorial07/today.html</a> (plus 35 condolences)	21 photos
Planet Blacksburg	<a href="http://blog.planetblacksburg.com/">http://blog.planetblacksburg.com/</a>	12 blog entries, book

**Table 9.** Public Sites with Entries for "Virginia Tech April 16"

Site	URL	Estimated Entries
Facebook	<a href="http://www.facebook.com">http://www.facebook.com</a>	400,000 users
Flickr	<a href="http://www.flickr.com">http://www.flickr.com</a>	400 photos
MySpace	<a href="http://www.myspace.com/">http://www.myspace.com/</a>	700 results
YouTube	<a href="http://www.youtube.com/">http://www.youtube.com/</a>	4,000 videos
Technorati	<a href="http://technorati.com">http://technorati.com</a>	9,000 blog posts
Digg	<a href="http://Digg.com">http://Digg.com</a>	500 stories (link elsewhere), other

information about those who expressed their concern and interest electronically. Fortunately, many of the Web 2.0 services have APIs to help automate interaction, as is shown in Table 11. Our group will be working with these APIs as we develop our digital library. This will facilitate identification and in some cases collection of information, when access permissions allow.

**Table 10.** Some Facebook Groups Larger Than 500, for “Virginia Tech April 16”

Size	Title
1,531	A Canadian Tribute To All Those Involved In The Virginia Tech Shooting
1,070	A DEDICATION TO THOSE HURT BY THE INCIDENTS AT V-TECH
752	A Heartfelt Thank You to the Heroes of 4.16.07
8,129	All will be missed—VA Tech: April 16, 2007
3,384	Always remember the Virginia Tech Massacre
873	Always Remember 4/16/07—Virginia Tech School Shootings
3,896	America Stands in Support of Virginia Tech
1,409	AMERICA SUPPORTS VIRGINIA TECH
1,201	April 16, 2007—A Moment of Silence “TODAY WE ARE ALL HOKIES”
40,322	April 16, 2007—A Moment of Silence
898	April 16th 2007 Will Be Remembered Forever
324,653	A tribute to those who passed at the Virginia Tech Shooting
561	College Students United for Virginia Tech
859	Condolences For Virginia Tech . . .
688	Deepest condolences from England to the dead of Virginia Tech
730	Georgians (and others) Praying for VA TECH
506	04*16*07 God Bless the Hokies
633	God bless the Virginia Tech 33, We will always remember.
2,195	God Bless the Virginia Tech students
573	God Bless VA Tech—Current Events
3,673	God Bless VA Tech—Beliefs & Causes
1,176	God Bless Virginia Tech—Beliefs & Causes
1,921	God Bless Virginia Tech—History
795	God Bless Virginia Tech—Current Events
848	High Schooler’s Taking A Stand. We’re Here For You Virginia Tech!
621	Hoosiers Pray for Virginia Tech Tragedy
1,379	I’ll be praying for Virginia Tech.
765	In Honor of the VA Tech Victims
1,882	In loving memory of all those who lost their lives at Virginia Tech ~1,560
1,560	In Memory of the Victims of the Virginia Tech Shootings

**Table 11.** Web 2.0 Sites and the URLs for Their APIs

Site + Description	URL for API
Flickr photo sharing service	<a href="http://www.flickr.com/services/">http://www.flickr.com/services/</a>
Google search	<a href="http://www.google.com/apis/">http://www.google.com/apis/</a>
YouTube videos	<a href="http://www.youtube.com/dev">http://www.youtube.com/dev</a>
del.icio.us bookmarks	<a href="http://del.icio.us/doc/api">http://del.icio.us/doc/api</a>
Facebook social networking	<a href="http://developers.facebook.com/">http://developers.facebook.com/</a>
MySpace social networking	<a href="http://www.readwriteweb.com/archives/myspace_evolves_developer_platform.php">http://www.readwriteweb.com/archives/myspace_evolves_developer_platform.php</a>
AOL video search	<a href="http://developer.searchvideo.com/index.php">http://developer.searchvideo.com/index.php</a>
Yahoo video search	<a href="http://developer.yahoo.com/search/video/V1/videoSearch.html">http://developer.yahoo.com/search/video/V1/videoSearch.html</a>
Technorati blogs search	<a href="http://technorati.com/developers/">http://technorati.com/developers/</a>
Microsoft Virtual Earth	<a href="http://dev.live.com/virtualearth/">http://dev.live.com/virtualearth/</a>
Creative Commons Licensing	<a href="http://api.creativecommons.org/docs/">http://api.creativecommons.org/docs/</a>
Digg social bookmarks	<a href="http://www.digg.com/tools/">http://www.digg.com/tools/</a>
NewsGator news, Web sites, and blogs	<a href="http://www.newsgator.com/ngs/api/">http://www.newsgator.com/ngs/api/</a>

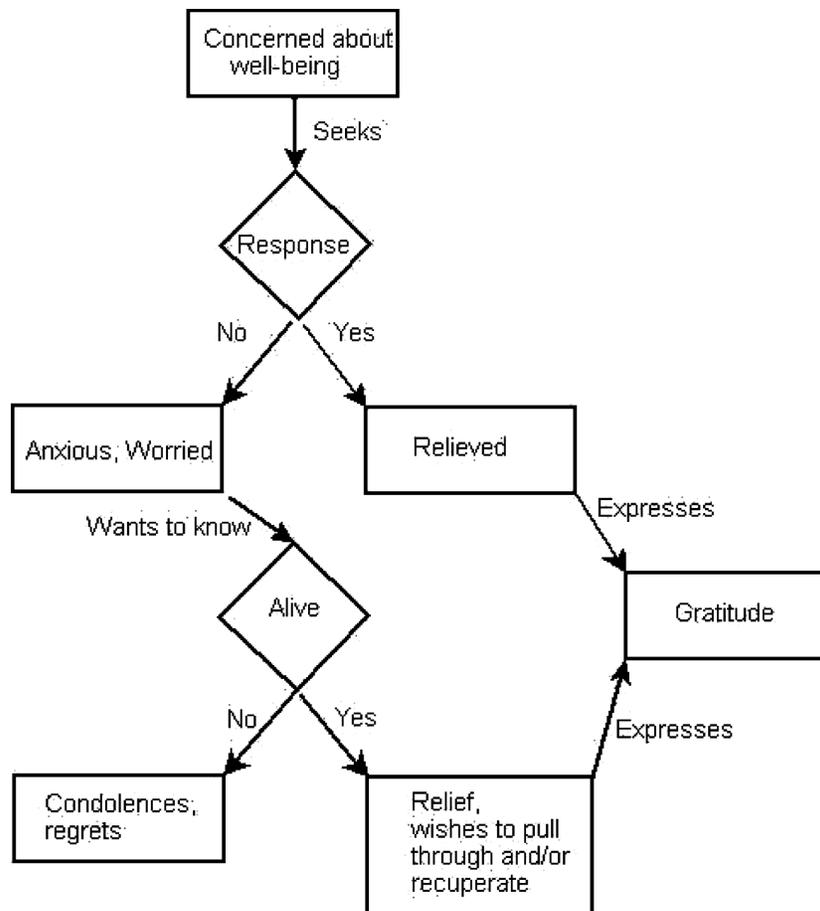
## Preliminary Analysis

We are just beginning to analyze some of the data available to us, and we are just starting to explore hypotheses and models that have been proposed with regard to the April 16 tragedy. Below we discuss

some preliminary observations to illustrate the potential of the approach.

## Example Pattern

One observation, made by coauthor Ananya Kassahun, who is a close friend of a person severely



**Figure 5.** Patterns of Facebook blog postings on April 16 and afterwards.

injured on April 16, is that frequent patterns of behavior emerged that day, such as that illustrated in Figure 5. Those who were in Blacksburg on April 16 were contacted by family and friends concerned about our safety, using a wide variety of communication schemes. Depending on the responses received, those making inquiries engaged in appropriate follow-on behaviors.

We seek to validate this pattern, understand what types of communication schemes were used, and connect actions with stress and trauma and other reactions. It will be hard to collect the data needed, but we are involved in survey efforts to help, and another research project is seeking voluntary submission of phone logs (using both mobile and land lines) to help with related analyses.

### Tags

Fortunately, some other types of analyses are easier to undertake, with less labor-intensive data collection

requirements. One line of research has to do with categories employed in connection with Web sites for collecting content, in particular the April 16 Archive (CDDC, 2007). That site allows people to tag files being uploaded. We considered about 800 items uploaded, extracted all the tags (i.e., words or short phrases) used with those, and summarized some key results in Table 12. This shows pairs of tags that co-occurred often in the collection of items. Thus, in 55 of the items in the collection, blog and commentary both appeared, showing the popularity of blogs for the purpose of making comments. As a second example, we note that the co-occurrence of Professor Librescu's name and the name of a leading Romanian magazine can help guide us to the news coverage in the country of origin of one of the heroes killed on April 16.

We can learn more about the cloud of tags connected with the collection by examining Figure 6, an information visualization that complements the

**Table 12.** Some High-Frequency Tag Pairs  
From the April 16 Archive

Tag 1	Tag 2	Count
Blog	Commentary	55
Candlelight	Vigil	43
University	Vigil	35
Hokie rock	Hokie stone	25
Drill field	Memorial	21
Blog	Cho	17
Second life	Memorial	16
Librescu	Omagiu	13
Mourning	Vigil	12
Omagiu	Alexandru Ioan Cuza	11
Memorial	Balloon	9
Art	Drawing	9
Blog	Gun control	9

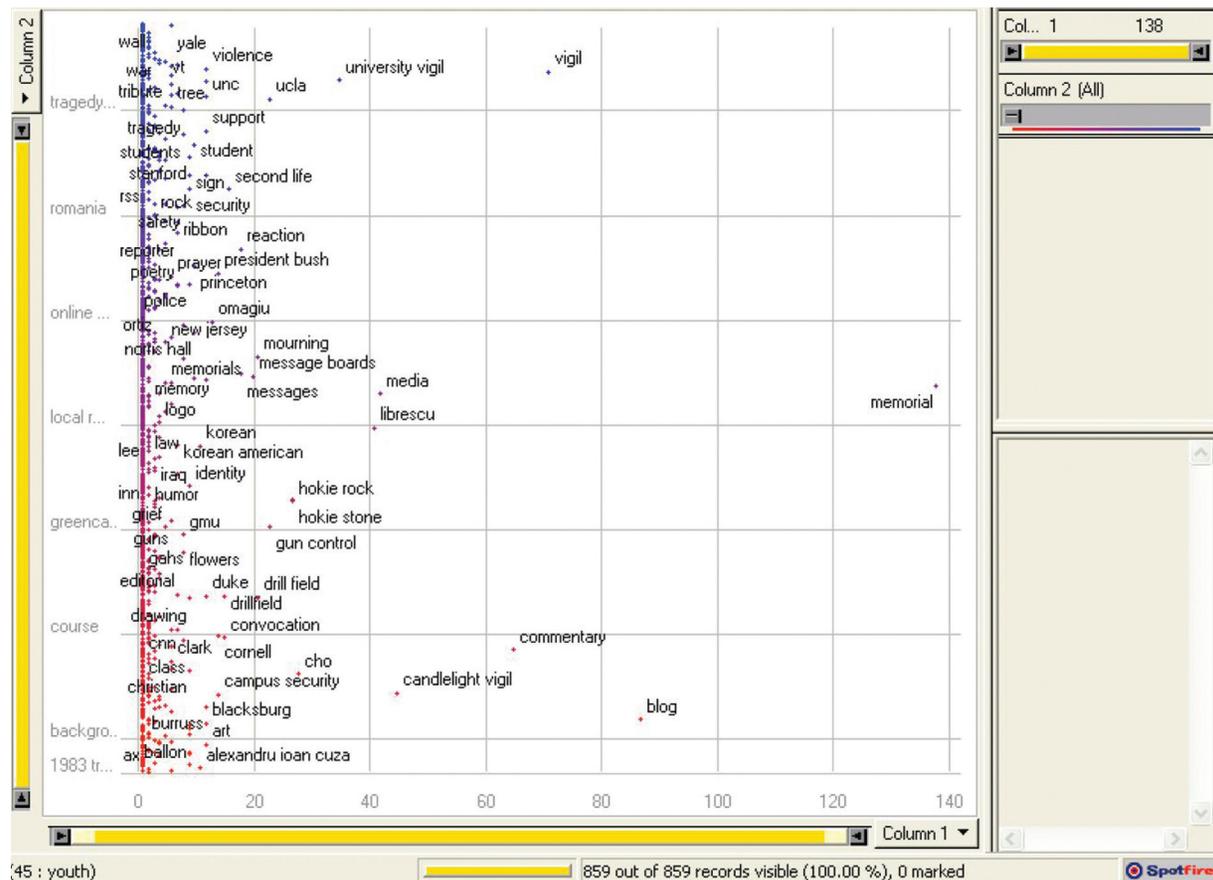
tabular presentation of Table 12. Here, we miss out on co-occurrence information but see the relative frequency of individual tags, the most popular of which is memorial. We see also that people were involved not

only in events such as campus vigil but also in electronic communication using blogs and other mechanisms to prepare commentaries (a trend mentioned above that also was spotted using a different analysis).

### Clusters of Text Words

To illustrate how texts in a digital library can be analyzed and presented using information visualization methods, we made use of the IN-SPIRE software suite (McQuerry, 2004; Wise et al., 1995). We used all the text from 800 of the contributions to the April 16 Archive (CDDC, 2007). The contributions are clustered according to a textual analysis of words and phrases in the text, and the clusters are labeled with frequently used words within the cluster.

Figure 7 provides a Themescapes view, where clusters of contributions are shown as mountains that form a landscape of themes. Figure 8 provides a Galaxy view, where each dot represents one contribution. Other than the differences in visual representation, these two



**Figure 6.** How often tags appear in the submitted works of the April 16 Archive.

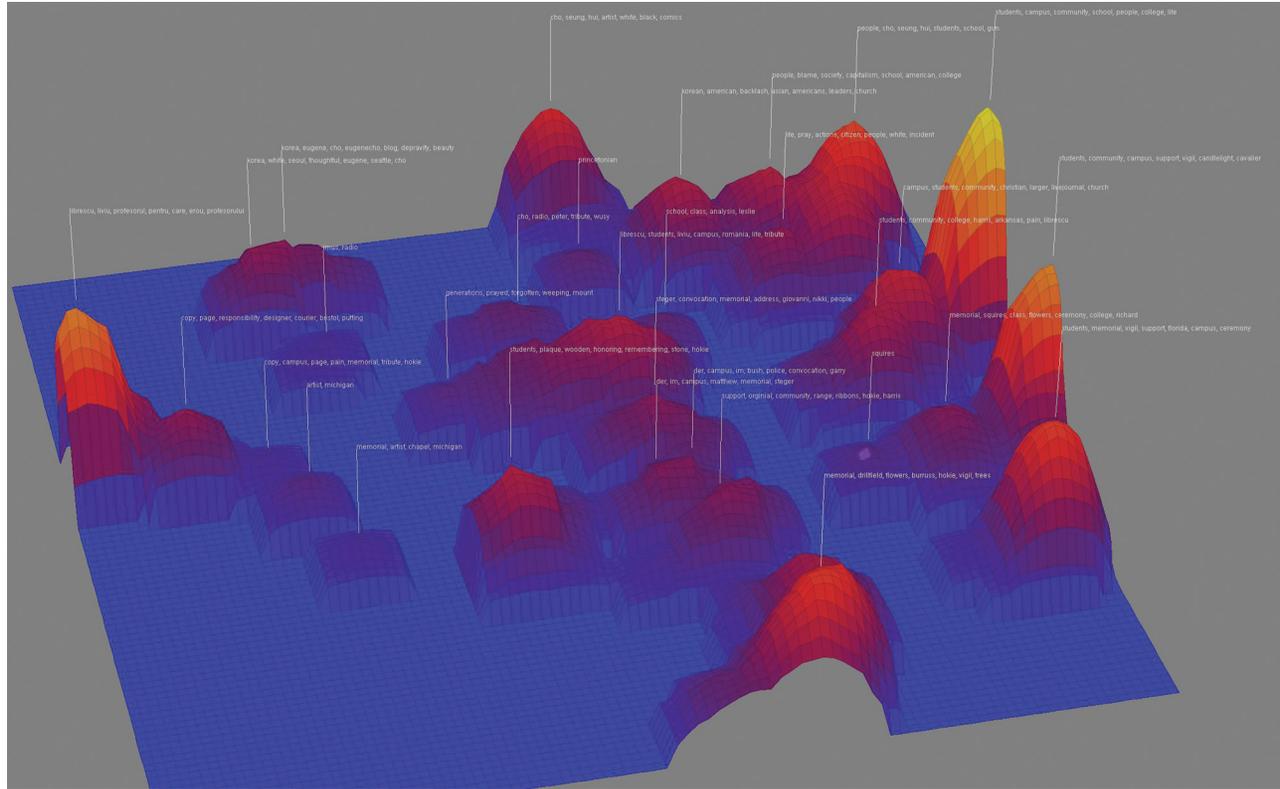


Figure 7. Themescape view from IN-SPIRE for all contributions to the April 16 Archive.

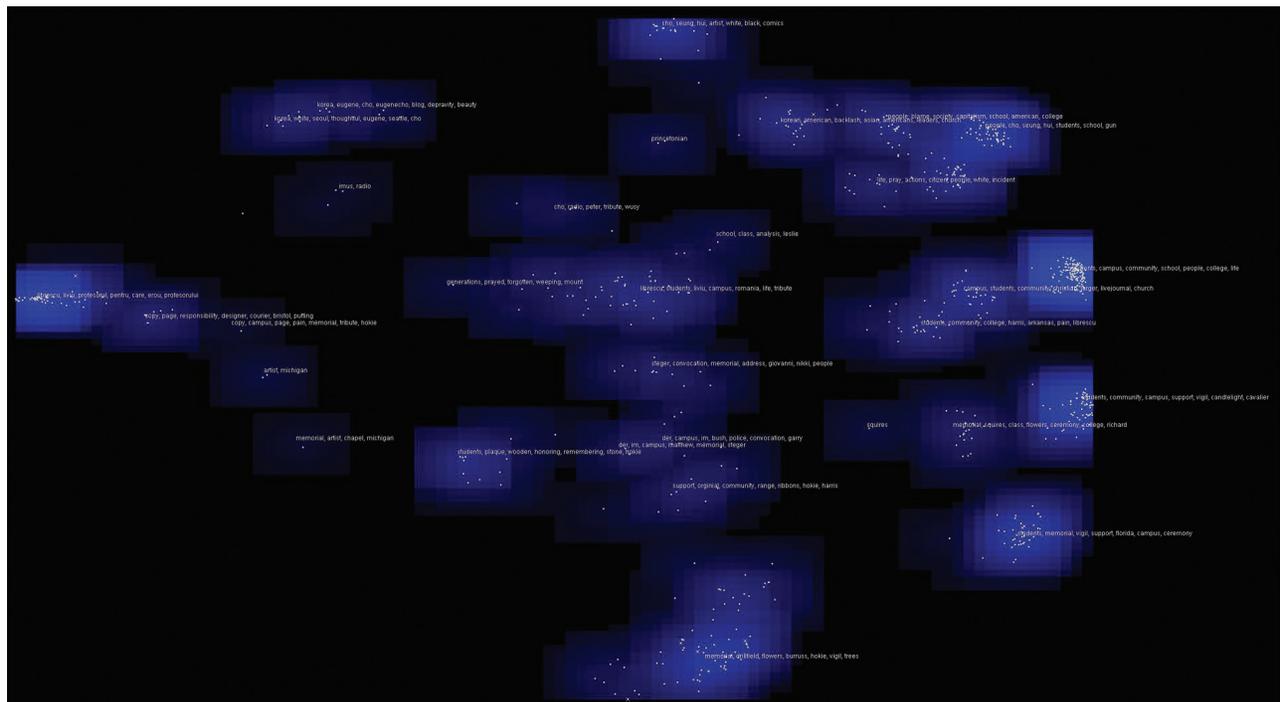


Figure 8. Galaxy view from IN-SPIRE for all contributions to the April 16 Archive.

views present essentially the same analysis and clustering layout. In particular, the placement of contributions and their groupings reflect a multidimensional scaling analysis, whereby closely related works are placed nearby and sections of the visualization reflect aspects or issues popular in the collection.

Thus, for example, on the right side is a large cluster about the student community, with some subclusters for the vigil and memorials. In the top-right is a cluster about societal issues. The middle areas are mainly about various events and memorials, such as the convocation, Drillfield, and so on. The left side seems to deal with external issues and foreign languages. Interest in Professor Librescu led to a sizeable cluster on the far left. Issues related to Korea are concentrated toward the top, especially on the left.

## Plans and Prospects

A digital library project such as that described herein, besides enabling the objectives discussed, undoubtedly also will result in an increase in the amount of information available to terrorists, as well as to nondestructive elements of society. Our group is conscious of the underlying trade-offs between free access to information and increased vulnerabilities. We reserve the right to restrict access to information, such as images of a graphic nature, and content for which preservation and scholarly study is approved, but where global access is not authorized by the contributors. We will adopt best practices of balancing security and cyber-rights in networked digital libraries, as recognized, for example, by Conway (2005).

This project had an August 15, 2007, starting date for our NSF-funded initiative, IIS-0736055, entitled "SGER: DL-VT416: A Digital Library Testbed for Research Related to 4/16/2007 at Virginia Tech." Although we worked between May and August to launch this effort, only in November could we begin to install equipment able to record and provide access to the terabytes of videos, audio files, high-quality photos, and other content collected through the many schemes outlined above. By early 2008, we had running in the Virginia Tech Computing Center a powerful server able to store 7 terabytes of information.

We plan to explore, demonstrate, and help others in their use of a growing suite of services attached to the digital library. Of particular benefit as we work with the local community to demonstrate the potential benefits is the very large display laboratory at Virginia Tech, including the gigapixel display (Ball & North,

2007). In the area of information visualization, we will make use of further tools in the IN-SPIRE suite, such as ThemeRiver (Havre, Hetzler, Whitney, & Nowell, 2002), to show themes and changes over time. In addition, we will also use other tools, such as Spotfire (Ahlberg & Shneiderman, 1994). In connection with the work on social network analysis, we will use information visualization tools scaled to work with Internet collections (Viegas et al., 2007).

Ultimately, we hope that the systems approach that has been so successful in science and engineering, and which has evolved to provide effective support in biology as part of the systems science trend (Saraiya et al., 2005), can lead to "systems behavioral and social science." We hope that powerful digital libraries can help during the ongoing recovery from April 16 and that when other tragedies occur in the future tailored digital libraries will be easier to devise and even easier to deploy, for even greater benefit.

## Acknowledgments

We thank three students taking Virginia Tech Department of Computer Science course CS4624 (multimedia, hypertext, and information access), who completed a fall 2007 term project related to this article and our effort: Phuc Lai, Kevin Franklin, and Brian Ho. They have helped us in our efforts to collect information for the digital library. We also thank NSF for the support of grant IIS-0736055. Finally, we thank the many people who have participated in focus group and other discussions and the Virginia Tech staff who have assisted with purchasing, setting up, and supporting the equipment involved in our work.

## References

- Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In S. Dumais & J. Olson (Eds.), *Proceedings of CHI'94. The SIGCHI Conference on Human Factors in Computing Systems: Celebrating interdependence* (pp. 313-317), Boston. New York: ACM Press.
- Amavizca, M., Sánchez, J. A., & Abascal, R. (1999). *3D tree: Visualization of large and complex information spaces in the floristic digital library*. Paper presented at the Proceedings of Segundo Encuentro de Computación (ENC'99), Pachuca, Hidalgo, México.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Arms, W. Y. (2000). *Digital libraries*. Cambridge: MIT Press.

- Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., et al. (2003). *Revolutionizing science and engineering through cyberinfrastructure* (Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure). Arlington, VA: National Science Foundation.
- Ball, R., & North, C. (2007). Realizing embodied interaction for visual analytics through large displays. *Computers & Graphics*, 31, 380-400.
- Breslau, N., Davis, G., Andreski, P., & Peterson, E. (1991). Traumatic events and posttraumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry*, 48, 216-222.
- Candela, L., Manghi, P., & Pagano, P. (2007, September). *An architecture for type-based repository systems*. Paper presented at the Proceedings of the Second Workshop on Foundations of Digital Libraries, in conjunction with the 11th European Conference on Research and Advanced Technologies on Digital Libraries (ECDL 2007), Budapest, Hungary.
- Carroll, D., Ebrahim, S., Tilling, K., Macleod, J., & Smith, G. D. (2002). Admissions for myocardial infarction and World Cup football: Database survey. *British Medical Journal*, 325, 1439-1442.
- CDDC. (2007). *The April 16 archive*. Blacksburg, VA: Virginia Tech Center for Digital Discourse and Culture. Retrieved January 25, 2008, from <http://www.april16archive.org/>.
- Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. *IEEE Computer*, 37, 50-56.
- Chen, S., & Fox, E. A. (1996). Guest editors' introduction to special issue on digital libraries. *Journal of Visual Communication and Image Representation*, 7, 1.
- Conway, M. (2005, September). Terrorist use of the Internet and fighting back. Paper presented at the Oxford Internet Institute conference: Safety and Security in a Networked World: Balancing Cyber-Rights and Responsibilities, Oxford University, UK.
- Domingos, P., & Richardson, M. (2001). Mining the Network Value of Customers, in *KDD'01: Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 57-66), San Francisco. New York: ACM Press.
- Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429, 180-184.
- Everson-Rose, S. A., & Lewis, T. T. (2005). Psychosocial factors and cardiovascular diseases. *Annual Review of Public Health*, 26, 469-500.
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49, 76-82.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39, 27-34.
- Fox, E., & France, R. (1987). Architecture of an expert system for composite document analysis, representation and retrieval. *International Journal of Approximate Reasoning*, 1, 151-175.
- Fox, E., Suleman, H., Madalli, D., & Cassel, L. (2004). Digital libraries. In M. P. Singh (Ed.), *Practical handbook of internet computing* (p. 4-1). Boca Raton, FL: Chapman & Hall/CRC Press.
- Fox, E. A. (1993). Digital libraries ("hot topics"). *IEEE Computer*, 26, 79-81.
- Fox, E. A. (1999, August). *The 5S framework for digital libraries and two case studies: NDLTD and CSTC*. Paper presented at Proceedings NIT'99, Taipei, Taiwan.
- Fox, E. A. (2007a, July). *LIKES: Living in the knowledge society*. Invited talk delivered at the University of Arizona, Tucson.
- Fox, E. A. (2007b). *Living In the KnowEdge Society (LIKES) Community Building Project*. Virginia Tech. Retrieved January 25, 2008, from <http://likes.dlib.vt.edu/>.
- Fox, E. A., Akscyn, R. M., Furuta, R. K., & Leggett, J. J. (1995). Guest editors' introduction to digital libraries. *Communications of the ACM*, 38, 22-28.
- Fox, E. A., Eaton, J., McMillan, G., Kipp, N., Weiss, L., Arce, E., et al. (1996, September). National Digital Library of Theses and Dissertations: A scalable and sustainable approach to unlock university resources. *D-Lib Magazine*, 2(8). Retrieved January 25, 2008, from <http://www.dlib.org/dlib/september96/theses/09fox.html>.
- Fox, E. A., & Marchionini, G. (1998). Toward a worldwide digital library: Guest editors' introduction to special section on digital libraries: Global scope, unlimited access. *Communications of the ACM*, 41, 29-32.
- Fox, E. A., & Sornil, O. (1999). Digital libraries. In R. Baeza-Yates & B. Ribeiro-Neto (Eds.), *Modern information retrieval* (pp. 415-432). Harlow, UK: ACM Press/Addison-Wesley-Longman.
- Fox, E. A., & Sornil, O. (2000). Digital libraries. In A. Ralston, E. D. Reilly, & D. Hemmendinger (Eds.), *Encyclopedia of computer science* (4th ed., pp. 576-581). London: Nature Publishing.
- Fox, E. A., & Urs, S. (2002). Digital libraries. In B. Cronin (Ed.), *Annual review of information science and technology (ARIST)* (Vol. 36, pp. 503-589). Silver Spring, MD: American Society for Information Science.
- Fox, E. E. D. (2003). *NDLTD: Networked Digital Library of Theses and Dissertations*. Retrieved January 25, 2008, from <http://www.ndltd.org>.
- Friedman, T. L. (2006). *The world is flat [updated and expanded]: A brief history of the twenty-first century*. New York: Farrar, Straus, & Giroux.
- Goncalves, M. A. (2004). *Streams, structures, spaces, scenarios, and societies (5S): A formal digital library framework and its applications*. Unpublished doctoral dissertation, Virginia Tech, Blacksburg. Retrieved January 25, 2008, from <http://scholar.lib.vt.edu/theses/available/etd-12052004-135923/>.

- Goncalves, M. A., & Fox, E. A. (2002, July). 5SL—A language for declarative specification and generation of digital libraries. In G. Marchionini (Ed.), *Proceedings of JCDL'2002, Second ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 263-272), Portland, OR. New York: ACM Press.
- Goncalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, structures, spaces, scenarios, societies (5S): A formal model for digital libraries. *ACM Transactions on Information Systems*, 22, 270-312.
- Gorton, D. C. (2007). *Practical digital library generation into DSpace with the 5S framework*. Unpublished master's thesis, Virginia Tech, Blacksburg. Retrieved January 25, 2008, from <http://scholar.lib.vt.edu/theses/available/etd-04252007-161736/>.
- Green, B. L., Wilson, J. P., & Lindy, J. D. (1985). Conceptualizing post-traumatic stress disorder: A psychosocial framework. In C. R. Figley (Ed.), *Trauma and its wake* (pp. 53-69). New York: Brunner/Mazel.
- Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8, 9-20.
- Hearst, M. A. (1995). Tilebars: Visualization of term distribution information in full text information access. In *Conference Proceedings on Human Factors in Computing Systems, CHI'95* (pp. 59-66), Denver, CO. New York: ACM Press.
- Hodges, S., Williams, L., & Wood, K. (2007). *Introduction to SenseCam*. Microsoft Research. Retrieved January 25, 2008, from <http://research.microsoft.com/sendev/projects/sensecam/>.
- Jung, V. (1999). *MetaViz: Visual interaction with geospatial digital libraries* (Tech. Rep. No. TR-99-017). Berkeley: ICSI, University of California.
- Kark, J. D., Goldman, S., & Epstein, L. (1995). Iraqi missile attacks on Israel: The association of mortality with a life-threatening stressor. *Journal of the American Medical Association*, 273, 1208-1210.
- Kaza, S., Marshall, B., Xu, J., Wang, A. G., Gowda, H., Atabakhsh, H., et al. (2005, May). *BorderSafe: Cross-jurisdictional information sharing, analysis, and visualization*. Paper presented at the Intelligence and Security Informatics Conference, Atlanta, GA.
- Lagoze, C., & Van de Sompel, H. (2007). *Compound information objects: The OAI-ORE perspective*. Open Archives Initiative. Retrieved January 25, 2008, from <http://www.openarchives.org/ore/documents/CompoundObjects-200705.html>.
- LaGreca, A. M., Silverman, W. K., Vernberg, E. M., & Prinstein, M. J. (1996). Symptoms of posttraumatic stress in children after Hurricane Andrew: A prospective study. *Journal of Consulting and Clinical Psychology*, 64, 712-723.
- Leor, J., Poole, W. K., & Kloner, R. A. (1996). Sudden cardiac death triggered by an earthquake. *New England Journal of Medicine*, 334, 413-419.
- Lesk, M. (1997). *Practical digital libraries: Books, bytes and bucks*. San Francisco: Morgan Kaufmann.
- Lesk, M. (2005). *Understanding digital libraries* (2nd ed.). San Francisco: Morgan Kaufmann.
- Lovullo, W. R. (2005). *Stress and health: Biological and psychological interactions* (2nd ed.). Thousand Oaks, CA: Sage.
- McQuerry, D. (2004). *IN-SPIRE™ visual document analysis*. Richland, WA: Pacific Northwest National Laboratory.
- Michaels, A. J., Michaels, C. E., Moon, C. H., Smith, J. S., Zimmerman, M. A., Taheri, P. A., et al. (1999). Posttraumatic stress disorder after injury: Impact on general health outcome and early risk assessment. *Journal of Trauma: Injury, Infection, and Critical Care*, 47, 460-467.
- MIT. (2003). *DSpace: Durable digital depository* [Web site]. Retrieved January 25, 2008, from <http://dspace.org>.
- North, C. (2005). Information visualization. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 1222-1246). New York: John Wiley.
- Nowell, L. (1997). *Graphical encoding for information visualization: Using icon color, shape and size to convey nominal and quantitative data*. Unpublished doctoral dissertation, Virginia Tech, Blacksburg. Retrieved January 25, 2008, from <http://scholar.lib.vt.edu/theses/available/etd-111897-163723/>.
- Nowell, L. T., & Fox, E. A. (1995, July). *Envision: Information visualization in a digital library*. Paper presented at Proceedings of ACM SIGIR'95, Seattle, WA.
- NSF. (2007). *Cyberinfrastructure vision for 21st century discovery* (NSF 07-28). Arlington, VA: National Science Foundation Cyberinfrastructure Council.
- O'Reilly, T. (2005). *What is Web 2.0: Design patterns and business models for the next generation of software*. Sebastopol, CA: O'Reilly Media. Retrieved January 25, 2008, from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.
- OCLC. (2004). *XTCat NDLTD union catalog*. Retrieved January 25, 2008, from <http://alcme.oclc.org/ndltd/index.html>.
- Petinot, Y., Giles, C. L., Bhatnagar, V., Teregowda, P. B., Han, H., & Councill, I. G. (2004, November). *A service-oriented architecture for digital libraries*. Paper presented at the Proceedings of the Second International Conference on Service Oriented Computing, New York City.
- Porges, S. W. (1995). Cardiac vagal tone: A physiological index of stress. *Neuroscience and Biobehavioral Reviews*, 19, 225-233.
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9, 1-6.
- Ramakrishnan, N., & Grama, A. Y. (1999). Data mining: From serendipity to science. *IEEE Computer*, 32, 34-37.
- Raymond, E. S. (1999). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. Sebastopol, CA: O'Reilly and Associates.
- Saraiya, P., North, C., & Duca, K. (2005). Visualization of biological pathways: Requirements analysis, systems

- evaluation, and research agenda. *Information Visualization*, 4, 191-205.
- Saraiya, P., North, C., Lam, V., & Duca, K. (2006). An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 12, 1511-1522.
- Seligman, M. E. P., Maier, S., & Solomon, R. L. (1971). Unpredictable and uncontrollable aversive events. In F. R. Brush (Ed.), *Aversive conditioning and learning* (pp. 347-400). New York: Academic Press.
- Shen, R., Vemuri, N. S., Fan, W. G., da S Torres, R., & Fox, E. A. (2006, June). *Exploring digital libraries: Integrating browsing, searching, and visualization*. Paper presented at the Proceedings of JCDL, Chapel Hill, NC. Retrieved January 25, 2008, from <http://doi.acm.org/10.1145/1141753.1141755>.
- Shen, R., Wang, J., & Fox, E. A. (2002). A lightweight protocol between digital libraries and visualization systems. In *JCDL Workshop on Visual Interfaces to Digital Libraries, Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries* (p. 425). Portland, OR: ACM Press.
- Spence, R. (2007). *Information visualization: Design for interaction* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Thayer, J. F., & Lane, R. D. (2007). The role of vagal function in the risk for cardiovascular disease and mortality. *Biological Psychology*, 74, 224-242.
- Thomas, J. J., & Cook, K. A. (Eds.). (2005). *Illuminating the path: Research and development agenda for visual analytics*. Washington, DC: IEEE Computer Society Press.
- Tollis, I. (1996). Graph drawing and information visualization. *ACM Computer Surveys*, 28, 19. Retrieved January 25, 2008, from <http://doi.acm.org/10.1145/242224.242247>.
- Van de Sompel, H., & Lagoze, C. (2000). Open Archives Initiative [Web site]. Retrieved January 25, 2008, <http://www.openarchives.org>.
- Viegas, F., Wattenberg, M., Ham, F. V., Kriss, J., & McKeon, M. (2007, October). *Many eyes: A site for visualization at Internet scale*. Paper presented at the Proceedings of IEEE Information Visualization (InfoVis) 2007, Sacramento, CA.
- Wang, J. (2002). *VIDI: A lightweight protocol between visualization systems and digital libraries*. Unpublished master's thesis, Virginia Tech, Blacksburg. Retrieved January 25, 2008 from <http://scholar.lib.vt.edu/theses/available/etd-07012002-145841/>.
- Weibel, S. (1995, July). Metadata: The foundations of resource description. *D-Lib Magazine*, 1(7). Retrieved January 25, 2008, from <http://www.dlib.org/dlib/July95/07weibel.html>.
- Weiss, J. M. (1991). Stress-induced depression: Critical neurochemical and electrophysiological changes. In I. J. Madden (Ed.), *Neurobiology of learning, emotion, and affect* (pp. 123-154). New York: Raven.
- Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., et al. (1995). Visualizing the non-visual: spatial analysis and interaction with information from text documents. Paper presented at InfoVis, Proceedings on Information Visualization (p. 51), Atlanta, GA.
- Wong, P. C., Whitney, P., & Thomas, J. (1999, October). Visualizing association rules for text mining. In *IEEE Information Visualization '99* (pp. 120-123), San Francisco. Los Alamitos, CA: IEEE Computer Society Press.
- Zhu, Q. (2002). *5SGraph: A modeling tool for digital libraries*. Unpublished master's thesis, Virginia Tech, Burlington. Retrieved January 25, 2008, from <http://scholar.lib.vt.edu/theses/available/etd-11272002-210531/>.