The Journal of Systems and Software 83 (2010) 1701–1714

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

Survey of data management and analysis in disaster situations

Vagelis Hristidis*, Shu-Ching Chen, Tao Li, Steven Luis, Yi Deng

School of Computing and Information Sciences, Florida International University, United States

ARTICLE INFO

Article history:
Received 6 December 2008
Received in revised form
13 November 2009
Accepted 6 April 2010
Available online 15 June 2010

Keywords:
Disaster management
Business continuity
Data analysis
Information retrieval

ABSTRACT

The area of disaster management receives increasing attention from multiple disciplines of research. A key role of computer scientists has been in devising ways to manage and analyze the data produced in disaster management situations.

In this paper we make an effort to survey and organize the current knowledge in the management and analysis of data in disaster situations, as well as present the challenges and future research directions. Our findings come as a result of a thorough bibliography survey as well as our hands-on experiences from building a Business Continuity Information Network (BCIN) with the collaboration with the Miami-Dade county emergency management office. We organize our findings across the following Computer Science disciplines: data integration and ingestion, information extraction, information retrieval, information filtering, data mining and decision support. We conclude by presenting specific research directions.

1. Introduction

Disaster management has been attracting a lot of attention by many research communities, including Computer Science, Environmental Sciences, Health Sciences and Business. The authors' team comes from a computer science background, and particularly from the area of data management and analysis. When we started working on the Business Continuity Information Network (http://www.bizrecovery.org/) we realized that it is hard to find information on the existing research works on data management and analysis for disaster management. One of the main reasons is that such works have been published in a wide range of forums, ranging from Computer Science to Digital Government to Health Science journals and conferences. Hence, we decided to take on the task of discovering and organizing all this knowledge in order to reveal the state-of-the-art research. We studied multiple sources, ranging from journal publications to government reports, and met with local disaster management officials, in order to come up with a set of future research directions for Computer Scientists – specifically from the areas of data management and analysis – that will greatly benefit the disaster management community. That is, we make an effort to bridge basic research and practice in this domain.

The Federal Emergency Management Agency (FEMA) classifies disasters as Natural (e.g., earth-quakes, hurricanes, floods, wildlife fires etc.) or Technological (e.g., terrorism, nuclear power plant emergencies, hazardous materials etc.). Other classification schemes exist, but whatever the cause, certain features are desirable for management of almost all disasters (also see http://grants.nih.gov/grants/guide/pa-files/PA-03-178.html):

- Prevention
- Advance warning
- Early detection
- Analysis of the problem, and assessment of scope
- Notification of the public and appropriate authorities
- Mobilization of a response
- Containment of damage
- Relief and medical care for those affected

Further, disaster management can be divided into the following four phases: (a) Preparedness; (b) Mitigation; (c) Response; and (d) Recovery. Mitigation efforts are long-term measures that attempt to prevent hazards from developing into disasters altogether, or to reduce the effects of disasters when they occur.

1.1. Disaster data management and analysis challenges

The importance of timely, accurate and effective use of available information in disaster management scenarios has been extensively discussed in literature. An excellent survey and recommendations on the role of IT in disaster management were recently composed by the National Research Council. Information management and processing in disaster management are particularly challenging due to the unique combination of characteristics (individual characteristics appear in other domains as well) of the

* Corresponding author.
E-mail addresses: vagelis@cis.fiu.edu (V. Hristidis), chens@cis.fiu.edu (S.-C. Chen), taoli@cis.fiu.edu (T. Li), luiss@cis.fiu.edu (S. Luis), deng@cis.fiu.edu (Y. Deng).

0164-1212/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2010.04.065

data of this domain, which include:

- Large number of producers and consumers of information
- Time sensitivity of the exchanged information
- Various levels of trustworthiness of the information sources
- Lack of common terminology
- Combination of static (e.g., maps) and streaming (e.g., damage reports) data
- Heterogeneous formats, ranging from free text to XML and relational tables (multimedia data is out of the scope of this paper)

The data that are available for disaster management consist of the following:

- News articles/announcements: these are mainly text data with additional attributes on time and location, etc.
- Business reports
- Remote sensing data
- Satellite imagery data and other multimedia data like video (outside the scope of this paper)

In addition to the content uniqueness, the data in disaster management are also having different temporal/spatial characteristics and they can be categorized into three different types: Spatial Data; Temporal Data; and Spatio-temporal Data.

1.2. The role of information science disciplines in disaster management

The management of disaster management data involves the integration of multiple heterogeneous sources, data ingestion and fusion. Data may have static (e.g., registered gas stations) or streaming (e.g., reports on open gas stations) nature.

The information analysis of disaster management data involves the application of well-studied information technologies to this unique domain. The data analysis technologies that we will review for disaster-related situations are:

- *Information extraction* (*IE*): disaster management data must be extracted from the heterogeneous sources and stored in a common structured (e.g., relational) format that allows further processing.
- *Information retrieval* (*IR*): users should be able to search and locate disaster-related information relevant to their needs, which are expressed using appropriate queries (e.g., keyword queries).
- *Information filtering* (*IF*): as disaster-related data arrives from the data producers (e.g., media, local government agencies), it should be filtered and directed to the right data consumers (e.g., other agencies, businesses). The goal is to avoid information overload.
- *Data mining* (*DM*): collected current and historic data must be mined to extract interesting patterns and trends. For instance, classify locations as safe/unsafe.
- *Decision support*: analysis of the data assists in decision-making. For instance, suggest an appropriate location as ice distribution center.

Fig. 1 shows how IE, IF, IR, DM technologies fit into the disaster management information flow process. IE, IF, IR and DM have been studied extensively in the past. In this paper, we survey the ways they have been applied to disaster management situations. A challenge of this survey is that to collect this information we had to do a literature survey of multiple research communities, including Computer Science, Business, Healthcare and Public Administration.

To limit the scope of the project we only study textual data, that is, we do not consider multimedia data. Management and analysis of multimedia data generally involves converting to textual knowledge using specialized techniques, and after that this knowledge is processed similarly to the textual data, as explained in this paper.

The rest of the paper is organized as follows. Section 2 discusses some of the intricacies of the workflow in disaster management situations, which we learned from our interaction with government and private agencies during our Business Continuity Information Network (BCIN) project. Sections 3–8 present the disaster management-related challenges and proposed solutions in the areas of data integration, Information Extraction, Information Retrieval, Information Filtering, Data Mining and Decision Support, respectively. Section 9 discusses our experiences in creating a Business Continuity Information Network for disaster management. Section 10 discusses future research directions. Finally, we conclude in Section 11.

2. Disaster management in practice and available tools

2.1. Disaster management workflow

There are different types of communities that require disaster data management tools and systems to report on the current situation, share infrastructure and resource data, and exchange other emergency communications. From the perspective of the emergency management community, organizational actors include the

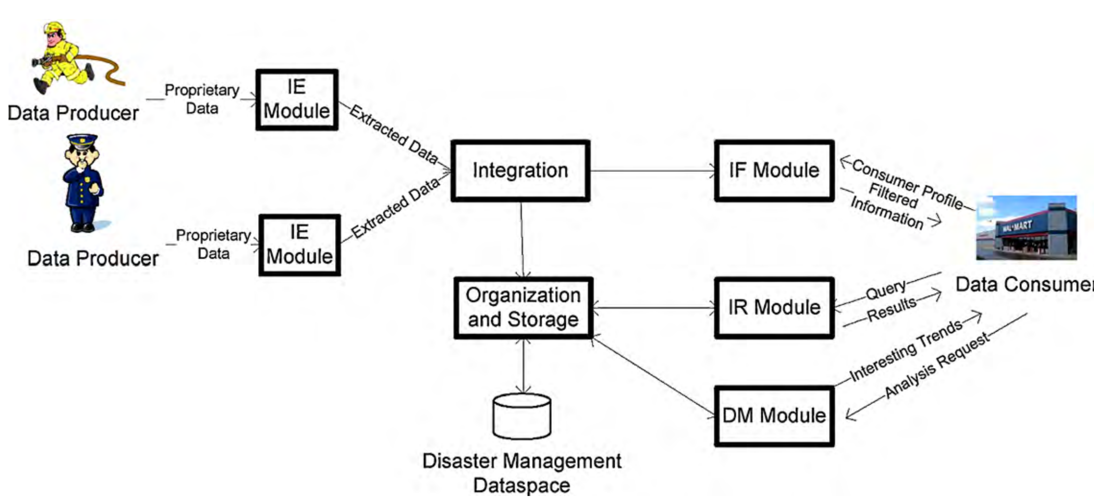

Fig. 1. Data flow in disaster management.

county emergency management offices, partnering agencies like police and fire departments, transportation departments, utility departments, and other critical private infrastructure like state-wide energy companies. Emergency management departments follow incident command systems standards like the National Incident Management System that provides protocols and processes for control, coordination and communication among the above mentioned agencies with jurisdiction that typically respond to a disaster event.

Emergency management departments use software systems to facilitate execution of their incident command systems and interact using intradepartmental, interdepartmental and external organization information sharing tools. The information shared is typically short unstructured messages with occasional tabular data, and is often encoded as PDF or Word documents. For example, a message could be "Points of Distributions (PODS) for ice and water will be open at 7 am tomorrow morning at the following locations throughout the county". The format of these messages is very limited in structure since the incident command system protocols do not try to engineer how to respond to a specific disaster type, but ensure that the participating departments and agencies follow a specific protocol and chain of command.

In a large-scale disaster like a hurricane, a disaster management system records near-real-time situation reports from hundreds of sources. As prescribed by incident command systems, typical reports include an incident action plan and situation report. These reports are generated by the participating agency and county division and provide information concerning the goals of the responding agency during the reporting period, the actions taken to achieve these goals and what was actually achieved. For example, when the county emergency operations center is activated in advance of a hurricane threat, the regional water management agency may report in an incident action plan that it is monitoring precipitation and canal levels to mitigate flood risk while reporting in the situation report what the recorded precipitation and canal water levels are and how that may affect the community. The branch director reviews and summarizes these reports from each agency and produces a summary report that is forwarded to the operations director. The operations director combines each branch reports to produce an overall situation report for the incident commander. Tools are used to record such response and recovery actions to apply and receive federal disaster recovery funds. Other communications mechanisms are used to transmit and document situational status such as e-mails, mailing lists, faxes and even paper.

For the private sector, maintaining a business continuity (or disaster recovery) plan insures the organization is considering the risks that may impede the continuity of operations and defines specific actions to mitigate these risks. Based on the type of disaster event and the amount of effort invested, the business continuity plan provides guidance on what steps the business continuity team will take based on the phase of the disaster event. Business continuity teams are responsible to prepare the appropriate communications and reports on situational status and preparation/response measures prescribed by the business continuity plan to management and employees. For instance, an organization may establish a set of assessment tasks to be conducted prior to 72 h of a hurricane landfall. Employees may be required to update contact information and provide operational status, vendors contracted for special disaster recovery services are placed on stand-by, non-essential equipment is moved out of harm's way.

2.2. Available disaster management tools

Providing a complete record of the event, of all actions, decisions, and resulting situational timeline is still a challenge, whether following an event in progress, or trying to reconstruct event actions after a disaster for evaluation. The approaches and the tools used for information sharing vary based on the mission and scale of the participating agencies. States have developed in-house tools for information sharing, but commercial tools are also used. Emergency Management departments located in large urban areas have access to more resources to acquire commercial software system such as Web EOC and E-Teams. DHS has developed a free Disaster Management Information System that is available to county emergency management offices. These products provide an effective computer based report/document sharing environment for county emergency management and participating agencies. Other leading and notable efforts include: The Florida Disaster Contractors Network is a web accessible database deployed in Florida by FEMA and the State of Florida in the aftermath of Hurricane Wilma to help citizens locate a contractor for home and business repairs; The National Emergency Management Network is a tool to allow local governments share resources and information about disaster needs; The RESCUE Disaster Portal is a web portal for emergency management and disseminating disaster information to the public; Sahana is a disaster and resource management tool for base camp humanitarian aid, communication and coordination.

These useful situation-specific tools store reports, provide query interfaces and GIS capabilities to search for reports and visualize results, and have several canned queries and views that help to simplify the users' interaction. The primary goal of these systems are message routing, resources tracking and document management for the purpose of achieving situational awareness, demonstrating limited capabilities for automated aggregation, data analysis and mining.

Unlike most US counties that are mandated by the federal government to have a county emergency management plan and demonstrate their ability to address disasters, there are very few industries that are regulated to maintain a disaster recovery or more broadly speaking a business continuity plan. Such industries that are regulated are financial institutions such as banks and energy productions facilities like nuclear power plants. Many studies have shown that the vast majority of businesses in the US do not have a business continuity plan (Office Depot Brochure, 2007). As a result of the US 911 Commission, the Department of Homeland Security is developing a Voluntary Private Sector Preparedness Accreditation and Certification Program to help the business community become more prepared for disasters (Voluntary Private Sector Preparedness Accreditation). It is unclear how the business community will receive this program and how the program will motivate the business community to act.

There are quite a large number of tools available to assist the business community plan for a disaster event. These software tools range from simple document templates and online web questionnaires to sophisticated, multi-user workflow and document management systems. The State of Florida provides a free web-based tool at www.floridadisaster.org, which through a questionnaire helps business operators assess their risks, and advises on actions to take to improve their readiness for a disaster event. A similar but more comprehensive tool is provided by the Institute for Business and Home Safety and their Open for Business program, providing educational materials, workbook, and website support. A plethora of stand-alone business continuity programs exist which provide businesses with process support for different continuity planning standards and techniques.

The Business Continuity and Disaster Recovery Associates provide a website (Business Continuity and Disaster Recovery Associates) which provides a listing of business continuity planning and disaster recovery planning software and services. In addition, the site maintains a list of sample business continuity and disaster recovery plans. Websites like these provide a good starting point to developing a business continuity plan. The Strohl Systems Group produces a variety of software systems of value to business continuity planners and continuity team support. Their products range from Incident Commander, incident command system based on WebEOC to Notifind, a system designed to locate your key employees. Such programs provide the continuity manager with a set of tools that can be used to allow a company's key employees to simultaneously report on critical information and actions taken as prescribed by the company's continuity plan. The Disaster-Recovery-Plan Group sells business continuity preparation software and a business continuity toolkit consisting of checklists, questionnaires, and auditing processes.

However these tools function as if the business operates on an island and does not account for broader community interaction with other businesses and county organizations. Further these tools do not allow for the integration of real-time information on the conditions resulting from a disaster and typically focus on distilling processes into action checklists. Decision support is achieved by the effective routing of messages and document sharing in support of continuity planning and execution. These tools do not provide sophisticated IE, IR, IF and DM techniques needed when considering how tens of thousands of may interact with their key employees, each other via the supply chain and with county, state and federal organizations. Of particular interest to the authors is how to provide new types of tools that leverage business continuity plans and preparation software? Requirements for the development of such community tools are presented in later sections of this document.

3. Data integration and ingestion in disaster management

Data integration is the process of combining data residing at different sources and providing the user with a unified view of these data (Lenzerini, 2002). This process emerges in a variety of situations both commercial (when two similar companies need to merge their databases) and scientific (combining research results from different bioinformatics repositories). The problem of data integration becomes more critical as the amount of data that need to be shared increases.

By data ingestion, we mean the process of inserting to a system data, which is coming from multiple heterogeneous sources. Scalability is an important issue in data ingestion, since the flow of information may be very high during the time of a critical event.

3.1. Challenges of data integration and ingestion in disaster management

Disaster data are extremely heterogeneous, both structurally and semantically, which creates a need for data integration and ingestion in order to assist the emergency management officials in rapid disaster recovery whenever disasters occur. Disaster data to be collected and integrated for analysis and management can be (i) information such as incident action plans and situation reports; (ii) damage analysis reports; (iii) geographic data and open/closure status about roadways/highways/bridges and other infrastructure such as fuel, power, transportation, emergency services (fire stations, police stations, etc.), schools, and hospitals; (iv) logistics data about vehicles, delivery times, etc.; (v) communication and message data; (vi) financial data needed to manage the collection and distribution of donations; and (vii) data in blogs (Naumann and Raschid, 2006; Saleem et al., 2008). All these data are in many different formats, have varied characteristics, and are available from different sources. Further, such data is often uncertain. Wachowicz and Hunter (2005) examined the uncertainty in the knowledge discovery process in disaster management and discussed approaches for addressing the uncertainty. The availability of such data poses

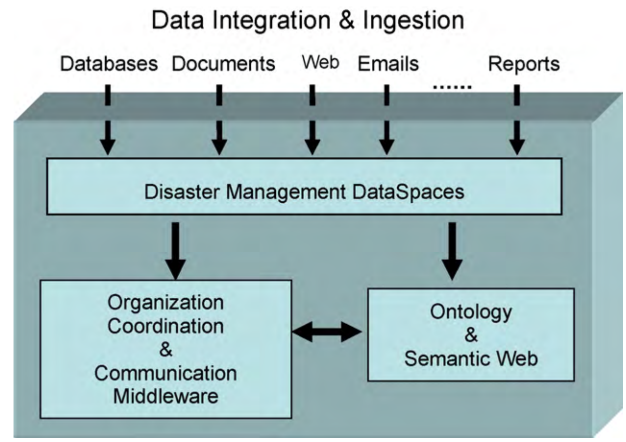

Fig. 2. Key components of data integration and ingestion for disaster data management.

a challenge for the design and development of effective strategies for data acquisition, ingestion, and organization. The key to derive insight and knowledge from the heterogeneous disaster-related data sources is to identify the correlation of data from multiple sources (Foster and Grossman, 2003).

Hence, technologies are needed for data integration and ingestion so that it is possible to query distributed information for disaster management. A disaster data management (DisDM) system was proposed to address the needs and requirements for information integration and information sharing solutions (Naumann and Raschid, 2006). There are two main challenges for the design of DisDM, namely how to enable automation and virtual integration for data management. While there is research work attempting to develop solutions for these challenges, researchers are still seeking an adaptive, flexible, and customizable solution that can be quickly deployed and evolved with the phases of the disaster (Naumann and Raschid, 2006).

3.2. Proposed solutions for data integration and ingestion in disaster management

To be able to handle the heterogeneous disaster data from multiple sources/formats including databases, documents, Web, e-mails, reports, etc., effective and efficient solutions are needed to address the data integration and ingestion issues. The solutions can be categorized into three main components that interact with one another, namely to employ disaster management dataspaces, ontology and semantic Web technologies, and organization coordination and communication middleware. As shown in Fig. 2, the disaster management dataspaces are established to acquire, ingest, organize, and represent the heterogeneous data. Once the dataspaces are established, the ontology and semantic Web component and the organization coordination and communication middleware component interact with each other to overcome the problem of semantic heterogeneity and to ensure the optimal provision of disaster-related information for fast decision-making in a highly coordinated manner.

3.2.1. Employing disaster management dataspaces

Disaster management requires methodologies for data aggregation of the aforementioned disaster-related data. Toward such a demand, the concept of Disaster Management Dataspace is established based on methodologies for disaster-related data acquisition, ingest, organization, and representation (Saleem et al., 2008). A dataspace is a loosely integrated set of data sources, where integration occurs on a need-based manner (Franklin et al., 2005). However, how to handle interrelated static and/or real-time disaster-related data in Disaster Management Dataspace is a challenge. Saleem et al. (2008) propose models for pre-disaster preparation and post-disaster business continuity/rapid recovery, which employ Disaster Management Dataspace as a key component for data aggregation. The proposed model is utilized to design and develop a Web-based prototype system called Business Continuity Information Network (BCIN) system. BCIN system facilitates collaboration among local, state, federal agencies, and the business community for rapid disaster recovery.

3.2.2. Employing ontology and semantic Web technologies

Another effort is to explicate knowledge by means of ontologies to overcome the problem of semantic heterogeneity (Klien et al., 2006). Their approach combines ontology-based metadata with an ontology-based search on geographic information to be able to estimate potential storm damage in forests. As mentioned earlier, the disaster-related data are extremely heterogeneous and different vocabulary could be used in different data sources. The reason why ontologies are employed is that they can be used to identify and associate semantically corresponding concepts in the disaster-related information so that the heterogeneous data can be integrated and ingested. In the context of ontologies, issues arise when integrating the disaster-related data that are described in different ontologies is needed. Since Semantic Web promises a machine-understandable description of the information on the Web, Michalowski et al. (2004) proposed to apply the Semantic Web technology so that a Semantic Web-enabled disaster management system can be developed. Such a system allows efficiently querying distributed information and effectively converting legacy data into more semantic representations. The authors proposed a running application called Building Finder that allows the users to find information about the affected area, including street names, buildings, and residents in the Semantic Web-enabled disaster management systems.

3.2.3. Employing organization coordination and communication middleware

Furthermore, the optimal provision of disaster-related information for fast decision-making in a highly coordinated manner is also critical. In Meissner et al. (2002), the design challenges for an integrated disaster management communication and information system during the response and recovery phases when a disaster strikes are discussed. The authors proposed a multi-level hierarchy that enables both intra and inter organization coordination in their proposed integrated disaster management system. The hierarchy is needed since the system must be able to provide relevant data for a *post-disaster* lessons learned analysis and for training purposes. In addition, the authors proposed the use of XML as a standard data interchange format due to the fact that XML documents are capable of containing all required information, from simple messages to complex maps (Meissner et al., 2002). A similar effort to address the data integration and ingestion issue via communication middleware was presented by Foster and Grossman (2003). They developed a distributed system middleware to allow distributed communities (or virtual organizations) to access and share data, networks, and other resources in a controlled and secure manner for applications such as data replication for business recovery and business continuity.

4. Information extraction in disaster management

Information extraction (IE) is the discipline whose goal is to automatically extract structured information, i.e., categorized and contextually and semantically well-defined data from a certain domain, from unstructured machine-readable documents. Usually, the goal is to populate tables, given a set of text documents.

4.1. Challenges of information extraction in disaster management

Information extraction (IE) is a type of knowledge discovery, whose goal is to automatically extract structured information – typically to be stored in a table – from unstructured documents.

From our experience with interacting with local and federal disaster management agencies, we have found that most of the available information is stored and exchanged in unstructured documents (Adobe pdf or Word files). For instance, each organization, like the Miami-Dade Emergency Office, the Coast Guard, the Fire Department, and so on, publishes a status report every few minutes or hours following a disaster event. Many of these reports have no predefined format, although typically the same organization follows a similar format for all its reports. Hence, it is feasible to automate IE if all the sources (organizations) are known in advance. However, when new sources appear in an ad hoc manner, it is necessary to have a human-in-the-loop to achieve high-quality IE.

As claimed in the excellent survey of Ciravegna (2001), it is widely agreed that the main barrier to the use of IE is the difficulty in adapting IE systems to new scenarios and tasks, as most of the current technology still requires the intervention of IE experts. Porting IE systems means coping with four main tasks: (a) adapting to the new domain information: implementing system resources, such as lexica, knowledge bases; (b) adapting to different sublanguages features: modifying grammars and lexical so to enable the system to cope with specific linguistic constructions that are typical of the application/domain; (c) adapting to different text genres: specific text genres (e.g., medical abstracts, scientific papers, police reports) may have their own lexis, grammar, discourse structure; (d) adapting to different types: Web-based documents can radically differ from newspaper-like texts, because of the existence of hyperlinks, which can be leveraged for ranking (Brin and Page, 1998) or IE: we need to be able to adapt to different situations.

4.2. Proposed solutions for information extraction in disaster management

4.2.1. General IE state-of-the-art

There has been a large corpus of work (Cowie and Lehnert, 1996; Soderland, 2004) on general IE systems. Wrappers are used before the IE patterns are applied on text documents like HTML pages (Kushmerick, 1997; Eikvil, 1999). IE patterns are typically user-specified parsing rules that define how to parse sentences to extract the desired structured (tabular) information. IE has been recently partitioned to Closed and Open IE. Cafarella et al. (2008) provides a recent overview of the IE area.

Closed IE requires the user to define the schema of the extracted tables along with rules to achieve the extraction. The recent work of Jain et al. (2007) shows how IE systems can be used to efficiently answer queries on documents. *Open IE* (Etzioni et al., 2008) generates RDF-like triplets with no input from the user. An RDF triplet is a knowledge representation model, e.g. (*Gustav, is category, 3*). A drawback of Open IE is that it leads to a huge number of triplets, which complicates the execution of structured queries. Another limitation of Open IE is that due to its lack on user input, it is not-focused. In all fully-automated IE systems, there are considerable precision and recall limitations.

Fig. 3 shows a general sequence of steps used in IE. The free text is first split to tokens (e.g., words). Then, these token are converted to entities (e.g., phrases) using a lexicon. Then, these entities are parsed and mapped to the entities of the IE rules (domain patterns). The final output is a set of tuples (records) that are used to populate relational tables.

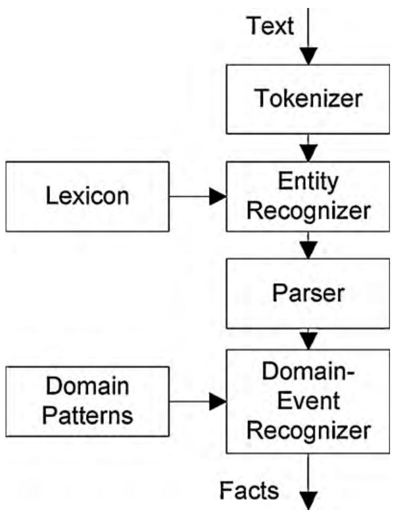

Fig. 3. General IE steps.

Table 1
Facts from Fig. 4.

Disease	Location	Victims
Ebola	Gulu and Masindi	26 infected
	Gulu and Masindi	15 dead
	Gulu and Masindi	7 sick
	Masindi	3 sick
	Gulu	4 sick

4.2.2. *Customization to disaster management scenarios*

Surdeanu and Harabagiu (2002) propose an IE system that can adapt to various domains by improving the handling of coreference of traditional IE systems. For instance, if "Michigan" appears in the same context as "Michigan Corp.", then "Michigan" is considered to be an organization and not a location. Even though this work is not entirely targeted to disaster scenarios, they use disaster management as their case study.

Huttunen et al. (2002) focus on Nature (Infectious Disease Outbreak and Natural Disaster) scenarios and present the challenges in delimiting the scope of a single event for this domain, and organizing the events into templates. First, the events, or components of events, seem to be much more widely spread in text, or *scattered*, in the Nature scenarios than in the Message Understanding Conference (MUC) tasks. Second, they interlock and relate to each other in complex ways, forming *inclusion* relationships.

For instance, the text of Fig. 4 should be translated to the facts in Table 1. However, this list of facts does not reflect the information in the text, since the structure and relationship between these numbers is lost. They tackle this problem by detecting inclusion relationships. For instance in Fig. 4 adding 15, 4, and 7 gives 26, even if the number 26 was not explicitly specified in Fig. 4. The inclusion problem here is to understand that the infected include the dead and the sick. They also create a separate incident for each event.

For the topic of *Terrorist Attacks*, investigated under MUC-3 (1991) and MUC-4 (1992), a system has to find terrorist incidents, and for each incident extract the place and time of the attack, a description of the victims, the responsible party, and the weapon used in the attack.

Grishman et al. (2002) present a case study of IE for infectious disease outbreaks based on the scenario customization algorithm presented in Yangarber and Grishman (1997). Yangarber and Grishman (1997) requires the user to input a few examples of IE, by indicating how some text fragments of interest would be mapped to tabular data. The system then automatically creates matching patterns from these examples.

5. Information retrieval in disaster management

Information retrieval (IR) (Salton and McGill, 1986) is the science of searching for documents relevant to a user query. The query is usually expressed as a set of keywords, and the documents are returned as a ranked list, ordered by decreasing relevance.

In traditional IR, documents are ranked based on a tf·idf ranking function, that is, a function that depends on the term frequency (tf) and the inverse document frequency (idf) of the query terms in the documents. tf denotes the number of occurrences of a term in a document – higher tf leads to higher score – and idf is proportional to the inverse of the number of documents that contain the term – higher idf is better because it means that the term is more infrequent in the data corpus. More recent IR work on searching the Web (PageRank (Brin and Page, 1998)), considers the hyperlinks between the Web pages as another ranking factor, where pages with more incoming hyperlinks generally receive higher score.

5.1. *Challenges of information retrieval in disaster management*

The challenges in IR for disaster management are similar to the ones for IE. In addition to the challenges discussed in Section 4.1, a challenge comes from the way data is stored in a disaster management system. In particular, there is a combination of static (e.g., addresses of gas stations) and streaming (e.g., reports, 911 calls) data. The IR system must be able to seamlessly combine them in order to provide high-quality of results. For instance, the inverse document frequency (idf) must be computed using both static and streaming sources; possibly different weight should be assigned to more recent events. In addition to the quality challenge, achieving efficient (fast response times) query answering is hard. Traditional IR systems rely on a periodically updated inverted index. In contrast, the most recent data must be readily available for querying in disaster management scenarios.

Another challenge is the personalization of the query results given the role or profile of the user. There has been work on personalized Web Search (Liu et al., 2004), which can be the basis of this effort. A key unique opportunity for disaster management personalized IR is the domain knowledge. That is, disaster management ontologies and rules can be employed to guide the IR process. Semantic Web technologies could be of great help here. For instance, a domain ontology may connect the terms "hurricane" and "storm", which will allow appropriately expanding a user query. Farfán et al. (2009) presents a technique to leverage medical ontologies for searching medical records. These ideas could be carried in the disaster domain, if comprehensive ontologies were available.

Moreover, an important parameter for the disaster management domain is time, which should receive special attention. For

A total of **26 health workers** have contracted the virus in *Gulu and Masindi*, according to the Health Ministry records. Of these cases, **15** have died , 4 have been discharged, while **7** are still undergoing treatment, **3** in *Masindi* and **4** in *Gulu*.

Fig. 4. Example of a disease outbreak report (Huttunen et al., 2002).

instance, old information tends to generally be of little importance in disaster management. Finally, novel domain-specific relevance feedback techniques must be studied. Current relevance feedback techniques typically expand the query by adding the most frequent terms from the top result documents. Spatio-temporal factors, which are critical in disaster situations, should be incorporated in the relevance feedback schemes.

5.2. Proposed solutions for information retrieval in disaster management

5.2.1. IR exploiting ontologies in disaster management

The need for disambiguation and standardization of disaster management terminology has been discussed by the early work of (Mondschein, 1994), where the focus is to support environmental emergency management systems. Interestingly, they support that the terminology issue is particularly challenging when the authorities communicate the environmental plans with the citizens, given that technical terms are not understood by citizens.

Klien et al. (2006) present a methodology to use ontologies to improve the quality of IR on Web Services. Their case study involves locating Web Services for estimating storm damage in forests. They use an ontology as follows. First, a shared vocabulary is created for the domain. Then, the Bremen University Semantic Translator for Enhanced Retrieval (BUSTER) (Vögele et al., 2003) is used to build an ontology and use it to expand the query results accordingly.

Clearly, the major obstacle in applying such ontology-assisted IR techniques is the construction of a high-quality domain ontology. Hence, it is imperative that a disaster management ontology is created. Halliday (2007) discusses the current efforts towards this direction. W3C has an active Wiki on this topic, although little material is available at the time this paper is written.

5.2.2. IR on mobile devices for disaster management

Generally, research efforts on Information Retrieval on mobile devices are focusing on the following three aspects:

(i) Utilizing context information where the context information includes location, personal profiles, time of day, schedule and browsing history. In disaster management, since the users are generally “mobile”, the location information is often very important. Gravano et al., 2003 study the problem of location-based search.
(ii) Improving the input. For example, Buyukkokten et al. (2000) improve the keyword input problem for mobile search by providing site-specific keyword completion; and indications of keyword selectivity within sites.
(iii) Improving the search results presentation. The techniques for improving search results presentation can be briefly categorized into two groups (see Chen et al., 2005; Wobbrock et al., 2002; Borning et al., 2000): transforming existing pages and generating new formats that are adapted to mobile display.

Hanna et al. (2003) present a framework to facilitate robust peer-to-peer IR on a distributed environment of mobile devices, with an application to disaster management. The motivating scenario is to allow emergency workers to access large collection of literature and documents (e.g., medical algorithms, maps, chemical hazard sheets, field manuals, area information) on-scene using networked mobile computers.

Arnold et al. (2004) survey the requirements and achievements of IR systems in out-of-hospital disaster response. In addition to the need for robust wireless transfer of the retrieved data, privacy is also critical in this scenario, and hence encryption techniques are necessary. A more recent article (Rosen et al., 2002) also discusses the importance of mobile, robust and secure IR in medical care.

5.2.3. Application of semantic web technologies

In contrast to traditional IR, where a query is evaluated against a collection of documents, the Semantic Web promises to provide higher quality of retrieved information. In particular, the vision is that the query will be expressed in a semantic language, and a software agent will consult ontologies, knowledge bases (e.g., RDF-based) and available Web Services to locate the information that best answers the query.

A key challenge in realizing this vision is converting currently available knowledge into formal semantics (RDF). An example of such a project, which can be applied to disaster management scenarios, is Building Finder (Michalowski et al., 2004), which helps users obtain satellite imagery, street information, and building information about an area. Building Finder combines information from various Web Services like Microsoft’s Web service TerraService (http://terraserver-usa.com). However, note that Building Finder can only accept location-based queries and not keyword-based as is common in IR. Kaempf et al. (2003) is another such project for flood management. They describe a methodology for building a flood-related ontology. Chapman and Ciravegna (2006) also employed semantic web and natural language technologies for emergency response. The First International Workshop on Agent Technology for Disaster Management discusses agent technologies in disaster management. Neches et al. (2002) presents a framework to combine Geo-spatial data and retrieved document collections in order to provide situation awareness in disaster situations. However, this work does not provide any details on the implementation, which means that most probably traditional tf·idf retrieval techniques were used.

5.2.4. Create topic hierarchies

Troy et al. (2008) present their experiences in building a local resource database of suppliers providing physical, information and human resources for use in disaster response. The wide variety of information in the database requires a systematic method for categorization that facilitates effective information searching and retrieval. The system directory provides a hierarchical structure for all data and information as recommended by Turoff et al. (2003). The taxonomy contains more than 8000 terms that are organized into a hierarchical structure that shows the relationships among terms. There are 10 service categories: Basic Needs, Consumer Services, Criminal Justice and Legal Services, Education, Environmental Quality, Health Care, Income Support and Employment, Individual and Family Life, Mental Health Care and Counseling, and Organizational/Community/International Services. Each category is broken down into increasingly specific levels.

6. Information filtering in disaster management

6.1. Overview of IF

According to Wikipedia, an Information filtering system is a system that removes redundant or unwanted information from an information stream using (semi)automated or computerized methods prior to presentation to a human user. Its main goal is the management of the information overload and increment of the semantic signal-to-noise ratio.

Information filtering is a function to select useful or interesting information for the user among a large amount of information. In general, IF systems filter data based on (a) the similarity between a user profile and the textual content of an event and (b) the user relevance feedback, that is, if a user likes an event then similar events should also be forwarded to that user.

IF has many similarities to IR. The key differences are that (a) in IF the data is streaming and (b) in IF the query is the user profile. A more detailed comparison of IR anf IF is available in the influential paper of Belkin and Croft (1992).

General-purpose IF systems include Tapestry and INFOSCOPE. In Tapestry (Goldberg et al., 1992), this problem is handled by using "cooperative information filtering"; specifically, users cooperate by recording their opinions of information entries that they read. In addition, in INFOSCOPE (Stevens, 1992), the usage patterns of individual users are recognized, and information content evaluation is performed by rule-base agents. Autodesk (Baclace, 1992) learns the user profile by exploiting user feedback.

Recommender systems, as defined in Wikipedia, are active information filtering systems that attempt to present to the user information items (movies, music, books, news, web pages) the user is interested in. These systems add information items to the information flowing towards the user, as opposed to removing information items from the information flow towards the user. Recommender systems typically use collaborative filtering approaches or a combination of the collaborative filtering and content-based filtering approaches, although content-based recommender systems do exist.

6.2. *Challenges of information filtering in disaster management*

The challenges of IF in disaster management data have many similarities with the challenges in IE and IR described in Sections 4.1 and 5.1, respectively. In addition to those, when a large-scale disaster happens, the problem of information flood can be very serious because a great deal of information occurs in a short time and is sent to a person or an organization that is responsible for managing the situation (Atoji et al., 2004). The order in which information arrives is not necessarily that in which events occur. Thus, an ordinary communication system cannot ensure fast analysis of causal relations, leading to a risk of misjudgment. Also, the consumer of the filtered data in a disaster environment has limited time to digest the produced data, in contrast to other IF applications like news filtering.

6.3. *Proposed solutions for information filtering in disaster management*

6.3.1. *Filtering tagged data*

Atoji et al. (2004) uses a semi-structured message format to describe information in case of disasters and other emergencies; automatic processing is made possible by attribute extraction. A semi-structured message is a format that combines a structured template part with a free description area, providing a practicable way of data representation. Specifically, there is a structured template providing fields for the reporter's name, event location, development of situation, and so on. Using such structured templates, a message can be written easily. In addition, attribute values are added automatically at the moment when a message is sent. After that, other information can be submitted via free description. Self-Organizing Maps (SOM) (Kohonen, 1995) are used to classify the events. SOM is a type of neural network that produces a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map. Three keywords are displayed in the user interface for each event, chosen using tf·idf ranking (Salton and McGill, 1986).

Lee and Bui (2000) also follow a template-based approach (Table 2), where customized templates are created for each type of expected disaster, which are filled during the disaster recovery stage. Clearly, this makes the filtering process much simpler than in unstructured text cases.

Table 2
An example of a template for the Kobe earthquake (Lee and Bui, 2000).

What	Have the rescue team in as soon as possible
Why	Rescue those in need of help
Where	Anywhere
Who	Military or Firefighters or Coast Guard or any other groups nearby trained for emergencies
When	Any Disaster where people need to be rescued
How	Contact (the list of potential rescue parties)
Activity decomposition	Direct communication, Block a road, Arrival with equipments needed
Resource Required	Heavy equipments
Success or failure	*Failed*
Reasons	Too late notice, blocked road, initial arrival with no equipment
Generalizability (context-sensitivity)	High

6.3.2. *Experiences from IF-related disaster systems*

Some disaster management projects (Housel et al., 1986; Bui and Sankaran, 2001) mention the support of IF without specifying any concrete method employed. Most probably such systems perform simple IF based on string matching. Bui and Sankaran (2001) analyzes the workflow typical in a disaster scenario and discusses the design considerations for a virtual information center that can both efficiently and effectively coordinate and process a large number of information requests for disaster preparation, management and recovery teams. According to this paper, workflows can be categorized depending on their organizational function and complexity level. This can help in subsequent system implementations since workflows of the same class tend to share coding techniques. Some of the workflows are the result of ad hoc requests common in decision support system environments. Concurrent workflows, on the other hand, coordinate a process among several system components in order to formulate an overall decision for a complex request for information. A third type of workflow can be categorized as jurisdictional workflows. They typically occur within an organizational function or unit alone such as sales or marketing. The workflows used in this paper are a combination of all the above types.

7. Data mining in disaster management

Overview of Data Mining: data mining or knowledge discovery is the nontrivial extraction of implicit, previously unknown, and potentially useful information from large collection of data (Han and Kamber, 2006; Tan et al., 2005). In practice, data mining refers to the overall process of extracting high-level knowledge from low-level data in the context of large databases.

In order to systematically review data mining research in disaster management, we use the framework introduced in Li et al. (2002) to describe the data mining process. The framework consists of an iterative sequence of the following steps: *data management*, *data preprocessing*, *data mining tasks and algorithms*, and *post-processing*.

First, *data management*, which is discussed in Section 3, concerns the specific mechanism and structures for how the data are accessed, stored and managed. The data management is greatly related to the implementation of data mining systems.

Next, *data preprocessing*, which is also discussed in Section 3, is an important step to ensure the data quality and to improve the efficiency and ease of the mining process. Real-world data in disaster management tend to be incomplete, noisy, inconsistent, high dimensional and multi-sensory etc. and hence are not directly suitable for mining. Data preprocessing usually includes data cleaning to remove noisy data and outliers, data integration to integrate data from multiple information sources, data reduction to reduce the dimensionality and complexity of the data and data transformation to convert the data into suitable forms for mining, etc.

Third, *data mining tasks and algorithms*, which are the focus of this section, are essential steps of knowledge discovery. There are many different data mining tasks such as association mining, exploratory data analysis, classification, clustering, regression, and content retrieval etc. Various algorithms have been used to carry out these tasks and many algorithms could be applied to several different kinds of tasks.

Finally, we need *post-processing* stage to refine and evaluate the knowledge derived from our mining procedure (Bruha and Famili, 2000). For example, one may need to simplify the extracted knowledge. Also, we may want to evaluate the extracted knowledge, visualize it, or merely document it for the end-user. We may interpret the knowledge and incorporate it into an existing system, and check for potential conflicts with previously induced knowledge. Since post-processing mainly concerns the non-technical work, such as documentation and evaluation, it is not discussed further.

7.1. Challenges of data mining in disaster management

The challenges for data mining in disaster management are grouped as follows.

7.1.1. Data preprocessing

The advances in storage technology and network architectures have made it possible and affordable to collect and store huge amounts of disaster data in various media types (text, image, video, graphics, etc.), formats, and structures from multiple information sources. These data typically include a significant amount of missing values and noises, and may have multi-level completeness, multi-level confidences, and may be inconsistent. In order to improve the efficiency and accuracy of knowledge discovery and data mining process, ensuring data quality is a big challenge.

7.1.2. Data mining tasks and algorithms

Since the data in disaster management may come from various sources and different users might be interested in different kinds of knowledge, data mining in disaster management typically involve a wide range of tasks and algorithms such as pattern mining for discovering interesting associations and correlations; clustering and trend analysis to understand the nontrivial changes and trends and classification to prevent future reoccurrences of undesirable phenomena. These different data mining tasks may use the same database in different ways. A first challenge is how to achieve the efficient data mining of different kinds of knowledge using different data mining algorithms.

Second, many datasets in disaster management are of geo-spatial type. How to effectively detect geo-spatial patterns with semantic awareness is not a trivial task. On one hand, it is not a trivial task to define a local neighborhood for mining geo-spatial patterns that includes space, time and semantic information. It is also challenging to incorporate domain-specific information (e.g., semantic ontology) into the mining process without compromising the underlying performance. On the other hand, many disaster phenomena may be localized to specific region at specific times. Therefore geo-spatial pattern mining for disaster management needs to be conducted across multiple regions with multiple spatial scales at different time periods.

Third, several special characteristics of disaster management data pose new challenges for applying traditional data mining methods: (1) the disaster data generally come from different sources and are of heterogeneous nature. Data mining across multiple information sources is a critical and challenging task because of the vast difference in data type, dimension and quality. (2) The disaster data contains inherent uncertainty and impreciseness due to the random nature of data generation and collection process. Although there has been much research work in data uncertainty management and in querying data with uncertainty, there is only limited research work in mining uncertain data. (3) In disaster management, unpredictable events often indicate suspicious situations. However, these events are extremely difficult to detect because they don't occur often or they occur at a time/location where they are not expected. In other words, there are only limited samples for the target class. (4) Disaster management applications are generally domain-specific. How to utilize the domain knowledge to guide data mining process or improve data mining performance is a challenging issue.

7.1.3. Post-processing

Challenges also exist in post-processing. For example, how to evaluate the discovered patterns and how to present the mining results to the domain experts in a way that is easy to understand and interpret? How to convert the discovered patterns into knowledge?

7.2. Proposed solutions for data mining in disaster management

7.2.1. Exploratory data analysis

For disaster management, many map-based tools intended to support exploratory analysis of spatial data and decision-making in a geographic context are developed for exploratory data analysis. For example, Andrienko and Andrienko (1999) proposed an integrated environment for exploratory analysis of spatial data that equips an analyst with a variety of data mining tools and provided the service of automated mapping of source data and data mining results. Best et al. (2005) described data analysis methods (classification, clustering, etc.) for mapping news events gathered from around the world. Web-based graphical map displays are used to monitor both the real-time situation, and longer term historical trends.

7.2.2. Clustering/classification

Correct forecasts and predictions of natural phenomena are vitally important to allow for proper evacuation and damage mitigation strategies. Li et al. (2008) used decision tree with the C4.5 algorithm to discover the collective contributions to tropical cyclones from sea surface temperature, atmospheric water vapor, vertical wind shear, and zonal stretching deformation. Chung and Fabbri (1999) used Bayesian prediction models for landslide hazard mapping. Chung et al. (2005) developed a three-stage procedure in spatial data analysis to estimate the probability of the occurrence of the natural hazardous events and to evaluate the uncertainty of the estimators of that probability. Chang et al. (2007) used generalized positive Boolean functions for landslide classification. Janeja et al. (2005) employed density-based clustering algorithms for alert management.

Many classification methods have been used to analyze the image data, remote sensing data, or laser scanning data for damage management and disaster prediction (see: Simizu et al.; Rehor; Di Martino et al., 2007; Gamba and Casciati, 1998). Barnes et al. (2007) developed a system-level approach based on image-driven data mining with sigma-tree structures for detection, classification, and attribution of high-resolution satellite image features for hurricane damage assessments and emergency response planning.

7.2.3. Spatial data mining

Many of the data available in disaster management are geo-spatial data. Hence many spatial data mining techniques have been applied in disaster management (Torun and Düzgün, 2006).

8. Decision support in disaster management

8.1. Overview of decision support

Decision Support Systems (DSS), according to Wikipedia, are a specific class of information systems that supports business and organizational decision-making activities. In particular, a properly-designed DSS is an interactive software-based system intended to help decision makers compile useful information from diverse data sources and/or business models to identify and solve problems and make decisions (Sprague and Watson, 1993; Turban et al., 2008).

As described in Marakas (1999), the primary components of a DSS include (1) the data management component, (2) the model management component, (3) the knowledge engine, (4) the user interface, and (5) the user(s). Basically, the data management component stores information from diverse sources including organization's data repositories and external data collections. The model management component handles representations of semantic events and situations and provides various optimization models, analytical and data mining tools. The knowledge engine manages the domain knowledge as well as the knowledge generated by various models and tools in the model management component. The knowledge engine can also provide inference capabilities and search engines for knowledge engineering. The user interface allows users to interact with the system and handles the dialog.

As we described in Section 1, disaster management can be divided into the following four phases: (a) Preparedness; (b) Mitigation; (c) Response; (d) Recovery. There are many decision-making activities associated with each phase. For example, in Preparedness phase, there are hazard assessment for vulnerability analysis, risk management for analyzing and evaluating disaster risks, and resource management and planning. In Mitigation phase and Response phase, mitigation plans and emergency response plans need to be developed, analyzed and evaluated. In Recovery phase, damage assessments and re-settlement issues should be addressed.

8.2. Challenges of decision support in disaster management

Generally, decision support in disaster management aims to:

- Support for more effective and comprehensive situation analysis and assessment.
- Assistance in finding appropriate recovery actions and scenarios.
- Help in handling multiple decision criteria and conflicting goals of multiple factors.

There are several challenges in developing decision support systems for disaster management:

- *Coordination*: a good decision support system for disaster management should be an integrated system containing the five primary components, i.e., (1) the data management component, (2) the model management component, (3) the knowledge engine, (4) the user interface, and (5) the user(s). In order to function well, the system needs coordination among these components.
- *Integration*: Traditionally, individual decision models handle specific decision-making needs. In current disaster management applications, there are typically many different decision scenarios. Therefore, one single model might not be sufficient for addressing the needs that arise in disaster management. Therefore there is a need to combine/aggregate individual decision models (Meissner et al., 2002).
- Model *adaption* along with time: Disaster management is an area with dynamic needs and an adaptive nature (Thomas and Larry, 2001). Therefore, it is necessary to adapt the decision models to the dynamic needs of disaster management area.
- Making decision with *uncertain* data: The data in disaster management may have multi-level completeness, multi-level confidences, and may be inconsistent. Making decisions with various sources of uncertainty is a very challenging task.

8.3. Proposed solutions for decision support in disaster management

Over the last few years, a large number of decision support systems have been developed for various types of disasters based on different models and decision support needs. A web-based decision support system for the response to strong earth-quakes is developed in Yong and Chen (2001). The system was focused on damage assessment and identification of effective response measures. A simulation model for emergency evacuation was developed in Pidd et al. (1996). In Gadomski et al. (2001), an agent-based system for knowledge management and planning is developed for suggesting plans in emergency scenarios. Wang et al. (2004) developed a DSS of flood disaster for property insurance. Decision support systems have also been developed for hurricane emergencies (Tufekci, 1995; Lindell and Prater, 2007), evacuation planning (Silva, 2001), oil spill (Pourvakhshouri and Mansor, 2003), forest fire (Kohyu et al., 2004), and rescue operations (Farinelli et al., 2003).

Andrienko and Andrienko (2005) discussed the intelligent decision support tools within the OASIS project whose goals are to improve situation analysis and assessment, provide assistance in finding appropriate recovery actions and scenarios, and help in handling multiple decision criteria. OASIS is an Integrated Project (IP) focused on the Crisis Management part of "improving risk management" of EC Information Society Technologies program (see www.oasis-fp6.org).

9. Case study: supporting business communities in disaster situations

In this section we focus on a subproblem of disaster management and discuss our experiences from working on the Business Continuity Information Network (BCIN, http://www.bizrecovery.org/) in South Florida with the collaboration of government and business entities, including the Miami-Dade Emergency Operations Center, IBM and Office Depot. This case study will help to motivate some of the research directions presented in Section 10. A screenshot of the BCIN prototype is shown in Fig. 5.

Current crisis management and disaster recovery systems/methodologies are aimed at fostering collaboration among the local, state and federal agencies for preparation and recovery process. These systems and methodologies have failed to include private businesses in the preparation and recovery processes. Access to time critical information about an impending hazard or under post-disaster conditions for business community is extremely limited and is made available to them after considerable delay which inhibits effective and efficient preparation along with planning and execution of precautionary and disaster recovery measures. Collaboration among the emergency management officials and the private business owners can assist in rapid recovery after a disaster and help mitigate the financial and economical losses associated with such disasters. Hence, there is an imperative need for a comprehensive business oriented disaster recovery information network that can facilitate collaboration among emergency management officials and private businesses, thus ensuring availability of and access to time critical information. Thus, we proposed a model for pre-disaster preparation and post-disaster business continuity/rapid recovery. The model involves elements

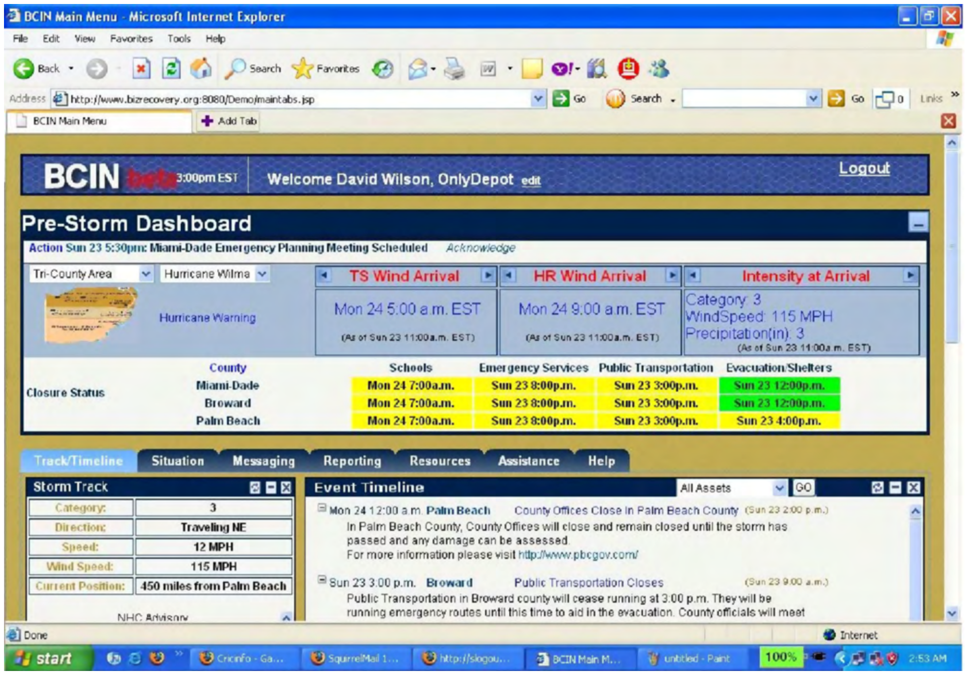

Fig. 5. Screenshot of BCIN application.

from Emergency Operations Centers in the South Florida Region along with the business community to provide a collaboration and communication channel facilitating disaster preparation and recovery plan execution, critical communications, local and regional damage assessment, dynamic contact management, situation awareness, intelligent decision support and identification/utilization of disaster recovery resources. We utilized our model approach to design and implement a web-based prototype implementation of our Business Continuity Information Network (BCIN) for rapid disaster recovery system, employing a fictitious Hurricane Tony as the case study. BCIN illustrates the pre-/post-disaster information exchange among the participants and helps in analyzing the effectiveness of such collaboration.

10. Future research directions on data management and analysis

In this section we present the data management and analysis research needed that will enable building disaster management systems of great impact and utility. These directions have been compiled from our own experiences and research, and from past federal grant solicitations (Informatics for Disaster Management, http://grants.nih.gov/grants/guide/pa-files/PA-03-178.html). Section 10.1 presents the challenges on disaster-related data management from the perspective of dataspaces, which is a recent paradigm that models heterogeneous data sources.

Directions related to data integration and ingestion in disaster management:

- While there are existing approaches attempting to address the challenges in integration and ingestion of disaster-related data, an adaptive, flexible, and customizable solution that can be quickly deployed and evolved with the phases of the disaster is being sought. Improvements upon the functionalities and/or methodologies of the existing disaster management systems, in particular Web-based systems, should be investigated to ensure a disaster preparation and recovery framework that can be utilized under different disaster conditions.
- The data in disaster management application domains are collected and stored in various media types (text, image, video, graphics, etc.), formats, and structures from multiple information sources. Many of these data are streaming data and collected in real-time, which may have temporal and spatial characteristics. These data may have different levels of completeness and confidence, and may be inconsistent. To make the situation worse, there is also information shift problem in these applications. That is, the information/knowledge may be constantly changed, outdated, and incomplete. To our best knowledge, there are no existing methodologies and/or algorithms that can properly address the aforementioned issues satisfactorily. Developing novel techniques for managing and analyzing data, that are from heterogeneous sources, having a mix of unstructured and structured types, of different temporal and spatial characteristics, with various sources of uncertainty, is a challenging task and is very important.

Directions related to IR, IE, IF in disaster management:

- Intelligent querying on heterogeneous, multi-source streaming disaster data. To support ad hoc or continuous queries on the streaming disaster data, effective and efficient information discovery algorithms must be created that take into consideration the context and the stakeholder's profile. In addition to plain keyword queries, predefined domain-specific parameterized queries must be supported, whose answer is computed by appropriately combined rules specified by a domain expert.
- High-throughput indexes and other precomputation techniques for streams must be developed to facilitate real-time query response times, and discover answers spanning across multiple streams. A two-phase (data collection, indexing) process is not adequate given the need for fresh data and advanced querying capabilities.
- Build IE algorithms that work in a semi-automatic way and learn from the interaction of a human with the IE system. For instance, these algorithms should learn the way that each organization reports its data, using some user IE hints.

- Exploit a combination of relevance feedback and context information to make precise decisions on what piece of information is appropriate for which stakeholder. Leverage and adapt previous works on modeling context and user characteristics.
- Incorporate uncertainty and possibly adversity in the handling and delivery of data. Authentication and social networking principles must be applied. Create reliability ranking mechanisms for the input sources. E.g., if most sources say that a bridge is open and only one says that it is closed, then the reliability rank of this latter source should be decreased.
- Adjust to the physical characteristics of the situation, e.g., to the network bandwidth, the computational power of the used devices. Novel criteria for the top answers of a query must be defined, considering the psychological state of the user and the device/environment capabilities.

Data mining and decision support-related research directions:

- Data mining from Multiple Information Sources. The disaster data are generally coming from different sources and are of heterogeneous nature. Data mining across multiple information sources is a critical and challenging task because of the vast difference in data type, dimension and quality (Wu et al., 1999; Grossman and Mazzucco, 2002). Two major issues must be addressed in order to perform data mining across multiple information sources: (i) scale issue: data types must be unified or scaled in order to allow comparison and combination. (ii) consistency issue: The data must be weighted or verified in a quantitative and consistent manner.
- Efficient pattern recognition, data mining, and knowledge extraction algorithms. To efficiently discover information from the huge amount of data collected in disaster management, the pattern recognition, data mining as well as knowledge extraction algorithms need to be efficient and scalable. The performance of the algorithms should be acceptable for large-scale datasets.
- Utilizing domain knowledge: In disaster management, domain knowledge is very useful for the preparation and recovery process. Thus an important yet challenging research direction is how to utilize the domain knowledge to guide data mining process, improve data mining performance, and make decisions.
- Community-based disaster management: disaster management is often community-based. Here a (virtual) community is defined as one of several communities whose members may include one or more company, government agency, and other non-governmental organizations which are related by their geographic location, their supply chain dependency, jurisdictional authority, etc. One of the research directions is how to build community (social network) and use community information for disaster management.
- Mining and making decisions on time-evolving and uncertain data: disaster management is an area with dynamic needs and an adaptive nature. In addition, the data in disaster management may have multi-level completeness, multi-level confidences, and may be inconsistent. An important research direction is how to mine and make decisions on time-evolving and uncertain data.

Finally, we present some more high-level directions from the National Academy report on "Improving Disaster Management: The Role of IT in Mitigation, Preparedness, Response, and Recovery" (Improving Disaster Management) that also apply to data management and analysis. First, disaster management organizations should work closely with technology providers to define, shape, and integrate new technologies as a coherent part of their overall IT system. Second, create metrics to inform cost-benefit decisions for investment in IT for disaster management and make enhanced end-user performance a primary objective of disaster management acquisition programs.

10.1. Disaster management dataspace support platform

In this section we present the challenges on management and analysis of disaster data using the recently proposed paradigm of *dataspaces* (Franklin et al., 2005). We select to present the problem from this perspective given that a disaster management system shares many challenges with the management of heterogeneous data sources in the dataspace paradigm. A dataspace is a loosely integrated set of data sources, where integration occurs in a lazy way, following a pay-as-you-go approach. In the case of disaster management, the sources include reports from government agencies, business entities and Non-Government Organizations (NGOs), media announcements, and web blogs.

With the increase in popularity of communication via different data types like structured reports, XML documents, unstructured e-mail communications, images, etc., having a common management system for all these data types becomes inevitable. The concept of dataspace (Franklin et al., 2005) thus became an alternative to traditional Database Management Systems (DBMSs) to handle diverse data types from varied sources. A DSSP can offer a suite of interrelated services on a set of large and loosely coupled data sources, where the data can be under different administrative control and not necessarily under the control of the DSSP, and the relationship between data sources need not be necessarily known. A DSSP provides data integration over all data sources, including those with different levels of complexity and based on different models. It facilitates a uniform interface that examines all available data to provide incremental integration improvement, and searches data sources to help users automatically develop data relationships (Franklin et al., 2005).

It has been shown that a dynamic disaster management system (Saleem et al., 2008) has the potential of being managed best by a loosely coupled, flexible, dynamic dataspace system as there is a constant influx of information, consisting of structured and unstructured data, of different types and from independent sources, which need to be efficiently stored, indexed, retrieved, and optimized. We argue that there is a need for a *Disaster Management Dataspace Support Platform (DM-DSSP)*, with elements suggested by Franklin et al. (2005).

However, the current DSSP approach requires manually building relationships/partitioning data, which can be referred to as the "man-in-the-loop" problem; this limitation makes it inadequate for our purposes. Therefore, there is a need and opportunity to exploit the limited scope of the disaster management domain to refine and extend the dataspace approach by developing novel methodologies to support DM-DSSP so that it can be applied to a disaster scenario, and provide rapid on-demand data integration within the disaster context tailored for virtual communities (social networking, communities of interest). The successful creation of a DM-DSSP will need to leverage knowledge from many research areas (such as data management, data integration, data mining, and information retrieval) and applications (such as digital government, social networks, homeland security, and military).

Another key addition to the model of (Franklin et al., 2005) is the existence and accommodation of *virtual communities* in a disaster management context. For instance, in the Business Continuity Information Network (BCIN) presented in Section 9, a dynamic virtual community is defined as one of several communities whose members may include one or more companies, government agencies, and other NGOs which are related by their geographic location, their supply chain dependency, jurisdictional authority, etc. Since a member of each community can participate in multiple communities simultaneously, different relationships can be developed, creating an explosion of interactions which in turn generate a rich set of participant data sets. These communities can be partitioned according to their relative relationships, and thus virtual communities, which are based on their common relational attributes, can be created. Further, new communities appear as a hazard impacts normal operations of a participating member and new relationships are created on-the-fly to respond to recovery needs. The existence of virtual communities provides opportunities to improve the data delivery, searching, reliability and integration in the DM-DSSP. Principles from collaboration filtering and wisdom of the crowd must be applied.

11. Conclusions

In this paper we present for the first time a comprehensive survey of the efforts on utilizing and advancing the management and analysis of data to serve disaster management situations. We organized our findings across the following Computer Science disciplines: data integration and ingestion, information extraction, information retrieval, information filtering, data mining and decision support. We also presented our own experiences from working with the Miami-Dade county Emergency Management Office building a Business Continuity Information Network. Finally, we presented concrete future research directions for computer scientists to advance the knowledge of applied data management and analysis in disaster management.

Acknowledgements

This work is partially supported by NSF Grants HRD-083309 and IIS-0811922.

References

Anon. First International Workshop on Agent Technology for Disaster Management. http://users.ecs.soton.ac.uk/sdr/atdm/, 2006.
Andrienko, N., Andrienko, G., 2005. A concept of an intelligent decision support for crisis management in the OASIS project. In: Geo-information for Disaster Management. Springer, Berlin, Heidelberg.
Andrienko, G., Andrienko, N., 1999. Knowledge-based visualization to support spatial data mining. In: Proceedings of the 3rd Symposium on Intel Ligent Data Analysis, Amsterdam, The Netherlands, August, pp. 149–160.
Atoji, Y., Koiso, T., Nakatani, M., Nishida, S., 2004. An information filtering method for emergency management. Electrical Engineering in Japan.
Arnold, J.L., Levine, B.N., Manmatha, R., Lee, F., Shenoy, P., Tsai, M.C., Ibrahim, T.K., O'Brien, D.J., Walsh, D.A., 2004. Information-sharing in out-of-hospital disaster response: the future role of information technology. Prehospital and Disaster Medicine 19 (July–September (3)), 201–207.
Baclace, P.E., 1992. Competitive agents for information filtering. Communication ACM 35, 50.
Belkin, N.J., Croft, W.B., 1992. Information filtering and information retrieval: two sides of the same coin? Communication ACM 35 (December (12)), 29–38.
Bruha, I., Famili, A.F., 2000. Postprocessing in machine learning and data mining. SIGKDD Exlporations.
Barnes, C.F., Fritz, H., Yoo, J., 2007. Hurricane disaster assessments with image-driven data mining in high-resolution satellite imagery. IEEE Transactions on Geoscience and Remote Sensing 45 (June (6)), 1631–1640.
Best, C., van der Goot, E., Blackler, K., Garcia, T., Horby, D., Steinberger, R., Pouliquen, B., 2005. Mapping world events. In: Geo-information for Disaster Management. Springer, Berlin, Heidelberg.
Buyukkokten, O., Garcia-Molina, H., Paepcke, A., 2000. Focused web searching with PDAs. In: 9th International World Wide Web Conference (WWW 2000), Amsterdam, Netherlands, May.
Borning, A., Lin, R.K., Marriott, K., 2000. Constraint-based document layout for the web. ACM Multimedia Systems Journal 8 (October (3)), 177–189.
Brin, S., Page.F L., 1998. The anatomy of a large-scale hypertextual web search engine. Computer Networks 30 (1–7), 107–117.
Bui, T.X., Sankaran, S.R., 2001. Design considerations for a virtual information center for humanitarian assistance/disaster relief using workflow modeling. Decision Support Systems 31 (2), 165–179.
Ciravegna, F., 2001. Challenges in information extraction from text for knowledge management. IEEE Intelligent Systems and Their Applications 16 (6), 88–90.
Chapman, S., Ciravegna, F., 2006. Focused Data Mining for decision support in Emergency Response Scenarios. In: Proceedings of the International Semantic Web Conference 2006 Workshop on Web Content Mining with Human Language Technologies, Athens, GA, USA, November 2006.
Chung, C.-J.F., Fabbri, A.G., Jang, D.-H., Scholten, H.J., 2005. Risk assessment using spatial prediction model for natural disaster preparedness. In: Geo-information for Disaster Management. Springer, Berlin, Heidelberg.
Chung, C.F., Fabbri, A.G., 1999. Probabilistic prediction models for landslide hazard mapping. Photogrammetric Engineering and Remote Sensing 65–12, 1389–1399.
Cowie, J., Lehnert, W., 1996. Information extraction. Communication ACM 39 (1 (January)), 80–91, http://doi.acm.org/10.1145/234173.234209.
Chang, Y.L., Liang, L.S., Han, C.C., Fang, J.P., Liang, W.Y., Chen, K.S., 2007. Multisource data fusion for landslide classification using generalized positive boolean functions. GeoRS 45 (June (6)), 1697–1708.
Cafarella, M.J., Madhavan, J., Halevy, A.Y., 2008. Web-scale extraction of structured data. SIGMOD Record 37 (4), 55–61.
Chen, Y., Xie, X., Ma, W.Y., Zhang, H.J., 2005. Adapting web pages for small-screen devices. IEEE Internet Computing 9 (January (1)).
Di Martino, G., Iodice, A., Riccio, D., Ruello, G., 2007. A novel approach for disaster monitoring: fractal models and tools. GeoRS 45 (June (6)), 1559–1570.
Disaster Management Information System by FEMA, http://www.disasterhelp.gov/disastermanagement/dmistools.
Anon. http://disasterrecoveryworld.com a website by Business Continuity & Disaster Recovery Associates, 2009.
Anon. http://www.disaster-recovery-plan.com/plan.htm products by Disaster-Recovery-Plan.com, 2009.
Eikvil, L., 1999. Information extraction from world wide web – a survey. In: Technical Report 945, Norweigan Computing Center, Oslo, Norway.
Etzioni, O., Banko, M., Soderland, S., Weld, D.S., 2008. Open information extraction from the web. Communication ACM 51 (December (12)), 68–74.
E-Teams, by NC4. http://www.nc4.us/ETeam.php.
Foster, I., Grossman, R.L., 2003. Data integration in a bandwidth-rich world. In: Communications of the ACM, Special Issue on Blueprint for the future of high-performance networking, vol. 46, November (11), pp. 50–57.
Farinelli, A., Grisetti, G., Iocchi, L., Lo Cascio, 2003. Design and evaluation of multiagent systems for rescue operations. In: Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, Nevada.
Franklin, M., Halevy, A., Maier, D., 2005. From databases to dataspaces: a new abstraction for information management. ACM SIGMOD Record 34 (4), 27–33.
Farfán, F., Hristidis, V., Ranganathan, A., Weiner, M., 2009. XOntoRank: ontology-aware search of electronic medical records. In: IEEE International Conference on Data Engineering (ICDE).
Disaster Contractors Network. http://www.dcnonline.org.
Gadomski, A.M., Bologna, S., Di Costanzo, G., 2001. Towards intelligent decision support systems for emergency managers: The IDA approach. International Journal of Risk Assessment and Management 2 (3/4), 224–242.
Gamba, P., Casciati, F., 1998. GIS and image understanding for near-real-time earthquake damage assessment. PhEngRS 64 (October (10)), 987.
Gravano, L., Hatzivassiloglou, V., Lichtenstein, R., 2003. Categorizing web queries according to geographical locality. In: 12th ACM Conference on Information and Knowledge Management (CIKM 2003), New Orleans, Louisiana, USA, November.
Grishman, R., Huttunen, S., Yangarber, R., 2002. Real-time event extraction for infectious disease outbreaks. In: Proceedings of the Second international Conference on Human Language Technology Research (San Diego, California, March 24–27, 2002). Human Language Technology Conference. Morgan Kaufmann Publishers, San Francisco, CA, pp. 366–369.
Grossman, R., Mazzucco, M., 2002. DataSpace – a web infrastructure for the exploratory analysis and mining of data. IEEE Computing in Science and Engineering, 44–51.
Goldberg, D., Nichols, D., Oki, B.M., Terry, D., 1992. Using collaborative filtering to weave an information TAPESTRY. Communication ACM 35, 61–70.
Halliday, Q., 2007. Disaster Management, Ontology and the Semantic Web. http://polaris.gseis.ucla.edu/qhallida/DisasterManagementFinal.html, UCLA.
Han, J., Kamber, M., 2006. Data Mining: Concepts and Techniques, 2nd ed. Morgan Kaufmann, San Francisco.
Hanna, K.M., Levine, B.N., Manmatha, R., 2003. Mobile distributed information retrieval for highly-partitioned networks. Network Protocols, 38–47.
Huttunen, S., Yangarber, R., Grishman, R., 2002. Diversity of Scenarios in Information Extraction. Third International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas, Canary Islands, Spain.
Improving Disaster Management: The Role of IT in Mitigation, Preparedness, Response, and Recovery, 2007. Computer Science and Telecommunications Board (CSTB). National Research Council Of The National Academies. The National Academies Press.
Janeja, V.P., Atluri, V., Gomaa, A., Adam, N., Bornhoevd, C., Lin, T., 2005. DM-AMS: employing data mining techniques for alert management. In: Proceedings of the 2005 National Conference on Digital Government Research (Atlanta, Georgia, May 15–18, 2005). ACM International Conference Proceeding Series, pp. 103–111, vol. 89, Digital Government Research Center.
Jain, A., Doan, A., Gravano, L., 2007. SQL Queries Over Unstructured Text Databases. ICDE, 1255–1257.
Kohonen, T., 1995. Self-organizing maps. Springer-Verlag.
Kushmerick, N., 1997. Wrapper Induction for Information Extraction. Doctoral Thesis. UMI Order Number: AAI9819266. University of Washington.
Kaempf, C., Ihringer, J., Daemi, A., Calmet:, J., 2003. Agent-based information retrieval for flood management research. In: Proceedings of EnviroInfo 17th, Cottbus.
Klien, E., Lutz, M., Kuhn, W., 2006. Ontology-based discovery of geographic information services-An application in disaster management. Computers, Environment and Urban Systems 30 (January (1)), 102–123.

Kohyu, S., Weiguo, S., Yang, K.T., 2004. A Study of Forest Fire Danger Prediction System in Japan. In: Proceedings of the 15th International Workshop on Database and Expert Systems Applications (DEXA'04).
Lenzerini, M., 2002. Data integration: a theoretical perspective. In: PODS 2002, pp. 233–246.
Lee, J., Bui, T., 2000. A template-based methodology for disaster management information systems. In: 33rd Annual Hawaii International Conference on System Sciences.
Li, T., Li, Q., Zhu, S., Ogihara, M., 2002. A survey on wavelet applications in data mining. SIGKDD Explor. Newsl. 4 (December (2)), 49–68.
Mondschein, L.G., 1994. The role of spatial information systems in environmental emergency management. Journal of the American Society for Information Science 45 (9), 678–685.
Lindell, M.K., Prater, C.S., 2007. A hurricane evacuation management decision support system (EMDSS). Natural Hazards 40 (3), 627–634.
Liu, F., Yu, C., Meng, W., 2004. Personalized Web Search For Improving Retrieval Effectiveness. IEEE Transactions on Knowledge and Data Engineering 16 (January (1)), 28–40.
Li, W., Yang, C., Sun, D., 2008. Mining geophysical parameters through decision tree analysis to determine correlation with tropical cyclone development. Computers & Geosciences.
Marakas, G.M., 1999. Decision support systems in the twenty-first century. Prentice Hall, Upper Saddle River, NJ.
Michalowski, M., Ambite, J.L., Thakkar, S., Tuchinda, R., et al., 2004. Retrieving and Semantically Integrating Heterogeneous Data from the Web. IEEE Intelligent Systems 19 (May (3)), 72–79.
Meissner, A., Luckenbach, T., Risse, T., Kirste, T., Kirchner, H., 2002. Design challenges for an integrated disaster management communication and information system. In: The First IEEE Workshop on Disaster Recovery Networks (DIREN 2002), in conjunction with IEEE INFOCOM 2002, New York, USA, June 24.
Message Understanding Conferences (MUCs) http://www-nlpir.nist.gov/related_projects/muc/.
National Emergency Management Network. http://www.nemn.net/.
National Incident Management System. http://www.fema.gov/emergency/nims.
Naumann, F., Raschid, L., 2006. Information Integration and Disaster Data Management (DisDM) Workshop on Information Integration, Philadelphia, USA, October 25–27.
National Research Council, 2007. Improving Disaster Management: The Role of IT in Mitigation, Preparedness, Response and Recovery. The National Academies Press, ISBN-13:978-0-309-10396-1.
Neches, R., Yao, K., Ko, I., Bugacov, A., Kumar, V., Eleish, R., 2002. GeoWorlds: integrating GIS and digital libraries for situation understanding and management. The New Review of Hypermedia and Multimedia 7 (1), 127–152.
Expecting the Unexpected. Office Depot Brochure, 2007. http://www.officedepot.com/speciallinks/us/od/docs/promo/pages/docs/onlinedisasterbrochure.pdf.
Pidd, M., De Silva, F., Eglese, R., 1996. A simulation model for emergency evacuation. European Journal of Operational Research 90, 413–419.
Pourvakhshouri, S., Mansor, S., 2003. Decision support system in oil spill cases. Disaster Prevention and Management 12 (3), 217–221.
Rehor, M., Classification of Building Damages Based on Laser Scanning Data, Laser 07(326).
RESCUE Disaster Portal. http://rescue-ibm.calit2.uci.edu/DisasterPortalProject/disasterportal.html.
Resource Description Framework (RDF), http://www.w3.org/RDF/.
Rosen, J., Grigg, E., Lanier, J., McGrath, S., Lillibridge, S., Sargent, D., Koop, C.E., 2002. The future of command and control for disaster response. Engineering in Medicine and Biology Magazine, IEEE 21 ((September/October) 5).
Sahana Open Source Disaster Management System. http://www.sahana.lk/.
Silva, F.N.d., 2001. Providing decision support for evacuation planning: a challenge in integrating technologies. Disaster Prevention and Management 10 (1), 11–20.
Soderland, S., 2004. Learning information extraction rules for semi-structured and free text. Machine Learning (November (30)), 233–272.
Stevens, C., 1992. Automating creation of information filters. Communication ACM 35, 48.
Simizu, H., Gotoh, T., Saji, H. Automatic detection of earthquake damaged areas from aerial images and digital map, ICIP03 (I: 829–832).
Sprague, R.H., Watson, H.J., 1993. Decision Support Systems: Putting Theory into Practice. Englewood Clifts, NJ, Prentice Hall.
Surdeanu, M., Harabagiu, S.M., 2002. Infrastructure for open-domain information extraction. In: Proceedings of the Second international Conference on Human Language Technology Research (San Diego, California, March 24–27, 2002). Human Language Technology Conference. Morgan Kaufmann Publishers, San Francisco, CA, pp. 325–330.
Thomas, S.D., Larry, C., 2001. Disaster Management and Preparedness. Lewis Publisher, New York.
Saleem, K., Luis, S., Deng, Y., Chen, S.-C., Hristidis, V., Li, T., 2008. Towards a business continuity information network for rapid disaster recovery. In: Proceedings of the 9th Annual International Conference on Digital Government Research, Montreal, Canada, May 18–21, pp. 107–116.
Salton, G., McGill, M.J., 1986. Introduction to Modern Information Retrieval. McGraw-Hill, Inc.
http://www.strohlsystems.com, products by Strohl Systems Group, Inc.
Tufekci, S., 1995. An integrated emergency management decision support system for hurricane emergencies. Safety Science 20, 39–48.
Turban, E., Aronson, J.E., Liang, T.-P., 2008. Decision Support Systems and Intelligent Systems. Pearson Prentice Hall.
Turoff, M., Chumer, M., Van del Walle, B., Yao, X., 2003. The design of a dynamic emergency response management information system (DERMIS). Journal of Information Technology Theory and Application 5 (4), 1–36.
Troy, D.A., Carson, A., Vanderbeek, J., Hutton, A., 2008. Enhancing community-based disaster preparedness with information technology: community disaster information system. Disasters 32 (March (1)), 149–165.
Torun, A., Düzgün, E., 2006. Using spatial data mining techniques to reveal vulnerability of people and places due to oil transportation and accidents: a case study of Istanbul strait. In: ISPRS Technical Commission II Symposium, Vienna, 12–14 July.
Housel, T.J., El Sawy, O.A., Donovan, P.F., 1986. Information systems for crisis management: lessons from southern California Edison. MIS Quarterly 10 (December (4)), 389–400.
Tan, P.-N., Steinbach, M., Kumar, V., 2005. Introduction to Data Mining. Addison Wesley.
Vögele, T., Hübner, S., Schuster, G., 2003. BUSTER – an information broker for the semantic web. KI-Künstliche Intelligenz.
Voluntary Private Sector Preparedness Accreditation and Certification Program administrated by FEMA. http://www.fema.gov/media/fact_sheets/vpsp.shtm, 2009.
WebEOC, Manufactured by ESi Acquisition, Inc. http://www.esi911.com/home.
Wobbrock, J.O., Forlizzi, J., Hudson, S.E., Myers, B.A., 2002. WebThumb: interaction techniques for smallscreen browsers. In: 15th Annual Symposium on User Interface Software and Technology (UIST 2002), Paris, France, October.
Wachowicz, M., Hunter, G.J., 2005. Dealing with uncertainty in the real-time knowledge discovery process. In: Geo-information for Disaster Management. Springer, Berlin Heidelberg.
Wu, L., Oviatt, S.L., Cohen, P.R., 1999. Multimodal integration – a statistical view. IEEE Transactions on Multimedia 1 (4), 334–341.
Wang, L., Qin, Q., Wang, D., 2004. Decision support system of flood disaster for property insurance: theory and practice. In: In Proceedings 2004 IEEE International Geosciences and Remote Sensing Symposium, 2004 IGARSS '04, vol. 7, 20–24, September, pp. 4693–4695.
W3C. Disaster Management Wiki. http://esw.w3.org/topic/DisasterManagement?highlight=%28management%29%7C%28disaster%29.
eXtensible Markup Language, http://www.w3.org/XML.
Yong, C., Chen, Q.F., 2001. Web based decision support tool in order to response to strong earth-quakes. In: Proceedings of TIEMS2001, Oslo, Norway.
Yangarber, R., Grishman, R., 1997. Customization of Information Extraction Systems. In: Proceedings of International Workshop on Lexically Driven Information Extraction, Frascati, Italy, July 16.

Vagelis Hristidis received the BS degree in Electrical and Computer Engineering from the National Technical University of Athens, and the MS and PhD degrees in Computer Science from the University of California, San Diego, in 2004. Since then, he is an Assistant Professor in the School of Computing and Information Sciences at Florida International University, in Miami. Vagelis has received funding from the National Science Foundation, the Department of Homeland Security and the Kauffman Entrepreneurship Center, including the NSF CAREER Award. His main research work addresses the problem of bridging the gap between databases and information retrieval. He has published more than 45 research articles, which have received more than 1,500 references as reported by Google Scholar. He also recently edited and co-authored a book on "Information Discovery on Electronic Health Records". For more information, please visit http://www.cis.fiu.edu/~vagelis/.

Dr. Shu-Ching Chen is a Full Professor in the School of Computing and Information Sciences (SCIS), Florida International University (FIU), Miami since August 2009. Prior to that, he was an Assistant/Associate Professor in SCIS at FIU from 1999. He received Master's degrees in Computer Science, Electrical Engineering, and Civil Engineering in 1992, 1995, and 1996, respectively, and the Ph.D. degree in Electrical and Computer Engineering in 1998, all from Purdue University, West Lafayette, IN, USA. His main research interests include distributed multimedia database management systems and multimedia data mining. Dr. Chen received the best paper award from the 2006 IEEE International Symposium on Multimedia. He was awarded the IEEE Systems, Man, and Cybernetics (SMC) Society's Outstanding Contribution Award in 2005 and was co-recipient of the IEEE Most Active SMC Technical Committee Award in 2006. He was also awarded the Inaugural Excellence in Graduate Mentorship Award from FIU in 2006 and the University Outstanding Faculty Research Award from FIU in 2004. He has been a General Chair and Program Chair for more than 30 conferences, symposiums, and workshops. He is the Editor-in-Chief of International Journal of Multimedia Data Engineering and Management and Associate Editors/Editorial Board for several journals.

Tao Li is currently an Associate Professor in the School of Computer Science at Florida International University. He received his PhD in Computer Science in 2004 from the University of Rochester. His research interests are in data mining, machine learning, and information retrieval. He is a recipient of NSF CAREER Award and IBM Faculty Research Awards.

Steven Luis is the Director of Technology and Business Relations in the School of Computing and Information Sciences (SCIS) at Florida international University. His research interests include disaster information management, cloud computing, and mobile computing. He is the project manager on several applied research projects, such as the Business Continuity Information Network, developing both software requirements and design and the business development for commercial engagement. He manages the SCIS IT services and infrastructure. He holds a Masters in Computer Science from Florida International University.