1 Purpose, Objectives and Research Questions

Crises and tragedies are, regrettably, part of life; a recent sample, showing the small number of collections preserved at the Internet Archive, is shown in Table 1. While always difficult, recovery from tragic events may be increasingly facilitated and supported by information and communication technology (ICT). Individuals, groups, and communities are using ICT in innovative ways to learn from these events and recover more quickly and more effectively. During and after a crisis, individuals and communities face a confusing plethora of data and information, and strive to make sense by way of that data [114]. They seek to carry out their usual activities, but want to be informed by new insights. They work to help others, or to receive help, but the context and technologies involved in communication today (e.g., Internet, WWW, online communities, mobile devices) make it exceedingly difficult to integrate content, community, and services. Accordingly, individuals and communities respond by attempting to meet their needs with the tools they have, e.g., creating a Facebook group to quickly inform members who is OK, and other groups to share pictures, comments, and additional contributions.

Table 1. Internet Archive Collections on World Wide Crises and Tragedies			
Year	Event	IA Collection	Wikipedia suffix, other URLs[1]
2004	Asian Tsunami	2004	Indian_Ocean_earthquake
2007	Burmese Uprising	937	2007_Burmese_anti-government_protests
2007	California Wildfires	877	California_wildfires_of_October_2007
2008	Georgia and Russia Conflict	1120	2008_Georgia–Russia_crisis
2005	Hurricane Katrina	174	Hurricane_katrina
2008	Iowa Flood	1092	Iowa_flood_of_2008
1998	Matthew Shepard murder	1075	Matthew_Shepard
2008	N. Illinois U. Shooting	970	Northern_Illinois_University_shooting
2008	Tibet protests	1044	Tibet_protests
2007	VT April 16 Shooting	694	Virginia_Tech_massacre
2008	Zimbabwean crisis	1048	Zimbabwe

We will research the urgent problem of integrating content, community, and services related to crisis, tragedy, and recovery (CTR). Digital libraries can manage data heterogeneity and a large volume of information. That information is critical to many users; Table 2 demonstrates the strong public interest in, and need for, such information. An integrated network of digital libraries, drawing content from direct submissions, special collections, and both Web and Web 2.0 sources, coupled with a CTR specific ontology, providing a rich suite of advanced services, can help people to acquire tailored domain specific information in crisis situations. This integrated network of resources also can support social science research and broader utilization by scholars, policy makers, and the general public. For example, the tragic shooting at Virginia Tech on April 16, 2007 motivated our university community in Blacksburg, Virginia to build such a digital library [30, 32], as we grappled with recovery. It brought researchers together – from the computing and information sciences, and the social and behavioral sciences – to seek effective approaches and solutions that also have wider applicability.

This research will build a key portion of our nation's cyberinfrastructure, which we call CTRnet. We move beyond conventional database and information systems into a comprehensive integration of content, community, and services – building upon and extending our ongoing research with digital libraries. We will use, adapt, and develop content, services, and applications that best connect people, specialized communities, and the world-wide supportive public. Those involved in crisis and tragedy situations

---

[1] Prefix for Wikipedia articles is http://en.wikipedia.org/wiki/. See there suffixes: List_of_hostage_crises, List_of_terrorist_incidents, Disaster, Natural_disaster, Man-made_hazards, etc. Regarding 9/11, see, e.g., http://911digitalarchive.org and http://www.poetry.com/us_tragedy/searchgroup.asp.

include emergency response workers, law enforcement agencies, government officials, and community activists who help serve the public during and after catastrophic incidents. An even broader range of people are involved in the years that follow, especially since we will archive a key part of the Web [4].

Table 2. Popular CTR-related queries (from Google Zeitgeist)	
Month	Keywords
April 2007	virginia tech; cho seung-hui; columbine; darfur
October 2007	ohio shooting; pensacola weather; california fires; ocean isle fire; earthquake
February 2008	Northern Illinois University Shooting; Cole Hall shooting; Kazmierczak
March 2008	Georgia-Russia Crisis; Kosovo independence; Georgian UAV shot down
July 2008	California Earthquake; Los Angeles 5.4; World's most seismically active region
November 2008	Mumbai shooting; India shooting; Deccan Mujahideen; Mumbai shootout photo

We seek to help people who have been affected by tragedies, and/or are interested in CTR related information. Our service aim is to demonstrate how next-generation digital libraries can support such needs. Our technical aim is to achieve the broader intelligent information integration "holy grail" of shifting to a new paradigm of digital library research where all sources (from uploading, collecting, Web, and Web 2.0 sources) are brought together with a rich suite of services to support broad domain specific individual and community needs. Our goals include:

- Integrating heterogeneous information in a specific domain, making it accessible, and preserving it for long-term reuse;
- Extending the scope of digital libraries so they are closely but flexibly coupled with a wide variety of services to support diverse emerging communities; and
- Supporting information exploration with advanced methods (SSP [13-15, 28, 139], PathRank [140], and Storytelling [71]) that facilitate searching, browsing, and discovery.

We will build a domain specific distributed digital library, integrating information from many sources. Two new access points, a digital library website and a social network application, will be developed as part of our test bed for CTRnet. Our social networking application will use the APIs of Facebook and other social networks, e.g., MySpace, to make CTR information accessible from within these systems in an integrated manner to a wider audience. We will begin with our April 16 system and extend it to serve more global CTR situations. Thus, in order to assist the CTR community in preserving and analyzing information, we will implement a CTR toolkit. Its software will provide services such as searching, browsing, recommending, personalization, multilingual access, statistical analysis, data set management, literature and linguistics analysis, and information visualization. Our advanced software for exploration will assist with finding relevant content: individual items, sequences of items, and explanatory stories. The questions we address include:

- How can parts of CTRnet be built semi-automatically, drawing upon related digital libraries, web pages, query logs, Web 2.0 applications, and other Internet resources?
- How can this CTRnet be utilized, efficiently and effectively, for a wide variety of tasks?
- What kind of user interfaces can facilitate building and utilizing the CTR network?
- How can we integrate information from other digital libraries, and from the Internet Archive and the Library of Congress, along with the current Web, to make it more accessible?
- How can our solution be evaluated and validated, leading to a widely used methodology?

The combination of CTRnet and the CTR toolkit could be a powerful aid in the CTR domain. With major tragedies seemingly occurring daily, our system will aid recovery in our nation and around the world.

2   Related and Prior Work
2.1   Crisis Informatics
Recently, a new area called "Crisis Informatics" has emerged, where researchers are studying connections between socially and behaviorally conscious ICT, and results from empirical research in crisis situations [81]. Being a union of information, social, and computer science, Crisis Informatics aims to bring

together the changing role of technology in social areas during times of emergency, disaster, and other forms of crisis scenarios [80]. The goal is to be aware of the information and technology needs of all members, including responders and the public, in disaster situations.

Many studies investigating emergency preparedness and crisis situations have been conducted, including Turoff and Hiltz's work with NLM [128], Schneider and Foot's investigation of information-seeking behavior after the September 11, 2001 attacks [108], and Kim, Jung, et al.'s study on the relationship between "internet connectedness" and communicative actions after September 11 [67]. The University of Colorado at Boulder hosts a Natural Hazards Center (http://www.colorado.edu/hazards/) with mission overlapping that of our initiative. Although many studies focus on analyzing crisis-related data and use of applications in response to crisis situations, there is limited research on using this data to build a network to dramatically enhance and integrate the currently highly varied user services. Traditional services such as browsing, searching, and recommending clearly do help, serving not just the general public but also specialized communities. But those services will be improved further through CTRnet, and other specialized services can be introduced to better support government officials, policy makers, learners, and researchers.

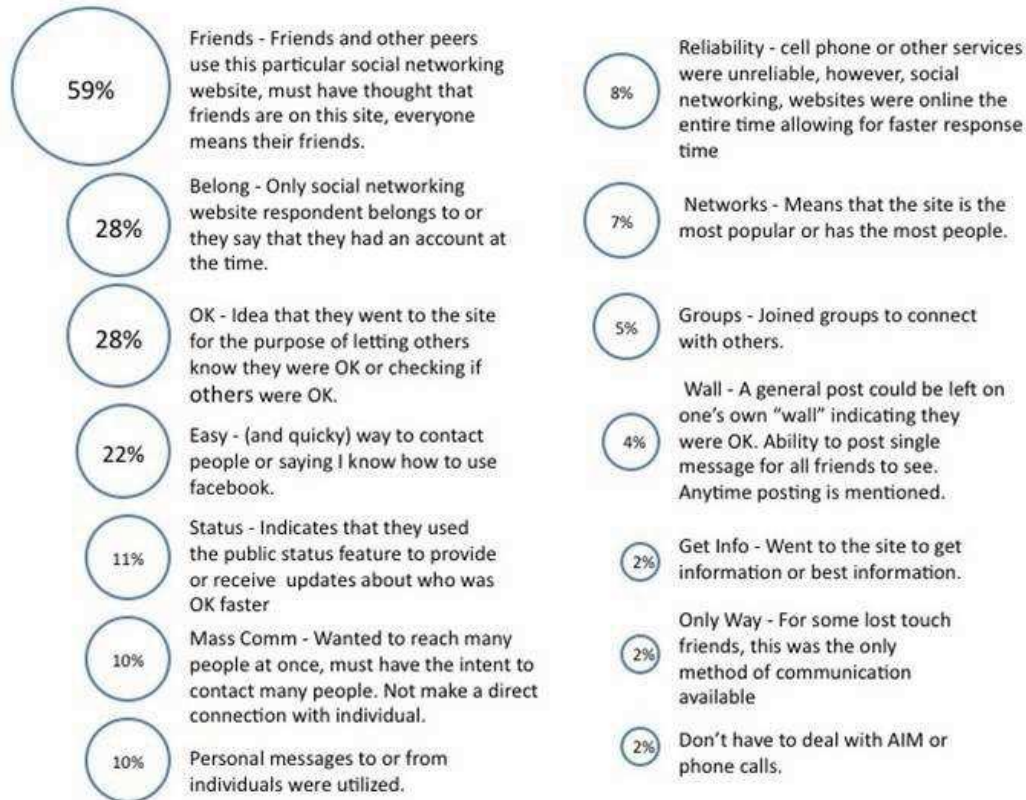## 2.2   Building Ontologies, Semantic Search

Conventional ontology building approaches involve knowledge acquisition, followed by manual topic and relationship identification. This results in high quality ontologies, but the process is time consuming. On the other hand, fully automatic approaches like Text2Onto [11], MoK [2], and OntoLearn [130] may yield low quality results. Semi-automatic approaches, as by Wang et al. [134] and Fortuna et al. [22], seem effective, but may be hard to scale. Folksonomies are flexible and scalable, but may have variable quality. Yet, a CTR ontology would be valuable for integrating information across tragic events.

Ontology driven navigation appears promising [54, 92]. The ontology structure is navigated (from general to more specific) to generate results matching the users' query. During browsing, the user could be guided by an ontology, and asked to select a more meaningful query [54]. We will employ ontologies for browsing and semantic query expansion (of continuing interest [6, 8, 23, 74, 133, 138]), in CTRnet.

2.4   Social Networks and Web 2.0

"Web 2.0" refers to collaborative and interactive value-added services, extending the WWW's content hosting services [79]. The collaborative aspect of Web 2.0 refers to the social interaction among users and their collective creation of content – which, during crises, is amplified as social networks activate to understand, and mobilize to cope. In response to needs for information, sites are created, messages are sent, and information is contributed in a rapid fashion, as ad hoc social networks form that merge existing network segments, e.g., students and local community members. Before the Internet, community members largely contributed through participation in events or impromptu memorials, in timeframes of days for widespread knowledge of an event, and longer for informing all involved. Today ICT brings breaking news live to the world. Each crisis impacts more people more quickly than ever before. More importantly, people have many more options for communicating more quickly than ever before. The wide distribution of cell phones that provide services of messaging, email, and instant-messaging – coupled with the extension of broadband always-on networks – have enabled the emergence of a culture of connectedness among segments of society. Sites such as Facebook, Flickr, and MySpace provide a platform for users not only to browse content contributed by their peers but also to participate in discussions, form focus groups, and upload and share their perspectives through messages, photos, and videos. These responses to tragedy occur through these platforms because they are the contexts of modern life, i.e., people respond by joining groups in Facebook, because they are already there and have the social desire to share their ideas and feelings. New advances in ICT have now enabled emergent phenomenon of dynamically self-organizing groups that build parallel information grids to facilitate community understanding and healing. Accordingly, in our study of the tragic shootings at VT on April 16th, we asked people why they used a social networking site in relation to the tragedy. The vast majority of respondents (students) identified Facebook as the site they chose, with only 10 references to other sites across 426 responses. A content

analysis approach was applied to derive keywords that capture intended meanings of the ideas, beliefs, and concepts contained in the responses. Three trained coders independently assigned keywords to each response. Fig. 1 shows the keyword definition and the percentage of responses to which it was assigned.

59% Friends - Friends and other peers use this particular social networking website, must have thought that friends are on this site, everyone means their friends.

28% Belong - Only social networking website respondent belongs to or they say that they had an account at the time.

28% OK - Idea that they went to the site for the purpose of letting others know they were OK or checking if others were OK.

22% Easy - (and quicky) way to contact people or saying I know how to use facebook.

11% Status - Indicates that they used the public status feature to provide or receive updates about who was OK faster

10% Mass Comm - Wanted to reach many people at once, must have the intent to contact many people. Not make a direct connection with individual.

10% Personal messages to or from individuals were utilized.

8% Reliability - cell phone or other services were unreliable, however, social networking, websites were online the entire time allowing for faster response time

7% Networks - Means that the site is the most popular or has the most people.

5% Groups - Joined groups to connect with others.

4% Wall - A general post could be left on one's own "wall" indicating they were OK. Ability to post single message for all friends to see. Anytime posting is mentioned.

2% Get Info - Went to the site to get information or best information.

2% Only Way - For some lost touch friends, this was the only method of communication available

2% Don't have to deal with AIM or phone calls.

Figure 1. Why Social Networking during Tragedy Content Analysis Coding Results

Respondents went to Facebook because they see the site as easy to use, their friends are there, and they had a page on the site prior to the tragedy; they wanted to let their friends know they were OK, and find out that others were OK. They used the site for both personal communications with individuals and to send a mass communication to everyone that looked at their page or the membership of a group they had joined. Another interesting aspect that emerged is that Facebook allowed people to maintain contact with multiple, mostly separate, social networks, e.g., friends from high school and friends from VT. In some cases this was the only way to make contact with some networks of friends for which the respondent no longer had correct contact information. It seems likely that similar uses of social networking sites will occur in future tragedies as they become more embedded in the lives of more members of society.

    This project will provide tools customized to the needs of communities enduring a crisis through a determination of the range of information needed in crisis and recovery situations. It also will define a digital library and system architecture with the necessary services. One aspect of the architecture is a package of services related to interactions with social networking websites. This is important, since approaches for capturing information from the Web 1.0 world, e.g., crawling, may not be allowed in the social networking etiquette. Web 2.0 is an expression of the voice of the people that must be recorded and preserved. Yet, such information is not systemically captured, and often disappears. Only through new embedded applications can data be captured, broader access be provided, and preservation ensured. Thus we will help with building applications and services within Web 2.0 sites. In so doing, we will leverage the tools provided within these sites, e.g., Flickr released a formatting convention – machine tags – that offers a way to assign semantics to Flickr images using common terms. Similarly, Google Maps [41]

provides application programming interfaces (APIs) to build services that integrate mapping functions. Further, OpenSocial is a software product by Google that provides developers with the ability to develop applications that work in multiple social networking sites, e.g., MySpace and Friendster, designed specifically to enable the development of applications such as that envisioned for this project, albeit focused on shopping or other more common activity. Services developed with these tools have the potential to provide great value in crisis situations by bringing information to people where they are.

A new paradigm involving a service oriented architecture has been applied [88] wherein different digital libraries (that are inherently heterogeneous, distributed, and geographically separated) can present a single uniform interface to users through Web services hosted at various sites. This project will build such a network for the CTR domain. 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 and access.

3   System Stakeholders

When a tragedy occurs, an information exchange network tends to form. We propose that these users communicate using our CTR digital library as in Fig. 2.
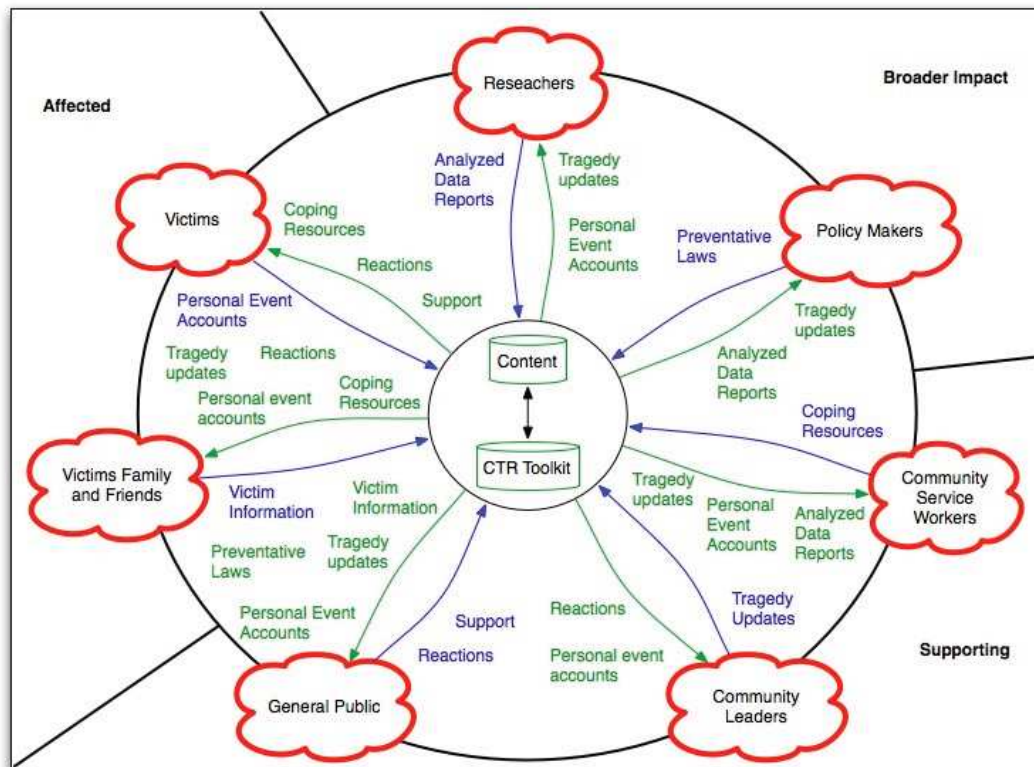

Figure 2. Local CTR Digital Library

The involved parties can communicate with one another using resources such as blogs, wikis, forums, Facebook, MySpace, Google or Yahoo groups, Flickr, etc. However, large amounts of information from these sources are unstructured or semi-structured. They often have issues with data heterogeneity, a large volume, and a lack of structure. There is need to filter, to obtain relevant and up-to-date information. Thus, Fig. 2 describes the system stakeholders plus the information exchange that occurs between them, during and after tragic events. Our CTR digital library includes content and software services packaged in our CTR toolkit, which will provide a rich suite of services. CTRnet will build cyberinfrastructure whereby these stakeholders can contribute as well as retrieve CTR-related information.

4   Approach

Our approach will be to build a distributed digital library network, CTRnet (Sec. 4.2), founded on our 5S theoretical framework (Sec. 4.1), where nodes can run our software toolkit (Sec. 4.3) affording a rich suite of services (Sec. 4.5), which also includes a full portal from each crisis Facebook site (Sec. 4.4). Since information integration is so important, it will be built into the CTRnet architecture (Fig. 3), and supported for end-users by our special software for exploration (finding pathways and stories, Sec. 4.5.1).

4.1   5S Framework as Theoretical Foundation for System Design, Implementation, and Evaluation

Recognizing the difficulties in understanding, defining, describing, and modeling digital libraries (DLs), Gonçalves et al. [24] have proposed and formalized the 5S (Streams, Structures, Spaces, Scenarios, and Societies) framework of digital libraries (DLs) [36, 37, 40]. 5S provides a formal framework to capture the complexities of DLs. The definitions in [37] unambiguously specify many key characteristics and behaviors of DLs. This also enables automatic mapping from 5S constructs to actual implementations as well as the study of qualitative properties of these constructs (e.g., completeness, consistency) [3].

We have used 5S as a guiding framework to successfully design and deploy a very large integrated DL for Archaeology – ETANA-DL [89, 94, 110, 111]. It has been used to guide efforts for the VT416DL project [31] as well as various educational initiatives [26, 27]. 5S will guide our Ensemble [29] NSDL computing pathway work, and work on CTRnet. Table 3 (discussed later) shows a modified version of our planned services taxonomy, based on the 5S framework [42]. We also will use 5S-based software for logging, evaluation, and reporting, so all parts of CTRnet can be continuously improved [37, 39, 75, 76].
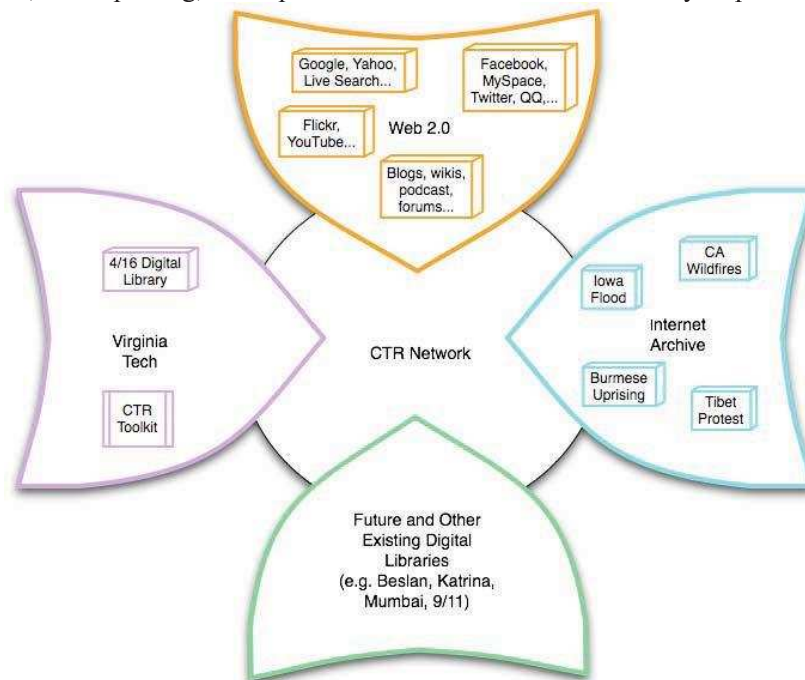

Figure 3. CTR Network

4.2   CTR Network

Fig. 3 depicts the design of the CTRnet architecture, illustrating four components, reflecting the broad scope of content and the variety of partners and services. The CTR Network is a distributed digital library, accessible through the Web, bringing together communities in and across crises and tragedies. CTRnet allows users to perform a search and return results from multiple digital libraries. It aims to inform and help individuals and communities heal.

The node on the left side of Figure 3 is the Virginia Tech component. There are two initial nodes within this component, 4/16 Digital Library and CTR Toolkit. The 4/16 Digital Library is related to the

6

April 16, 2007 shooting on the Virginia Tech campus [31]. The CTR Toolkit node provides a broad suite of services, and access to the 4/16 shooting digital library from CTRnet. Virginia Tech will offer to crisis sites temporary or long-term aid by way of additional digital library nodes and toolkit instances.

The Future and Other Existing Digital Libraries component supports expansion of CTRnet. An example partner could be from Mumbai, India. The recent attack there led to many dead or wounded, and caused pain throughout India. New CTRnet nodes are intended to serve communities like Mumbai. A tailored social networking application within Facebook could contain the latest news on the tragedy; plus information on local assistance, healing groups, and centers for people to go and talk about what happened; and could lead to a permanent digital memorial for all who suffered.

The Internet Archive (recall Table 1) component provides capabilities of Web crawling, acquiring information, and preservation. Local communities, likely without adequate technical or financial resources, do not need to host their own collections or crawl for content to include; we would coordinate work with the Internet Archive (see letter of support).

The final component is integrating Web 2.0, top center of Figure 3, into a seamless user interface that provides results from the CTR-net within social networking sites, and provides site users the ability to preserve their information by contributing it to the digital library on the CTR-net.

4.3   CTR Toolkit

We propose the use of a CTR Toolkit to help people affected by a crisis. During and after a tragic event, there are a series of needs that have to be addressed. The timeline in Fig. 4 highlights some of the data/information requirements related to longitudinal recovery from, and study of, such events.

Between the day of, and week 1 after, the tragedy, immediate response data are needed. Virginia Tech and others in CTRnet could assist. Or, when local archivists are activated and realize they want to capture digital information, Virginia Tech would send the CTR toolkit, allowing entry and preservation of information about the event into a digital library. People in the community would immediately be able to browse information about the tragedy. Between week 1 and month 1, the post-event initial recovery process begins. The community will be able to browse information about the tragedy using CTR-net, as community leaders and others share information.

Between month 1 and month 6, the community may launch long-term recovery planning. Data collected and analyzed by researchers will clarify how (if not why) the tragedy occurred and how the community has been dealing with it. Also, they will be able to compare it with prior similar events.

Between month 6 and month 12, anniversary events will occur. During this time, the public will reflect on the event that occurred within the last year. Lawmakers may consider how to prevent a similar tragedy from occurring. With major tragedies occurring worldwide, a CTRnet system can influence how people respond to and recover from these events.
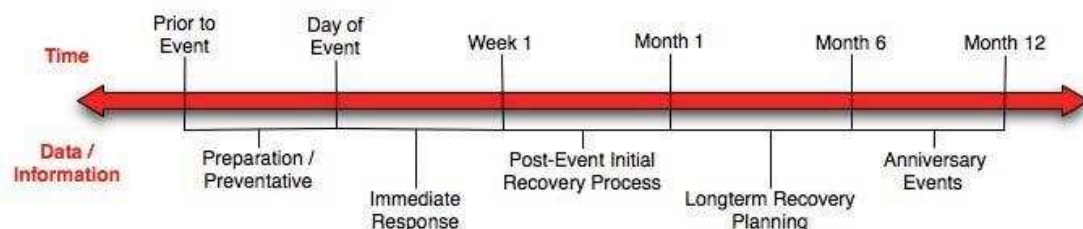

Figure 4. CTR Timeline

We can build upon lessons learned over the last years at Virginia Tech to provide immediate and ongoing support for related needs at other sites. The toolkit will provide a rich suite of services to help preserve, exchange, visualize, and analyze information; see Fig. 5.

To help avoid tragic events, preventative information should be widely disseminated. Thus, after

April 16, 2007, many universities investigated how to give emergency instructions to their communities. In case of looming natural disasters, evacuation plans can reduce damage, injury, and loss of life.
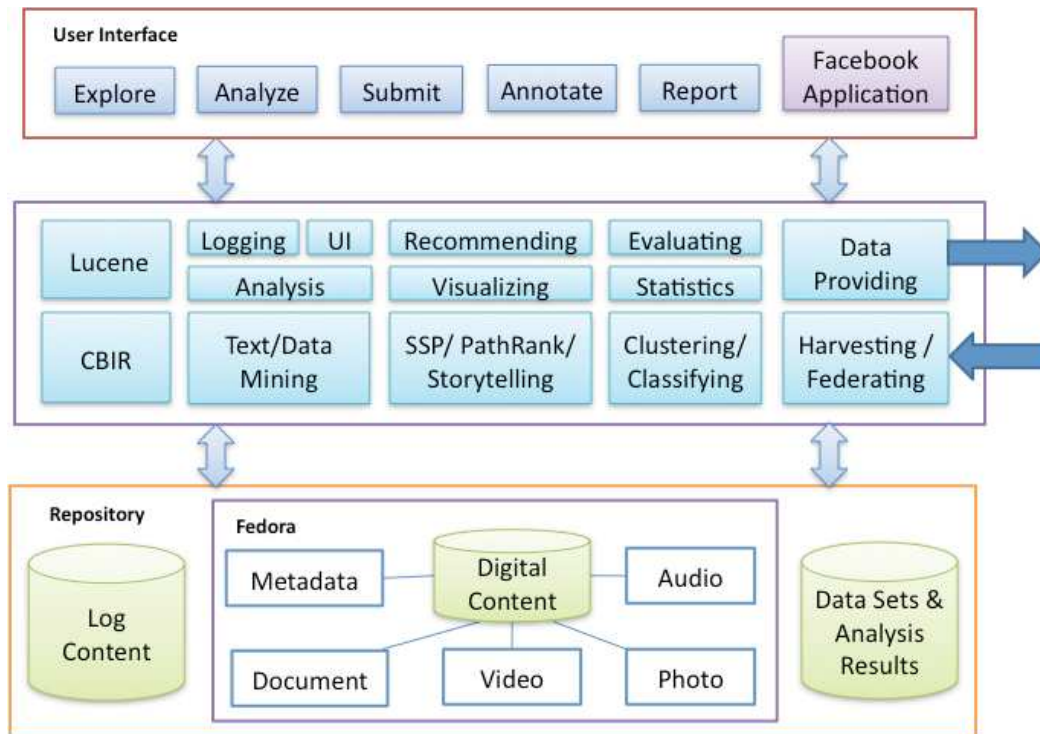
Figure 5: CTR Toolkit Services

The CTR Toolkit could be packaged to run on a modest computer, with necessary software components installed and configured. The server component will allow users to add photos, stories, and links to news coverage. There will be help files, pointing to local and global crisis and tragedy recovery groups (such as counseling or shelter providers), explaining for community leaders how to create healing groups for those going through the tragedy, and integrate crisis information sources.

The CTR Toolkit will include a repository storage system. Fedora offers many different services that accompany digital libraries. Besides repository services, Fedora [82, 116, 117] offers preservation, enterprise, and semantic services. The Fedora framework is highly scalable and configurable. Fedora offers an integrated search engine to help search the repository. It supports replication, format conversion, federation, and harvesting through the OAI-PMH protocol [118, 129]. It also will allow integration of a wide variety of tools developed in related projects at Virginia Tech [7, 18, 24, 31, 34, 35, 39, 40, 42, 43, 65, 66, 75-78, 86, 87, 90, 91, 95-97, 100, 102-105, 110, 112, 113, 124-127, 131, 136, 137, 139, 141-147].

The Toolkit will include data and text mining tools, as well as sample surveys for collecting data consistent with the CTR-net nodes. Social science experts will study the data, try to anticipate change, assess benefits, and answer questions such as: Why did this happen? How did this affect the community mentally? How can we prevent such in the future? Is prevention possible? If not, how can we prepare? What actions will derive the optimal outcome?

4.4   Social Networking Application

Facebook is a popular social network used in schools, colleges, and even professional settings today. We will create an application that runs within Facebook to provide any community concerned about a crisis or tragedy with an instantly available resource that is under their control. They will be able to create a group, for the crisis or tragedy just experienced. We will provide a manual, and a set of guidelines, that will

jumpstart the use of social networking to begin the recovery process. We will provide all the information necessary to set up their own page with a tested look and feel that includes enough information to attract users to the group. For example, providing a zip code in the "Location field" will link to a Google Map, local weather, lists of the food and housing shelters, aid groups, local counselors, government contact information, etc. We will anticipate these needs based on our experiences with digital library at VT.

Our social networking application will include many services for CTR events. The application will provide a range of information and specifics of the event informing the community what has happened, such as time of day, history of event, maps, descriptions, demographics, and community responses, as reported in the news and updated by community members. A timeline of the event will provide one method of navigating the information. The application will work within the context of the site to integrate various content areas common to social networking sites, e.g., Photo section, Wall, Video section, Posted Items, and Members View section. A list to allow people to post notes saying they are OK from the event will be available. Multiple news feeds, which can be dynamically modified by the community, will be available in a categorized table showing items such as: community announcements, Government announcements, police announcements, victims, etc. The panel will allow filtering based on the CTR ontology to provide information that users need. There will be a panel for artifacts from the event: poems, stories, event history, interviews, and so forth. Social network site members will be able to post, tag, and search videos, photos, posted items, comments, etc. Tagging of items using the CTR ontology will be emphasized to leverage the involvement of the community, and to attach shared meanings to the elements of the digital library.

The CTR digital library will be able to access the elements of the social networking site based on the permission given by individual users. When users post items, they will have the option to share the information with the digital library so it will not violate any user identification/legality issues, and to tag it using the CTR ontology. Once a social networking site user has shared photos, videos, comments and approved their distribution, they will be preserved in the digital library and by the Internet Archive.

4.5   Services

Table 3 defines the key CTR toolkit services. Columns 1 and 2 list infrastructure services. Column 2 identifies services that aggregate information from their inputs or connect objects together [34]. The output of the infrastructure services is the input for the information satisfaction services listed in column 3. As appropriate, other services can be incorporated into the CTR toolkit and CTRnet, building on a services oriented architecture [12, 88].

Table 3. Key services involved in CTR construction		
Infrastructure Services		**Information Satisfaction Services**
Repository-Building	**Add Value**	
Acquiring [131]	Classifying [144, 146, 147]	Browsing *(Website, Facebook)*
Cataloging and Annotating [77, 78]	Clustering [9, 25]	Collaborating
Crawling [10]	Entity Extraction and Integration	Customizing *(Website, Application, using CTRnet)* [84, 93, 96, 126, 127]
Digitizing	Evaluating [75, 76]	Filtering
Federating *(Integrator)* *[111, 112, 131]*	Auto Tagging [141-143]	Providing access *(Website, and Facebook application)*
Harvesting *(Wrappers,Integrator, Extraction)* [111, 113, 131]	Publicizing *(Sharing information on Facebook application, Website)*	Recommending [85]
Submitting *(Done by user on Website and application)*	Rating *(Social tagging, and inputs from Facebook application, Website, and common algorithms)*	Searching *(Semantic Query Expansion on Website and Facebook application)*
Indexing *(Ontologies, RDF Graphs)*	Ontology building [40]	Visualizing *(through website and Facebook application)*[66]

4.5.1 Using SSP, PathRank, and Storytelling to find connections among concepts and events in a CTR Digital Library

Currently there are already a multitude of articles, pictures, and other information organized in our CTR digital library [32]. Sometimes, when learning, researching, or exploring in general, a user may want to gain insight into the connections between two events, two people, two documents, two topics, etc. Alternatively, query splitting can yield two queries from a single long one [139]. Due to the amount of material, we need an automated approach to find the best connections among all the possible paths. We have developed several approaches, including SSP (Stepping Stones and Pathways) [13, 15, 28, 139], which is used to retrieve chains of relationships between articles. Xiaoyan Yu's doctoral research is further extending that work with the PathRank (PR) approach [140]. Separately, Ramakrishnan, Gresock, and Kumar [44, 45, 71] proposed the Storytelling approach; it yields interesting stories on a biomedical dataset, refined through a series of filtering and compression operations over the mined stories. The common characteristics of SSP, PathRank, and Storytelling are that they all can be treated as an automatic link finder, and hence can be used to connect the topics or events people are interested in.

So, given two endpoints, such as two events/topics/articles, SSP, PathRank, and Storytelling can identify the chain of intermediate information carriers from one to other, ensuring that any neighboring intermediate points have a relatively high similarity value. SSP uses interactive real-time information retrieval technology. Whenever two endpoints are identified, SSP enriches them separately, by adding keywords from the top relevant results to the initial endpoints. After that, the enriched endpoints are queried again, and the common sets of the results will be identified as the intermediate pathways and also be organized nicely by the clustering of these results. Storytelling works in a somewhat different way, employing redescription data mining technology. At the beginning, all the endpoints can be recorded, and the intermediate paths or stories can be mined by using a searching approach. So usually the stories are the shortest path from one endpoint to another. Fig 6 is an overview of how queries can lead to stories.
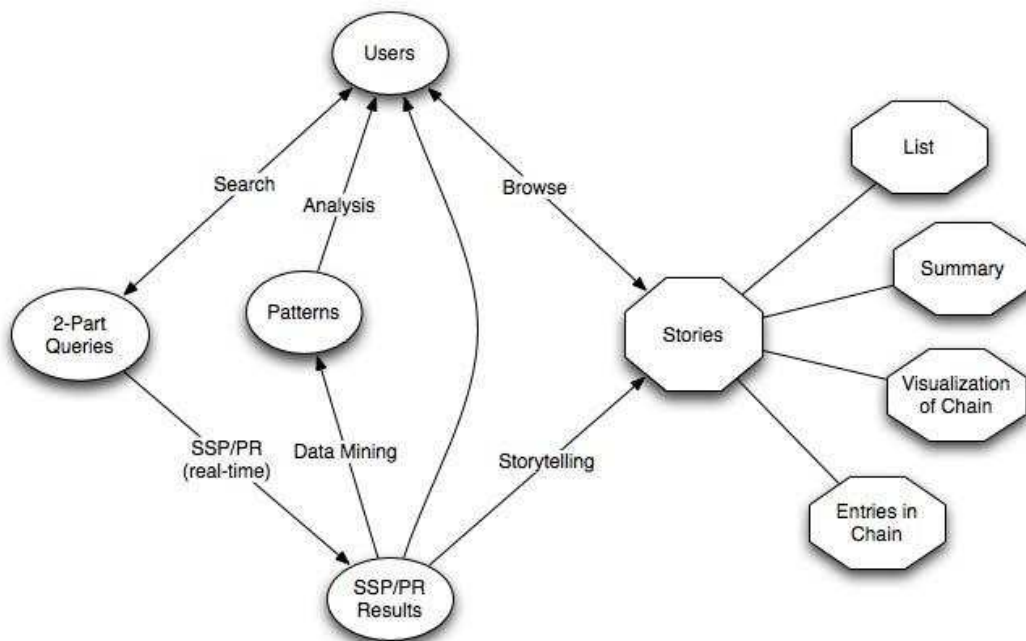
Figure 6. Exploring with SSP, Pathrank, and Storytelling

Our proposed research will integrate these methods using the process for submitting queries, then applying SSP, PR, and Storytelling to connect concepts or events. The user may submit two endpoint queries, such as two tragic events, or two names of individuals, to the system directly. SSP/PR will

process these queries online, and return the pathways to connect these two queries. These pathways will be returned to the users immediately, and at the same time, the system will record the users' queries, and use the Storytelling algorithm later to search for the best stories connecting those user queries. The stories can be displayed in a webpage as the mining result, or some other visualization technologies can be used. By using summarization techniques developed in the information extraction area, the stories also can be filtered and pruned further.

Fig. 7 shows results we produced by hand that summarize what would happen when a user seeks to see how the Virginia Tech shooting is related to a number of events. Thus, the storytelling capabilities of the CTR-toolkit will allow users to explore the ever-expanding CTRnet collection.
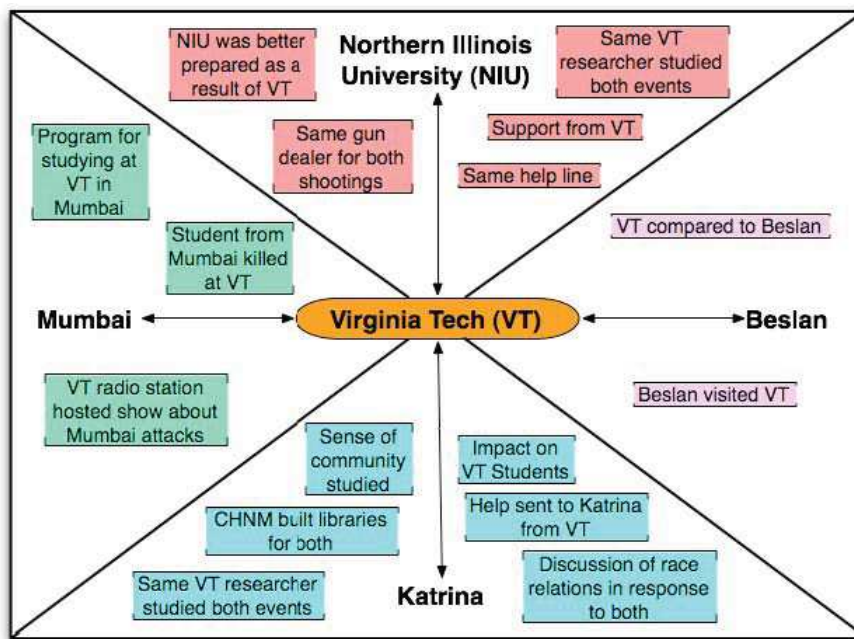

Figure 7: Event connection to VT ([5, 16, 19, 21, 46, 52, 72, 73, 107, 115, 119, 121, 123, 132, 135])

4.5.2   Ontology Building for CTR Domain

We will seed the ontology building process with data gathered from our initial focus group analyses, as well as knowledge gained from literature in this domain, such as, proceedings of the International Conference on Information Systems for Crisis Response and Management or ISCRAM [17, 70] and work on Crisis Informatics [69, 80, 81]. Table 4 shows part of a pruned list based on results from the ISCRAM proceedings (where there is limited coverage of long term community recovery issues). Thus, in a brief pilot exploration related to this proposal, we explored a simple technique for finding CTR related terms

Table 4. Selected distinctive word pairs identified from ISCRAM proceedings			
emergency response	decision support	information systems	teams participants
decision making	data models	disaster monitoring	teams maps
command teams	disaster plan	crisis management	sms text-message
flood alerts	information seeking	situational awareness	disaster registry
physical communication	human disaster	teams access	decision preference

and phrases, rather than use the more sophisticated methods we have deployed for summarizing electronic theses and dissertations [104]. We began by taking the abstracts and keywords from papers in the 2005 and 2007 ISCRAM proceedings. After removing stop words, we used the N-Gram Statistics Package [1, 83] with the t-score measure, to get the top keywords. The result was a list with approximately 51%

accuracy.

We conducted focus group interviews following the April 16 tragedy at Virginia Tech with researchers from Psychology, Sociology, and Computer Science. Each focus group employed the Group Cognitive Mapping System (developed at Virginia Tech) in the Pamplin College of Business computing lab to complete two activities [122]. In the first activity participants used electronic brainstorming to generate ideas, issues, and concepts in response to the following framing statement:

Thinking of your research interests and of the interests of other researchers in your discipline:
1) What are the most important research questions, issues, and/or concepts that are relevant to the events of April 16th? 2) What data are needed to address these research questions and issues?
3) What are ideal visualizations that would be most useful for understanding the key issues in your discipline?

Each group generated a substantive number of concepts. The next activity was for participants to review the concepts generated and suggest names of categories that would contain a set of similar concepts. Participants were asked to suggest category names using a round-robin approach. When the group was satisfied that all the relevant categories were identified, participants were asked to provide a definition or description of the concepts that fit into that category. The resulting categories appear in Figure 8.
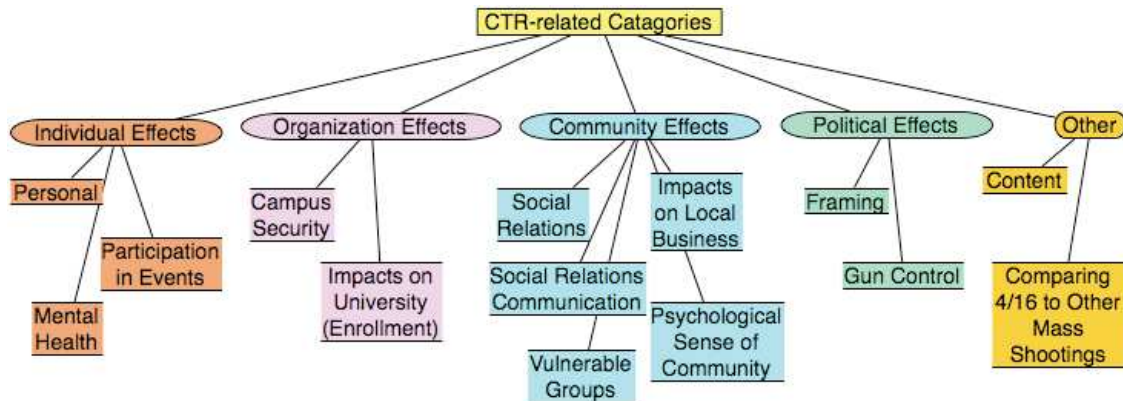

Figure 8: CTR Categories

The categories identified form the basis for defining an ontology of issues and concepts that the participants believe could be investigated in the context of this and other tragedies. A comprehensive ontology would be of great value, especially after being connected into a knowledge base that relates to events, studies, findings, and results / lessons learned / best practices. This would include: variables to measure, activities and behaviors to monitor, effects to test, and connections with the theories and knowledge of a variety of social and behavioral science disciplines. It could help with framing research for the future and our interest in engaging experts in developing a research agenda for such tragedies.

The ontology will be further refined and expanded based on user interactions and log data gathered from the CTR website and Facebook application. We will make use of social and automatic tagging to refine and expand it. The CTR domain-specific ontology will be used to enhance various user services such as searching, browsing, recommending, summarization [104, 105], and visualization.

5   Assessment, Evaluation and Testbeds

CTRnet consists of a variety of components and interfaces. Each type of component will have a different validation and evaluation plan. However, no single evaluation design can address all the evaluation questions. We will use a mix of qualitative and quantitative methods to address the evaluation questions. We list the main evaluation questions and their corresponding primary evaluation data collection methods in Table 5. For example, the services will have two-way validation and evaluation. We will evaluate the interfaces through online user questionnaires. The other evaluation will be through standard information retrieval evaluation techniques such as using TREC datasets (see, for example [47-51, 53]). We will

design and develop online surveys linked to the digital library collections we develop in order to ask users about their perceptions and experience in using information technology and the CTR toolkit. As such, our respondent population will be opportunity samples based on usage of the CTR toolkit and digital library related to a given event (e.g., community leaders, general public, interested outsiders). We will analyze the completed questionnaires using SPPS software. We will analyze the content of the Web 2.0 sources (e.g., Facebook, Twitter, blogs, Flickr) to identify major themes and patterns of crisis information fathering and sharing and the recovery process. We will follow Krippendorf's content analysis methodology [68] and use NVIVO software to analyze the social software entries (e.g., discussions, photos, blogs, Twitter posts).

Table 5. Evaluation Methods					
Questions	Heuristic Evaluation	Usability Testing	Expert Reviews	Log Analysis	Online Surveys
DL Expert Perspectives			X		
User Perspectives		X		X	X
Services & Enhancements	X	X	X	X	X
GUI Design	X		X		

Most of the evaluation questions and methods of the CTR toolkit and digital library will follow those of standard digital libraries [38, 62, 106]. That is, the data collection and coverage rate will be evaluated using standard information retrieval methods, such as precision and recall. The service and enhancement assessments will focus analysis on the use and usability of functions and services in the CTR digital library system. Further, we will address overarching questions such as: How effectively does the CTR project carry out its mission and serve the CTR community? How do the different parts of CTR function during different stages of CTR events? We will conduct these assessments by popular digital library evaluation methods, like usability testing and online user questionnaires. We will collect data for the evaluation from the typical users of our CTR system, such as the community leaders, victims, families and friends, and the general public. We will collect data for the evaluation from the perspective of the DL experts and researchers using primarily expert reviews. Researchers in psychology will provide specific research questions, such as how information technology and the CTR toolkit and digital library collection affected crisis management and recovery. Researchers in computer science will evaluate the use and impact of these technologies in the CTR community. Sociologists will analyze the social impact of the CTR resources on community wide information sharing and the crisis recovery process.

6   Prior NSF Work

This project builds on a current SGER grant related to April 16: NSF IIS-0736055: SGER: DL-VT416: A Digital Library Testbed for Research Related to 4/16/2007 at Virginia Tech, $199,993 + $12,000 REU supplement (with 3 students), PI Edward A. Fox, Co-PIs: Christopher L. North, Donald J. Shoemaker, Naren Ramakrishnan, Weiguo Fan, August 15, 2007 - January 31, 2009. Our main publication was a journal paper in Traumatology [30]. Our web site has been popular; see www.dl-vt-416.org. Another grant related to the tragedy is: NSF Small Grant for Exploratory Research (NSF-0738390) Capturing Ephemeral Data on the Virginia Tech Tragedy, A. Kavanaugh (PI), F. Quek, S. Sheetz, 7/1/07-8/31/09. We are investigating social network use of cell phones on the day of the shootings at Virginia Tech campus, using survey questionnaires and structured interviews with students, faculty and staff. Support has been provided for two PhD students and one undergraduate. Papers under review or in process include [60, 109].

    Fox served as PI at Virginia Tech on a subcontract from NSF (ITR) funding, through grant IIS-0325579, entitled Information Technology Research: Managing complex information applications: An archaeology digital library. This was launched by an archaeologist, Project PI James Flanagan (CWRU), with the IT aspects led by Fox and co-PI Fan. VT's subcontract was for $189,500, covering 9/1/03-

12/31/05, but a no-cost extension allowed continuation of research into the summer of 2007. The ETANA-DL (digital library – see http://www.etana.org for a link to the system) provides an integration framework and broad set of services operating on data from 15 sites in Jordan and Israel. Two dissertations, three theses, and a number of publications [20, 42, 90, 91, 95, 98, 99, 101, 110-113, 131] have demonstrated how information technology can be used effectively for the timely digitization and dissemination of archaeological [101] findings for the benefit of society.

Kavanaugh served as PI on NSF Digital Government Award NSF-0429274, Modeling Online Participation in Local Governance, with co-PIs M.A. Pérez-Quiñones, P. Isenhour, and D. Dunlap, 9/1/2004-8/31/2007. The project supported one post-doc, five graduate students and six undergraduate students on design and modification of innovative tools to support and foster online citizen discussion 'in the wild.' Published papers include [33, 55-59, 61, 63, 64, 120].

7   Project Personnel

This project will be guided by an advisory board, chaired by Professor Jerzy Nowak, whose wife was killed on April 16, 2007 (see letter). He is director of Virginia Tech's Center for Peace Studies and Violence Prevention. External members of the Advisory Board include Kristine Hanna (Internet Archive – see letter), Starr Roxanne Hiltz and Murray Turroff (NJIT), Susan Metros (USC – see letter), Ben Shneiderman (U. Md. – see www.cs.umd.edu/hcil/911gov/), Padmini Srinivasan (U. Iowa – see letter), and Eric Van de Velde (Caltech – see letter). Other local members of the board include Erv Blythe (VP for Information Technology), James Hawdon (Sociology – see letter), Russell Jones (Psychology), Timothy Luke (Political Science; Director, Center for Digital Discourse and Culture), Gail McMillan (Director, Digital Library and Archives, University Libraries), Christopher North (Computer Science), and John Ryan (Chair, Sociology – see letter). The external board will meet once yearly in person, perhaps at ISCRAM, as well as twice electronically. The local board will meet at least twice per year, in person. We will seek guidance and support more broadly too, and engagement of many locations in CTRnet.

Prof. Edward A. Fox has been at Virginia Polytechnic Institute and State University (VPI&SU or Virginia Tech) since 1983, where he serves as Professor of Computer Science and directs the Digital Library Research Laboratory. He has been PI and co-PI on over 100 research and development projects, leading to hundreds of publications and of presentations. He served from 1987-95 as vice chair and then chair of the Special Interest Group on Information Retrieval. He was Chairman of the IEEE-CS Technical Committee on Digital Libraries. Dr. Fox will serve as PI, providing managerial and technical leadership. He will liaise with the Internet Archive, and serve as point of contact with new CTRnet sites.

Donald J. Shoemaker, Professor of Sociology, has been at Virginia Tech since 1974. He has been involved in 37 funded research projects. In addition Professor Shoemaker has published extensively in the field of crime and delinquency, including a textbook on juvenile delinquency, published in September, 2008. Professor Shoemaker served on the Virginia Attorney General's Anti-Gang Task Force from 2003-2006 and is a member of the Oak Ridge Associated Universities National Security Expert Team, an advisory and consultant group which works with the Department of Homeland Security and other government agencies. He will coordinate liaison activities with other social and behavioral scientists.

Dr. Andrea Kavanaugh is a Senior Research Scientist in the Computer Science Department at Virginia Tech (VT). She is the former Director of Research for the community computer network known as the Blacksburg Electronic Village (BEV) under the direction of VT Information Systems. She also serves as the Associate Director of the Center for Human Computer Interaction. A social scientist, Dr. Kavanaugh has been employing quantitative and qualitative research methods and leading sponsored research projects since 1980 to evaluate the diffusion, adoption, use, and impact of information and communication technology. She leads a SGER grant on the social network use of cell phone technology during the tragic events of April 16 at VT. She will coordinate project evaluation activities.

Naren Ramakrishnan is a Professor and Associate head for graduate studies in the Department of Computer Science at Virginia Tech. He also serves as an adjunct professor at the Institute of Bioinformatics and Applied Biotechnology (IBAB), Bangalore, India. His research interests include

problem solving environments, mining scientific data, and information personalization. He is an area editor for IEEE Computer, an editorial board member of Knowledge and Information Systems and the Journal of Intelligent Information Systems. He is a general chair for the Eighth IEEE International Conference on Data Mining (ICDM'08), Pisa, Italy, Dec. 2008. In 2007, Computerworld named him to their list of '40 under 40' innovative IT people to watch. He will coordinate data mining and activities related to Storytelling.

Steven Sheetz is an Associate Professor in Accounting and Information Systems and Director of the Center for Global E-Commerce in the Pamplin College of Business. His research focuses on object-oriented software engineering, software measurement, and the adoption of information systems standards, e.g., XBRL (Extensible Business Rule Language). He has extensive industry experience with database management systems and software development. In response to the events of April 16th he led a study of the impact of technology on stress, collaborated closely with DL416 team, and was part of the SGER grant on cell phone use. He will design and direct the development of the software in the CTR-toolkit.

8 Summary

We will develop CTRnet, to serve those involved in or interested in crisis, tragedy, and recovery. This will be enabled by research in many aspects of intelligent information integration. Our approach will be to build upon our work on distributed digital libraries, information retrieval, HCI, data and text mining, database management, sociology, and other skills brought to bear by the co-PIs and our partners.

During the three years of the project, CTRnet will be gradually built and extended and evaluated, as new events occur and new partners are found. In the first year, the focus will be on constructing a first version of the CTR toolkit, including the Facebook application, and on testing it whenever an opportunity arises. Since VT already has received some data from the Texas A&M bonfire disaster, and the NIU shooting, preliminary testing will occur, with that data being integrated with VT information. When available, data recently promised from Beslan, and information from Dr. Srinivasan (see letter) related to the flood at Iowa, will be integrated. These efforts will proceed into year 2, when a second GRA is hired, who will concentrate on connecting SSP, PR, and Storytelling. Year 3 will focus on elaboration and broad dissemination of the CTR toolkit, expansion of CTRnet, and engaging a wide set of stakeholders in using the emerging and expanding cyberinfrastructure for CTR.

The **intellectual merit** of our research includes advancing digital library methods through intelligent information integration, building upon the 5S framework, leveraging Web 2.0 applications, providing powerful aids for exploration (for the first time integrating Stepping Stones and Pathways, PathRank, and Storytelling), and developing a CTR domain to benefit and facilitate healing for individuals and communities involved in tragic events. The use of ontologies, social networks, and advanced retrieval techniques will allow users to identify relationships among information items, and will enable communities to obtain necessary information for research and knowledge building.

The **broader impact** of our research is aimed towards aiding individuals and communities involved in crises and tragedies, especially with regard to long-term recovery. We do that with a new approach, which should work in many other domains as well: building (atop solid theoretical foundations) a domain specific digital library toolkit, so a distributed network can support a growing worldwide community, along with social and behavioral science researchers, learners, decision makers, law enforcement agencies, emergency personnel, and the public.

PS On 12/15/08, Fox was interviewed by a reporter of The Chronicle of Higher Education about the recent panel at the CNI Fall Meeting, on "Capturing Crisis"; Fox, along with two members of the Advisory Board (Hanna, Srinivasan), met with strong interest as they outlined key ideas of this proposal. On 12/16/08, an announcement appeared of Lucinda Roy's book "No Right to Remain Silent: The Tragedy at Virginia Tech"; she taught Seung-Hui Cho and alerted university officials about his troubled actions and writings, and addresses the violence affecting schools and lives.

References

1. Banerjee, S. and Pedersen, T., The Design, Implementation and Use of the Ngram Statistics Package. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics (Mexico City)*, (2003)

2. Bisson, G., Nedellec, C. and Canamero, L., Designing clustering methods for ontology building - The Mo'K workbench. In *Proc. of the ECAI Ontology Learning Workshop*, (2000)

3. Boden, M.A. Evolutionary Epistemology. *Synthese*, *85*. 185-197. 1990.

4. Brown, A. *Archiving Websites: A practical guide for information management professionals*. Facet Publishing, London, 2006.

5. Burton, F. The Virginia Tech Shootings: A Case for Redundant Communications, April 17th, 2007. http://www.april16archive.org/items/show/1397

6. Cai, D., Rijsbergen, C.J.V. and Jose, J.M., Automatic Query Expansion based on Divergence. In *10th International Conference on Information and Knowledge Management (CIKM'01)*, (New York, 2001), 419--426.

7. Calado, P.P., Goncalves, M.A., Fox, E.A., Ribeiro-Neto, B., Laender, A.H.F., Silva, A.S.d., Reis, D.C., Roberto, P.A., Vieira, M.V. and Lage, J.P. The Web-DL Environment for Building Digital Libraries from the Web. In *Proc. JCDL'2003, Third ACM / IEEE-CS Joint Conference on Digital Libraries, May 27-31, Houston, TX*, IEEE, 2003, 346 - 357.

8. Campos, L.M.d., Fernandez-Luna, J.M. and Huete, J.F., Query Expansion in Information Retrieval Systems Using a Bayesian Network-Based Thesaurus. In *14th Annual Conference on Uncertainty in Artificial Intelligence (UAI--98)*, (San Francisco, CA, 1998), 53-60.

9. Can, F., Fox, E.A., Snavely, C. and France, R.K. Incremental Clustering for Very Large Document Databases: Initial MARIAN Experience. *Information Systems*, *84*. 101-114. 1995.

10. Chen, Y. *A Novel Hybrid Focused Crawling Algorithm to Build Domain-Specific Collections*. Thesis. Virginia Tech, Computer Science, Blacksburg, 2007. http://scholar.lib.vt.edu/theses/available/etd-02162007-005107/

11. Cimiano, P. and Völker, J., Text2Onto: A Framework for Ontology Learning and Data-driven Change Discovery. In *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems*, (2005)

12. Curbera, F. and et_al. Unraveling the Web Services Web: An Introduction to SOAP, WSDL, and UDDI. *IEEE Distributed Systems Online*, *3* (4). 2002. http://dsonline.computer.org/0204/features/wp2spot.htm

13. Das-Neves, F., Fox, E.A. and Yu, X., Connecting topics in document collections with stepping stones and pathways. In *Proceedings of the 14th Conference on Information and Knowledge Management, CIKM 2005, 31st October - 5th November, Bremen, Germany*, (2005), ACM, New York, NY, USA, 91-98.

14. Das Neves, F. and Fox, E.A. Extending retrieval with stepping stones and pathways (web site for funded NSF proposal; see also first year report of Aug. 18), 2003. http://fox.cs.vt.edu/SSP/

15. Das Neves, F.A. *Stepping Stones and Pathways:Improving Retrieval by Chains of Relationships between Documents*. Thesis. Virginia Tech, Computer Science,

Blacksburg, VA, 2004. http://scholar.lib.vt.edu/theses/available/etd-11012004-003013/restricted/dissertation.PDF

16. Esposito, G. Shared struggles: Russian representatives visit Virginia Tech after 2004 Beslan school siege, November 17, 2008. http://www.roanoke.com/news/nrv/vtnews/wb/184533

17. Fahland, D., Gläßer, T.M., Quilitz, B., Weißleder, S. and Leser, U., HUODINI - Flexible Information Integration for Disaster Management. In *4th International Conference on Information Systems for Crisis Response and Management (ISCRAM), Delft, NL*, (2007)

18. Fan, W., Gordon, M.D., Pathak, P., Xi, W. and Fox, E.A. Ranking function optimization for effective web search by genetic programming: an empirical study. In *Proceedings of 37th Hawaii International Conference on System Sciences (HICSS), 5-8 Jan. 2004, Big Island, Hawaii*, IEEE, 2004, 105 - 112.

19. Fischman, J. Surprising Patterns of Coping After Virginia Tech and Other Calamities, February 14, 2008. http://chronicle.com/blogs/conference/1670/surprising-patterns-of-coping-after-virginia-tech-and-other-calamities

20. Flanagan, J.W., Fox, E.A. and Fan, W. Managing complex information applications: An archaeology digital library, ETANA-DL homepage, 2003. http://digbase.etana.org

21. Fleming, C. Committee assists Tech with April 16 research, March 11, 2008. http://www.collegiatetimes.com/stories/2008/03/11/committee_assists_tech_with_april_16_research

22. Fortuna, B., Mladenic, D. and Grobelink, M., Semi-Automatic Construction of Topic Ontology. In *Conference on Data Mining and Data Warehouses (SiKDD)*, (2005)

23. Fox, E. Lexical Relations: Enhancing Effectiveness of Information Retrieval Systems. *ACM SIGIR Forum*, *15* (3). 5-36. 1980.

24. Fox, E., Goncalves, M.A., Luo, M., Chen, Y., Krowne, A., Zhang, B., McDevitt, K., Perez-Quinones, M. and Cassel, L. Harvesting: Broadening the Field of Distributed Information Retrieval. In *SIGIR 2003*, 2003.

25. Fox, E.A. Some Considerations for Implementing the SMART Information Retrieval System under UNIX, Cornell Univ. Dept. of Comp. Science, Ithaca, NY, 1983.

26. Fox, E.A. The 5S Framework for Digital Libraries and Two Case Studies: NDLTD and CSTC. In *Proceedings NIT'99, Taipei, Taiwan, Aug. 18-20*, 1999. http://www.ndltd.org/talks/nit99.ppt, http://www.ndltd.org/pubs/nit99fox.doc

27. Fox, E.A. From Theory to Practice in Digital Libraries: 5S and Educational Applications (NDLTD, CSTC). In *NSF-CONACyT-ISTEC Workshop on Digital Libraries, Albuquerque, July 7-9*, 1999. http://www.ndltd.org/talks/nm990708.ppt

28. Fox, E.A., Das Neves, F., Yu, X.Y., Shen, R., Kim, S. and Fan, W. Exploring the computing literature with visualization and stepping stones & pathways. *Communications of the ACM*, *49* (4). 52-58. 2006. http://doi.acm.org/10.1145/1121949.1121982

29. Fox, E.A. Ensemble portal for NSF's NSDL computing pathway, Virginia Tech, 2008. http://www.computingportal.org

30. Fox, E.A., Andrews, C., Fan, W., Jiao, J., Kassahun, A., Lu, S.-C., Ma, Y., North, C., Ramakrishnan, N., Scarpa, A., Friedman, B.H., Sheetz, S.D., Shoemaker, D., Srinivasan, V., Yang, S. and Boutwell, L. A Digital Library for Recovery, Research, and Learning from April 16, 2007 at Virginia Tech *Traumatology*, *14* (1). 64-84. 2008.

31. Fox, E.A. and Murthy, U. Digital Library Research Laboratory (home page), DLRL staff, Blacksburg, VA, 2008. http://www.dlib.vt.edu

32.   Fox, E.A., Chen, Y., Andrews, C., North, C., Sheetz, S., Shoemaker, D., Jiao, J., Ewers, J., Hauptman, R., Wooldridge, K. and Srinivasan, V. The 4/16 Digital Library, Virginia Tech, Blacksburg, VA, 2009. http://www.dl-vt-416.org/

33.   Godara, A.K.J. and Isenhour, P., The Efficacy of Knowledge Sharing in Centralized and Self-Organizing Online Communities: Weblog Networks Vs. Discussion Forums. In *Hawaii International Conference on System Sciences, (Waikoloa, Hawaii)*, (2009), IEEE Computer Society

34.   Gonçalves, M., Fox, E. and Watson, L. Towards a digital library theory: a formal digital library ontology. *International Journal on Digital Libraries*, *8* (2). 91-114. 2005.

35.   Goncalves, M.A., Panchanathan, G., Ravindranathan, U., Krowne, A., Fox, E.A., Jagodzinski, F. and Cassel, L. The XML Log Standard for Digital Libraries: Analysis, Evolution, and Deployment. In *Proc. JCDL'2003, Third ACM / IEEE-CS Joint Conference on Digital Libraries, May 27-31, Houston, TX*, IEEE Computer Society, 2003, 312-314.

36.   Goncalves, M.A. *Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Digital Library Framework and Its Applications*. Thesis. Virginia Tech, Computer Science, Blacksburg, VA, 2004. http://scholar.lib.vt.edu/theses/available/etd-12052004-135923/

37.   Goncalves, M.A., Fox, E.A., Watson, L.T. and Kipp, N.A. Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. *ACM Transactions on Information Systems*, *22* (2). 270-312. 2004. http://doi.acm.org/10.1145/984321.984325

38.   Goncalves, M.A., Fox, E.A., Zhang, B. and Watson, L.T., Towards a Quality Model for Digital Libraries. In *Proceedings of the DELOS Workshop on the Evaluation of Digital Libraries*, (University of Padua, Italy, 2004), 39-40.

39.   Goncalves, M.A., Moreira, B.L., Fox, E.A. and Watson, L.T. "What is a good digital library?" - A quality model for digital libraries. *Inf. Process. Management*, *43* (5). 1416-1437. 2007. http://dx.doi.org/10.1016/j.ipm.2006.11.010

40.   Goncalves, M.A., Fox, E.A. and Watson, L.T. Towards a Digital Library Theory: A Formal Digital Library Ontology. *Int. J. Digital Libraries*, *8* (2). 91-114. 2008. DOI: 10.1007/s00799-008-0033-1

41.   Google. Google Maps, Google, Mountain View, VA, 2007. http://maps.google.com/

42.   Gorton, D., Shen, R., Vemuri, N.S., Fan, W. and Fox, E.A. ETANA-GIS: GIS for archaeological digital libraries. In *Proc. 6th ACM/IEEE Joint Conference on Digital Libraries (JCDL), Jun 11-15, 2006, Chapel Hill, NC*, 2006, 379-379.

43.   Gorton, D.C. *Practical Digital Library Generation into DSpace with the 5S Framework*. Thesis. Virginia Tech, Computer Science Master's Thesis, Blacksburg, 2007. http://scholar.lib.vt.edu/theses/available/etd-04252007-161736/

44.   Gresock, J. *Finding Combinatorial Connections between Concepts in the Biomedical Literature*. Thesis. Virginia Tech, Computer Science, 2007.

45.   Gresock, J., Kumar, D., Helm, R., Potts, M. and Ramakrishnan, N. Mining Novellas from PubMed Abstracts using a Storytelling Algorithm, Virginia Tech, 2007.

46.   Haegele, K. The April 16 Archive tells a new kind of Virginia Tech story, May 6, 2007. http://www.thelalatheory.com/dig7.html

47.   Harman, D.K. *The First Text REtrieval Conference (TREC-1), NIST Special Publication 500-207*. National Inst. of Standards and Technology (NIST), Gaithersburg, MD, 1992.

48. Harman, D.K. *The Second Text REtrieval Conference (TREC-2), NIST Special Publication 500-215*. National Inst. of Standards and Technology (NIST), Gaithersburg, MD, 1993.

49. Harman, D.K. *The Third Text REtrieval Conference (TREC-3), NIST Special Publication 500-225*. National Inst. of Standards and Technology (NIST), Gaithersburg, MD, 1994.

50. Harman, D.K. *The Fourth Text REtrieval Conference (TREC-4), NIST Special Publication 500-236*. National Inst. of Standards and Technology (NIST), Gaithersburg, MD, 1995.

51. Harman, D.K. *The Fifth Text REtrieval Conference (TREC-5), NIST Special Publication 500-238*. National Inst. of Standards and Technology (NIST), Gaithersburg, MD, 1996.

52. Harris, P. The killer of room 2121, April 22, 2007. http://www.april16archive.org/items/show/1023

53. Hawking, D. and Craswell, N. Overview of the TREC-2001 web track. In Voorhees, E.M. and Harman, D.K. eds. *the Proceedings of the Tenth Text Retrieval Conference*, 2001, 61-67.

54. Heflin, J. and Hendler, J. A Portrait of the Semantic Web in Action. *IEEE Intelligent Systems*, *16* (2). 54-59. 2001. http://www.cs.umd.edu/projects/plus/SHOE/search/

55. Kavanaugh, A., Carroll, J.M., Rosson, M.B., Reese, D. and Zin, T.T. Participating in Civil Society: The case of networked communities. *Interacting with Computers*, *17*. 9-33. 2005a.

56. Kavanaugh, A., Carroll, J.M., Rosson, M.B., Zin, T.T. and Reese, D.D. Community networks: Where offline communities meet online *Journal of Computer-Mediated Communication*, *10* (4). 2005b. http://jcmc.indiana.edu/vol10/issue4/kavanaugh.html

57. Kavanaugh, A., Zin, T., Carroll, J., Schmitz, J., Pérez-Quiñones, M. and Isenhour, P., When opinion leaders blog: new forms of citizen interaction. In *dg.o '06: Proceedings of the 2006 international conference on Digital government research*, (2006a), ACM Press, 79-88. http://portal.acm.org/citation.cfm?id=1146598.1146628

58. Kavanaugh, A., Kim, H.N., Pérez-Quiñones, M. and Isenhour, P., Models of local government blogging: Design trade-offs in civic engagement. In *3rd Communities and Technologies Conference*, (East Lansing, MI, 2007), Springer, 419-438.

59. Kavanaugh, A., Kim, B.J., Pérez-Quiñones, M. and Schmitz, J. Net gains in political participation: Secondary effects of Internet on community. *Information, Communication, and Society*, *11* (7). 933-963. 2008.

60. Kavanaugh, A., Sheetz, S. and Quek, F., The Social Network Use of Cell Phones during the Virginia Tech Tragedy. In *Fourth International Communities and Technologies Conference*, (State College, PA, (under review))

61. Kavanaugh, A.L., Zin, T.T., Rosson, M.B., Carroll, J.M., Schmitz, J. and Kim, B.J. Local Groups Online: Political Learning and Participation *Computer Supported Cooperative Work*, *16*. 375-395. 2006.

62. Kengeri, R., Seals, C.D., Harley, H.D., Reddy, H.P. and Fox, E.A. Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. *International Journal on Digital Libraries*, *2* (2/3). 157-169. 1999. http://link.springer.de/link/service/journals/00799/bibs/9002002/90020157.htm

63. Kim, B.J., Kavanaugh, A. and Hult, K.M. Local Community Groups and Internet Use: Examining Contents of Local Community Groups' Websites *Journal of Technology,*

Knowledge and Society, *2* (7). 207-222. 2007.
http://ijt.cgpublisher.com/product/pub.42/prod.272

64. Kim, H.N., Kavanaugh, A. and Smith-Jackson, T., Implementation of Internet Technology for Local Government Website: Design Guidelines. In *Hawaii International Conference on System Sciences*, (Waikoloa, Hawaii, 2007).

65. Kim, S., Murthy, U., Ahuja, K., Vasile, S. and Fox, E.A. Effectiveness of Implicit Rating Data on Characterizing Users in Complex Information Systems. In *Proc. European Conference on Digital Libraries, ECDL 2005, Vienna, Sept. 18-23*, Springer, 2005.

66. Kim, S., Lele, S., Ramalingam, S. and Fox, E. Visualizing User Communities and Usage Trends of Digital Libraries Based on User Tracking Information. In *Digital Libraries: Achievements, Challenges and Opportunities*, 2006, 111-120. http://dx.doi.org/10.1007/11931584_14

67. Kim, Y.-C., Jung, J.-Y., Cohen, E. and Ball-Rokeach, S. Internet connectedness before and after September 11 2001. *New Media Society*, *6* (5). 611-631. 2004. http://dx.doi.org/10.1177/146144804047083

68. Krippendorff, K. *Content analysis: an introduction to its methodology*. Sage Publications, Thousand Oaks, CA, 2004.

69. Kristensen, M., Kyng, M. and Palen, L., Participatory Design in Emergency Medical Service: Designing for Future Practice. In *Proceedings of the SIGCHI Conference of Human Factors in Computing Systems*, (New York, 2006), 161-170.

70. Kruchten, P., Woo, C., Monu, K. and Sotoodeh, M., A Human-Centered Conceptual Model of Disasters Affecting Critical Infrastructures. In *4th International Conference on Information Systems for Crisis Response and Management (ISCRAM), Delft, Netherlands May 13-16*, (2007).

71. Kumar, D., Ramakrishnan, N., Helm, R. and Potts, M., Algorithms for Storytelling. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2006)

72. Lydersen, K. Illinois College Applied Lessons From Massacre At Virginia Tech. Vargas, T. ed., February 16, 2008. http://www.washingtonpost.com/wp-dyn/content/story/2008/02/15/ST2008021501113.html

73. McCreary, G.Y. Glendon Y. McCreary: Discussing a tragedy, 4/24/07. http://www.april16archive.org/items/show/1200

74. Mitra, M., Singhal, A. and Buckley, C., Improving Automatic Query Expansion. In *21st Annual International ACM SIGIR Conference on Research and Development in Information*, (Melbourne, Australia, 1998), 206--214.

75. Moreira, B.L., Goncalves, M.A., Laender, A.H.F. and Fox, E.A. Evaluating Digital Libraries with 5SQual. In *Research and Advanced Technology for Digital Libraries, ECDL 2007, LNCS 4675*, Springer, Berlin / Heidelberg, 2007, 466-470. http://dx.doi.org/10.1007/978-3-540-74851-9_44

76. Moreira, B.L., Goncalves, M.A., Laender, A.H.F. and Fox, E.A. Automatic Evaluation of Digital Libraries with 5SQual. *Journal of Informetrics*, *3*. in press. 2009.

77. Murthy, U. and Fox, E. The Superimposed Information Project at Virginia Tech (homepage), Virginia Tech, Blacksburg, VA, 2006. http://si.dlib.vt.edu/

78. Murthy, U., Richardson, R., Fox, E.A. and Delcambre, L., Enhancing Concept Mapping Tools Below and Above to Facilitate the Use of Superimposed Information. In *Conference on Concept Mapping*, (San José, Costa Rica, 2006).

79. O'Reilly, T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, O'Reilly Media, Sebastapool, CA, 2005. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

80. Palen, L. and Liu, S., Citizen communications in crisis: anticipating a future of ICT-supported public participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA, April 28 - May 03, 2007). CHI '07*, (2007), ACM Press, New York, NY, USA, 727-736. http://www.cs.colorado.edu/~palen/palen_papers/palen-crisis.pdf

81. Palen, L., Vieweg, S., Sutton, J., Liu, S. and Hughes, A. Crisis Informatics: Studying Crisis in a Networked World. *Third International Conference on e-Social Science*. 2007. http://www.cs.colorado.edu/~palen/palen_papers/palen-crisisinformatics.pdf

82. Payette, S. and Staples, T. The Mellon Fedora Project. *Proc. 6th European Conference on Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science 2458*. 406--421. 2002. http://link.springer.de/link/service/series/0558/bibs/2458/24580406.htm; http://link.springer.de/link/service/series/0558/papers/2458/24580406.pdf

83. Pedersen, T. Ngram Statistics Package (NSP), University of Minnesota, 2007.

84. Pérez-Quiñones, M.A., Fox, E.A., Cassel, L. and Fan, W., Work in Progress: Personalizing a Course Website Using the NSDL. In *36th Annual Frontiers in Education Conference*, (San Diego, CA, 2006).

85. Perugini, S., Goncalves, M.A. and Fox, E.A. Recommender Systems Research: A Connection-Centric Survey. *Journal of Intelligent Information Systems*, *23* (2). 107-143. 2004. http://dx.doi.org/10.1023/B:JIIS.0000039532.05533.99

86. Perugini, S., McDevitt, K., Richardson, R., Perez-Quinones, M., Shen, R., Ramakrishnan, N., Williams, C. and Fox, E.A. Enhancing Usability in CITIDEL: Multimodal, Multilingual, and Interactive Visualization Interfaces. In *Proceedings Fourth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL2004), Tucson, AZ, June 7-11*, ACM, 2004, 315-324.

87. Perugini, S., McDevitt, K., Richardson, R., Pérez-Quiñones, M.A., Shen, R., Ramakrishnan, N., Williams, C. and Fox, E.A., Enhancing Usability in CITIDEL: Multimodal, Multilingual, and Interactive Visualization Interface In *4th ACM/IEEE-CS Joint Conference on Digital Libraries* (Tucson, AZ, 2004), 315-324.

88. Petinot, Y., Giles, C.L., Bhatnagar, V., Teregowda, P.B., Han, H. and Councill, I.G., A service-oriented architecture for digital libraries. In *Proceedings of the Second International Conference on Service Oriented Computing*, (2004), 263-268.

89. Raghavan, A. *Schema Mapper: A Visualization Tool for Incremental Semi-automatic Mapping-based Integration of Heterogeneous Collections into Archaeological Digital Libraries: The ETANA-DL Case Study*. Thesis. Virginia Tech, Computer Science, Blacksburg, VA 24061, 2005. http://scholar.lib.vt.edu/theses/available/etd-05182005-114155/

90. Raghavan, A., Rangarajan, D., Shen, R., Goncalves, M.A., Vemuri, N.S., Fan, W. and Fox, E.A. Schema mapper: A visualization tool for DL integration. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, June 7-11, 2005, Denver, CO*, 2005, 414.

91. Raghavan, A., Vemuri, N.S., Shen, R., Goncalves, M.A., Fan, W. and Fox, E.A. Incremental, Semi-automatic, Mapping-Based Integration of Heterogeneous Collections

into Archaeological Digital Libraries: Megiddo Case Study. In *Research and Advanced Technology for Digital Libraries, Proc. European Conference on Digital Libraries, ECDL 2005, Vienna, Sept. 18-23, LNCS 3652*, Springer, 2005, 139-150.

92.   Ramachandran, R., Movva, S., Graves, S. and Tanner, S., Ontology based Semantic Search Tool for Atmospheric Science. In *22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, (2006)

93.   Ramakrishnan, N. PIPE: Web Personalization By Partial Evaluation. *IEEE Internet Computing*, *4* (6). 21-31. 2000.

94.   Ravindranathan, U. *Prototyping Digital Libraries Handling Heterogeneous Data Sources - An ETANA-DL Case Study*. Thesis. Virginia Tech, Computer Science, Blacksburg, VA, 2004. http://scholar.lib.vt.edu/theses/available/etd-04262004-153555/

95.   Ravindranathan, U., Shen, R., Goncalves, M., Fan, W., Fox, E.A. and Flanagan, J.W. Prototyping digital libraries handling heterogeneous data sources - the ETANA-DL case study. In *The Proceedings of the European Conference on Digital Libraries (ECDL 2004)*, Bath, UK, 2004.

96.   Ravindranathan, U., Shen, R., Goncalves, M., Fan, W., Fox, E.A. and Flanagan, J.W. ETANA-DL: A Digital Library For Integrated Handling Of Heterogeneous Archaeological Data. In *Proceedings Joint Conference on Digital Libraries, JCDL'2004, Tucson, AZ, June 7-11*, 2004, 76-77.

97.   Ravindranathan, U., Shen, R., Goncalves, M., Fan, W., Fox, E.A. and Flanagan, J.W. ETANA-DL: managing complex information applications – an archaeology digital library (Demo). In Chen, H., Christel, M. and Lim, E.P. eds. *Proceedings of the Fourth ACM-IEEE Joint Conference on Digital Libraries - Global Reach and Diverse Impact, JCDL 2004, June 7-11, Tucson, AZ*, Tucson, AZ, 2004, 414. http://feathers.dlib.vt.edu/

98.   Ravindranathan, U., Shen, R., Goncalves, M., Fan, W., Fox, E.A. and Flanagan, J.W. ETANA-DL: managing complex information applications – an archaeology digital library (Demo). In *the Proceedings of 2004 ACM-IEEE Joint Conference on Digital Libraries*, Tucson, AZ, 2004.

99.   Ravindranathan, U., Shen, R., Goncalves, M., Fan, W., Fox, E.A. and Flanagan, J.W. ETANA-DL: a digital library for integrated handling of heterogeneous archaeological data. In *the Proceedings of 2004 ACM-IEEE Joint Conference on Digital Libraries*, Tucson, AZ, 2004.

100.   Ravindranathan, U., Shen, R., Goncalves, M.A., Fan, W., Fox, E.A. and Flanagan, J.W., ETANA-DL: A Digital Library For Integrated Handling Of Heterogeneous Archaeological Data. In *Joint Conference on Digital Libraries (JCDL 2004)*, (Tucson, AZ, 2004), 76-77.

101.   Ravindranathan, U., Shen, R., Goncalves, M.A., Fan, W., Fox, E.A. and Flanagan, J.W. Prototyping Digital Libraries Handling Heterogeneous Data Sources – The ETANA-DL Case Study. In Heery, R. and Lyon, L. eds. *Research and Advanced Technology for Digital Libraries: Proc. 8th European Conference on Digital Libraries, ECDL2004, September 12-17, 2004, U. Bath, UK, Lecture Notes in Computer Science, vol. 3232*, Springer-Verlag GmbH, Berlin, 2004, 186-197. http://springerlink.metapress.com/openurl.asp?genre=article&issn=0302-9743&volume=3232&spage=186

102. Richardson, R. and Fox, E.A. Using Concept Maps as a Cross-Language Resource Discovery Tool. In *Proceedings of the Fifth ACM/IEEE Joint Conference on Digital Libraries (JCDL2005), June 7-11, 2005, Denver*, 2005, 256-257.

103. Richardson, R. and Fox, E.A. Evaluating concept maps as a cross-language knowledge discovery tool for NDLTD. In *Proceedings ETD2005, Sydney, Sept. 28-30*, 2005.

104. Richardson, R., Goertzel, B., Pinto, H. and Fox, E.A., Automatic Creation and Translation of Concept Maps for Computer Science-Related Theses and Dissertations. In *2nd Concept Mapping Conference CMC 2006*, (San José, Costa Rica, 2006), University of Costa Rica, 160-163.

105. Richardson, W.R., Srinivasan, V. and Fox, E.A. Knowledge Discovery in Digital Libraries of Electronic Theses and Dissertations: An NDLTD Case Study. *Int. J. Digital Libraries*, *9* (2). 163-171. 2008. http://www.springerlink.com/content/w3182840w7j17117/

106. Saracevic, T. Digital Library Evaluation: Toward Evolution of Concepts. *Library Trends*, *49* (2). 350--369. 2000.

107. SAULNY, S. Grief and Questions After Deadly Shootings. BAILEY, J. ed., February 16, 2008. http://www.nytimes.com/2008/02/16/us/16shooting.html?_r=2

108. Schneider, S.M. and Foot, K.A. Crisis communication and new media: The Web after September 11th. In Howard, P.N. and Jones, S.J. eds. *Society Online The Internet in Context* Thousand Oaks: Sage, 2004, 137-153.

109. Sheetz, S., Kavanaugh, A., Quek, F. and Kim, B.J., (in process) Expectation of Connectedness: Cell Phone Use during the Virginia Tech Shootings. In *IISCRAM2009 The 6th International Conference on Information Systems for Crisis Response and Management*, (Gothenburg, Sweden, 2009), ACM.

110. Shen, R., Goncalves, M.A., Fan, W. and Fox, E.A. Requirements Gathering and Modeling of Domain-Specific Digital Libraries with the 5S Framework: An Archaeological Case Study with ETANA. In *Proc. European Conference on Digital Libraries, ECDL 2005, Vienna, Sept. 18-23*, Springer, 2005, 1-12. http://dx.doi.org/10.1007/11551362_1

111. Shen, R. *Applying the 5S Framework To Integrating Digital Libraries (Doctoral Dissertation)*. Virginia Tech, Blacksburg, VA, USA, 2006. http://scholar.lib.vt.edu/theses/available/etd-04212006-135018/

112. Shen, R., Vemuri, N.S., Fan, W.G., da S Torres, R. and Fox, E.A. Exploring digital libraries: Integrating browsing, searching, and visualization. In *6th ACM/IEEE Joint Conference on Digital Libraries (JCDL), June 11-15, 2006, Chapel Hill, NC*, 2006, 1-10.

113. Shen, R., Vemuri, N.S., Fan, W. and Fox, E.A. Integration of Complex Archaeology Digital Libraries: An ETANA-DL Experience. *Information Systems*, *33* (7-8). 699-723. 2008. http://dx.doi.org/10.1016/j.is.2008.02.006

114. Smelser, N.J. *Collective Trauma and Collective Identity*. Berkley: University of California Press, 2004.

115. Staff, C. Community relief efforts praiseworthy, September 8, 2005. http://www.collegiatetimes.com/stories/2005/09/08/community_relief_efforts_praiseworthy

116. Staples, T. and Wayland, R. Virginia Dons FEDORA: A Prototype for a Digital Object Repository. *D-Lib Magazine*, *6* (7/8). 2000. http://www.dlib.org/dlib/july00/staples/07staples.html

117.   Staples, T., Wayland, R. and Payette, S. The Fedora Project - An Open-source Digital Object Repository Management System. *D-Lib Magazine*, *9* (4). 2003. http://www.dlib.org/dlib/april03/staples/04staples.html

118.   Suleman, H. and Fox, E.A. The Open Archives Initiative: Realizing Simple and Effective Digital Library Interoperability. *Journal of Library Automation, special issue on "Libraries and Electronic Resources: New Partnerships, New Practices, New Perspectives"*, *35* (1/2). 125-145. 2001. http://www.dlib.vt.edu/projects/OAI/reports/jla_2001_article_oai.pdf

119.   Sutphin, M. Hurricane Katrina impacts students at Virginia Tech, August 31, 2005. http://www.collegiatetimes.com/stories/2005/08/31/hurricane_katrina_impacts_students_at_virginia_tech

120.   Tauro, C., Ahuja, S., Pérez-Quiñones, M.A., Kavanaugh, A. and Isenhour, P., Deliberation in the Wild: A Visualization Tool for Blog Discovery and Citizen-to-Citizen Participation. In *9th Annual International Conference on Digital Government Research (dg.o. 2008) (Montreal, Canada)*, (2008), ACM, 143-152

121.   Tech, V. Ongoing support for Northern Illinois University, Virginia Tech, 15 Feb 2008. http://www.april16archive.org/items/show/1678

122.   Tegarden, D.P. and Sheetz., S.D. Group Cognitive Mapping: A Methodology and System for Capturing and Evaluating Managerial and Organizational Cognition. *Omega*, *31*. 113-125. 2003.

123.   Thompson, G. Counseling Help Line Available for People Coping with the Northern Illinois University... Fri Feb 15, 2008. http://www.reuters.com/article/pressRelease/idUS260648+15-Feb-2008+BW20080215

124.   Torres, R.d.S., Medeiros, C.B., Goncalves, M.A. and Fox, E.A. A Digital Library Framework for Biodiversity Information Systems. *International Journal on Digital Libraries*, *5*. 2005.

125.   Torres, R.S., Falcao, A., Goncalves, M.A., Papa, J.P., Zhang, B., Fan, W. and Fox, E.A. A genetic programming framework for content-based image retrieval. *Pattern Recognition*, *42* (2). 283-292. 2009. http://dx.doi.org/10.1016/j.patcog.2008.04.010

126.   Tungare, M., Yu, X., Teng, G., Perez-Quinones, M.A., Cassel, L., Fan, W. and Fox, E.A., Towards a standardized representation of syllabi to facilitate sharing and personalization of digital library content. In *Proceedings of the 4th International Workshop on Applications of Semantic Web Technologies for E- Learning (SW-EL)*, (Dublin, Ireland, 2006).

127.   Tungare, M., Yu, X., Cameron, W., Teng, G., Perez-Quinones, M.A., Cassel, L., Fan, W. and Fox, E.A., Towards a syllabus repository for computer science courses. In *Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education*, (Covington, Kentucky, USA, 2007), ACM, 55-59. http://doi.acm.org/10.1145/1227310.1227331.

128.   Turoff, M. and Hiltz, S.R. Information Seeking Behavior and Viewpoints of Emergency Preparedness and Management professionals concerned with Health and Medicine, Final Report to the National Library of Medicine (NLM), March 6, 2008, NJIT, 2008. http://web.njit.edu/~turoff/Papers/FinalReportNLMTuroffHiltzMarch19.pdf

129.   Van de Sompel, H. and Lagoze, C. Open Archives Initiative, Cornell University, Ithaca, NY, 2000. http://www.openarchives.org

130. Velardi, P., Navigli, R., Cuchiarelli, A. and Neri, F. Evaluation of ontolearn, a methodology for automatic population of domain ontologies. In Buitelaar, P., Cimiano, P. and Magnini, B. eds. *Ontology Learning from Text: Methods, Evaluation and Applications*, IOS Press, Amsterdam, 2005.

131. Vemuri, N.S., Shen, R., Tupe, S., Fan, W. and Fox, E.A. ETANA-ADD: An interactive tool for integrating archaeological DL collections. In *Proc. 6th ACM/IEEE Joint Conference on Digital Libraries (JCDL), June 11-15, 2006, Chapel Hill, NC*, 2006, 161-162.

132. VirginiaTech. Masters of Information Technology in India, 2005. http://www.international.pamplin.vt.edu/mitIndia/

133. Voorhees, E.M. Query expansion using lexical-semantic relations. In *Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, July 3 - 6, 1994, Dublin, Ireland*, ACM, 1994, 61-69. http://www.acm.org/pubs/citations/proceedings/ir/188490/p61-voorhees/

134. Wang, Y., Völker, J. and Haase, P., Towards Semi-automatic Ontology Building Supported by Large-scale Knowledge Acquisition. In *AAAI Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition*, (2006)

135. WDBJ7.com. UPDATED: Two Nelson County residents confirmed dead in India, November 28, 2008. http://www.wdbj7.com/Global/story.asp?S=9425587

136. Xi, W., Zhang, B., Chen, Z., Lu, Y., Yan, S., Ma, W.-Y. and Fox, E.A. Link Fusion: A Unified Link Analysis Framework for Multi-type Inter-related Data Objects. In *Proc Thirteenth Intl World Wide Web Conf, WWW2004, New York, U.S.A. 19-22 May*, 2004, 319-327.

137. Xi, W. and Fox, E., SimFusion: Measuring Similarity using Unified Relationship Matrix. In *Proceedings of the 28th annual intl ACM SIGIR conf on R&D in information retrieval*, (2005).

138. Xu, J. and Croft, W.B. Query expansion using local and global document analysis. In *Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1996, 4-11.

139. Yu, X., Das-Neves, F. and Fox, E.A. Hard Queries can be Addressed with Query Splitting Plus Stepping Stones and Pathways. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, *28* (4). 29-38. 2005. http://sites.computer.org/debull/A05dec/yu.pdf

140. Yu, X. Mining Evidence for Relationships in Heterogeneous Object Spaces, Dissertation Preliminary Report (advisor E.A. Fox), Virginia Tech, Dept. of Computer Science, Blacksburg, VA, 2007.

141. Yu, X., Tungare, M., Fan, W., Perez-Quinones, M.A., Fox, E.A., Cameron, W., Teng, G. and Cassel, L., Automatic syllabus classification. In *Proceedings of the 2007 joint conference on digital libraries*, (Vancouver, BC, Canada, 2007), ACM, 440-441 http://doi.acm.org/10.1145/1255175.1255265

142. Yu, X., Tungare, M., Fan, W., Yuan, Y., Perez-Quinones, M.A., Fox, E.A., Cameron, W. and Cassel, L., Using automatic metadata extraction to build a structured syllabus repository. In *the Proceedings of the 10th International Conference on Asian Digital Libraries*, (Hanoi, Vietnam, 2007).

143. Yu, X., Tungare, M., Fan, W., Yuan, Y., Perez-Quinones, M., Fox, E.A., Cameron, W. and Cassel, L. Automatic Syllabus Classification using Support Vector Machines. In

Song, M. and Wu, Y.-F. eds. *Handbook of Research on Text and Web Mining Technologies*, Idea Group Inc., 2008.

144. Zhang, B., Gonalves, M., Fan, W., Chen, Y., Fox, E., Calado, P. and Cristo, M., Combining structural and citation-based evidence for text classification. In *CIKM '04: Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, (2004), ACM, 162-163.

145. Zhang, B., Goncalves, M.A., Fan, W., Chen, Y., Fox, E.A., Calado, P. and Cristo, M. Combining Structural and Citation-Based Evidence for Text Classification. In *Proc. of the 13th Conference on Information and Knowledge Management, CIKM 2004, Nov. 8-13, Washington, D.C.*, ACM Press, 2004, 162-163.

146. Zhang, B., Chen, Y., Fan, W., Fox, E.A., Goncalves, M.A., Cristo, M. and Calado, P. Intelligent Fusion of Structural and Citation-Based Evidence for Text Classification (poster). In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2005, Salvador, Brazil*, 2005, 667-668. http://doi.acm.org/10.1145/1076034.1076181

147. Zhang, B., Chen, Y., Fan, W., Fox, E.A., Gonçalves, M.A., Cristo, M. and Calado, P., Intelligent GP fusion from multiple sources for text classification. In *14th ACM International Conference on Information and Knowlege Management*, (2005), ACM Press, 477-484.