eSMS - a Semantics-assisted Emergency Information System Based on Social Media

Juan Li, Qingrui Li, Ashok Ginjala, Nazia Zaman
North Dakota State University, Computer Science Department, Fargo, 58078, USA
Email: {j.li, qingrui.li, ashok.ginjala, nazia.zaman}@ndsu.edu

Abstract— Natural and man-made disasters pose an ever-present threat to our society. Providing real time information is vital in emergency management. Current disaster information systems only use and manage relatively limited information, such as, within an organization or with a few organizations which have an agreement on information content and format. As a result, a large amount of potentially precious information and natural volunteer workers are ignored. On the other hand, social media have been used by people to propagate emergent situations. Therefore, in this project, we aim to maximally utilize all of the available emergency-related information from various sources of social media to better manage natural and man-made disasters. In particular, we propose a semantics-extended social search engine that can fetch highly relevant information from social media for emergency management purpose.

Index Terms— Emergency Management, Information System, Social Media, Semantics, Semantic Web

I. Introduction

Natural and man-made disasters may involve massive amounts of data coming from many sources, such as the geographic scope of the impacted area and the path of damage, data about debris and infrastructure impact, and information regarding numbers of deaths, injuries, and damaged houses/businesses. In response to the growing number of recent hazard events, such as the, Pakistani Floods, Gulf of Mexico oil spill, Haitian earthquake, it will be very beneficial to provide broad interoperability of emergency-related datasets to facilitate researchers, engineers and educators to access, share, and study various emergency-related data. In addition, disasters require that a system exists to allow users to respond as quickly and efficiently as possible with the most updated and accurate information. However, when disaster-related information is generated and shared across the community, problems of information collection, integration, storage, and query are very challenging due to data diversity, quantity, and geographical distribution.

Built on the ideological and technological foundations of Web 2.0, Social media allows the creation and exchange of user-generated content [1]. With Social Network Sites (SNS), blogs, wikis, and video-sharing, the general public is able to interact or collaborate with each other in a social media dialogue as creators of user-generated content in a virtual community. For example, there are growing trends of using social media to create entities in popular SNSs, such as Facebook, Twitter, and LinkedIn. Moreover, Social media have been used by people to propagate emergent situations such as the Pakistan flood, the Haiti earthquake, and the Gulf of Mexico oil spill. Using social media, emergent information can be quickly disseminated out to the public. For example, numerous emergency related organizations, such as Federal Emergency Management Agency (FEMA) and Information System for Crisis Response and Management (ISCRAM) have been set up at Facebook. Many hospitals have Twitter accounts for residents to follow their status in case there is an emergency. Expert groups have been established on LinkedIn to facilitate emergency planning and wellness exercising techniques. Furthermore, the ubiquity of mobile wireless devices facilitates the general public's involvement in the generation, propagation, and consumption of information – the anywhere and anytime paradigm.

Moreover, in the case of a disaster, existing information infrastructure at the disaster location may break down and information may not be available through the usual channels. Social media may become the only platform for persons to report and acquire information about events and damages in their neighborhood which is not available by other means. Furthermore, the state of the situation may change extremely quickly. For instance, hospitals might shut down abruptly due to electricity shortcuts. Information about such events are highly distributed among people, crews, and coordinating centers. Social networking sites such as Twitter are ideal media to update such status. From these examples above, we can see that using social media, emergent information can be quickly disseminated out to the public. In addition, organizations can collect information related to the disaster to assist decision makings. On the other hand, it is dangerous not to engage social media and let rumors and misinformation to spread. Therefore, gathering important disaster information from social media is another major task of our system in support of disaster management.

The aforementioned factors motivate us to maximally utilize all of the available information and human power to make information available to heterogeneous user groups; thus, enhancing the current information and communication systems to better manage natural and man-made disasters. In particular, in this paper, we propose an information system based on the concept of "crowdsourcing" to incorporate social media in their efforts to reach a wider audience and communicate more

efficiently and effectively than ever before. We hypothesize that a large number of people, communities, and organizations can be effectively utilized managing large-scale disasters. On the other hand, information and knowledge also is critical for creating the desired impact on the individual and community behavior [1]. Correct and real-time information also assists in effective and coordinated communication to avoid confusion and mayhem [2, 3].

To utilize the large amount of valuable information from social media to assist disaster management, we collect information from a number of freely available social data sources on the web, such as Facebook and Twitter. The crawling is based on keywords extended the disaster-related ontology. We distinguish authoritative information from emergency related organizations, such as National Weather Service and Federal Emergency Management Agency at Facebook. Because disaster is time-sensitive, we collect information from social media at finer time granularities as if they are sensor streams.

The remainder of this paper is organized as follows: We review the current state of the art in information and communication technologies (ICT) for emergency management in Section 2. Our proposed information system is presented in Section 3, in which we collect relevant information from the social media for emergency management use. In Section 4, we evaluate the proposed methods and show their effectiveness. Finally, we conclude the paper in Section 5.

II. RELATED WORK

Over the last few decades, the importance of effective management of information is being increasingly recognized in disaster management [4, 5]. A number of countries have set up disaster management information systems according to their own specific needs. The applications of these systems range from emergency response planning to short-range early warning to long-range mitigation and prevention planning [5]. A variety of information and communication technologies (ICT) have been proposed for managing national, regional, or local natural disasters and man-made crises [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. ICT can be applied during different stages of a disaster, including disaster prevention, mitigation, preparedness, disaster response, and disaster recovery. Basically, IT technologies can be used for: (a) effective warning of disasters using different communication channels, (b) integrating information on necessary supplies and other sources; (c) coordinating disaster relief work; (d) encouraging social, institutional, and public responses; (e) evaluating the damages caused by a disaster and the need for disaster relief.

The development of ICT over the last few years has supported disaster management efforts by introducing numerous collaborative tools at different levels. For example, some open source disaster management tools have become very popular [18], e.g., Ushahidi [19], Sahana [20], and SwiftRiver [21]. Ushahidi was developed to report on the violence during the 2008 Kenyan general election. The idea behind the website was to harness the benefits of crowdsourcing information (using a large group of people to report on a story) and to facilitate the sharing of information in an environment where rumors and uncertainties were dominant. Since then, Ushahidi has been deployed more than 20 times around the world to assist in similar situations where little or no support is provided by governmental authorities responsible for emergency management. Ushahindi has been used to manage aid, manage volunteers, and track disaster relief camps and victims. Sahana is a web-based collaboration tool that addresses common coordination problems during a disaster, such as finding missing people, managing aid, managing volunteers, tracking disaster relief camps and the victims. SwiftRiver [21] is a free and open source platform that complements Ushahidi's mapping and visualization products. The goal of the project is "to democratize access to the tools for making sense of information" [21]. Therefore, SwiftRiver helps users to understand and act upon a stream of massive amounts of crisis data.

Non-open source software products have also become popular—particularly within the United States. For instance, Web EOC, DLAN, EM2000, and others allow all responding organizations within a membership area (typically purchased on a state wide basis) to monitor and manage information related to the event including sharing information about the disaster, providing situation updates, managing disaster related projects and resources, and more. Yet, There are also products like Odysseus™ that are available to facilitate collaborative projects in advance of a disaster like creating disaster exercises to test preparedness and build multi-jurisdictional mitigation and response plans. Yet, these products do not use the heterogeneous legacy data in different organizations related to an emergency. Additionally, because none of these products are open-source only those jurisdictions with significant resources are able to purchase the software leaving some of the most impoverished and poor communities more vulnerable. Moreover, since scientists, engineers, and educators are not typically considered part of the disaster response network, they do not typically have access to data about the event. This situation prevents empirical research from being conducted on disasters and the dissemination of findings that would benefit other communities in future events.

More recently, there have appeared tools to search social media to get valuable information. Social media search tools use keywords and fetch results from one or more sources of social media and aggregate the results into a single stream. Some social media tools such as WhosTalkin and SocialMention are available on the web. WhosTalkin allows users to search for conversations surrounding the topics that they care about most, whether it is a favorite sport, favorite food, celebrity, or your company's brand name [22]. SocialMention allows a user to easily track and measure what people are saying about a company, a new product, or any topic across the web's social media landscape in real-time [23]. Though these social media tools address a wide range of aspects, a

dedicated social search engine related to emergencies and disasters is not available. Additionally geographic and location based information is the need of the hour.

III. System Design

To utilize the large amount of valuable information from social media to assist disaster management, we should collect information from a number of freely available social data sources on the web, such as news feeds, personal blogs, tagged images, and mini feeds. Data from social media are heterogeneous in representation and semantics; therefore, we integrate the collected social data, and provide a unified query interface to users. We use Semantic Web technologies to address both representational and semantic heterogeneity from different media sources. In this part, we detail the procedures to construct the social media search engine.

First, we need to examine several aspects of social media as represented by a search engine based on certain queries. These aspects included: (1) The social media websites that would be searched. Although there are many social media websites, the most popular ones such as Facebook, Tweeter, and Linkedin are considered in this project. Due to privacy and security reasons, we are not able to search arbitrary posts from Facebook. Instead, we can collect some posts from public groups. Tweeter provides good search API for all of its posts. Linkedin can be used for searching of experts. (2) The way the search engine represented social media websites across different search result. Integration of location based information through a map interface will further enhance peoples' awareness in their geographic area of crisis, which is especially important in an emergency environment. (3) The appropriate keywords that would be used for search. This is critical as many users may not have detailed expertise or knowledge, and their resultant query formulations may be deficient and lead to incorrect responses. We expand user queries with related concepts in a pre-defined ontology as shown in Fig. 1.
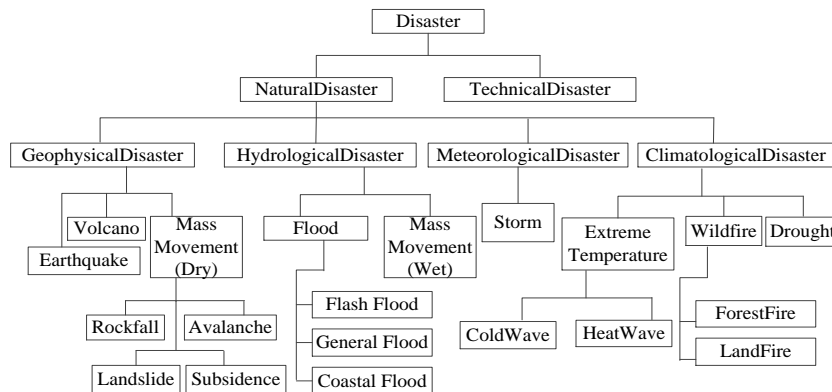

Fig. 1. Part of an ontology used in the system

The design and implementation of eSMS tool is mainly based on PHP server side scripting language in combination with Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript. MySQL is used as backend database.

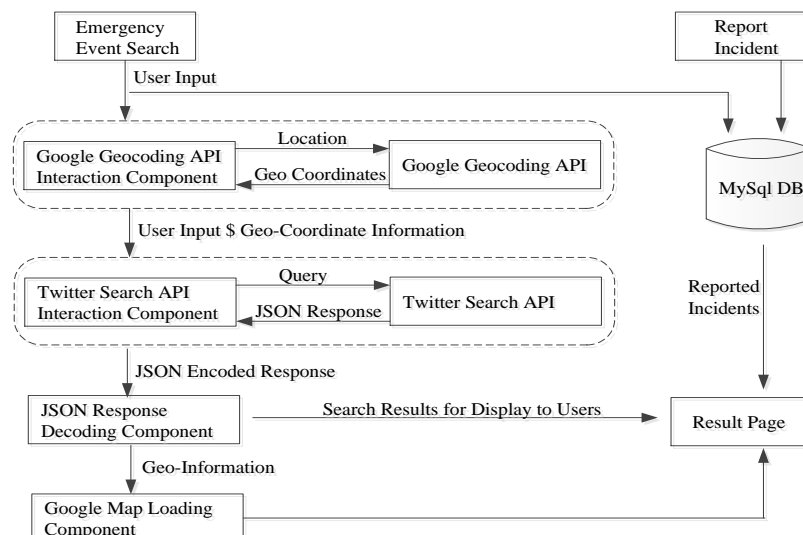The conceptual framework of information processing in the proposed system is shown in Fig. 2.

Fig. 2. System architecture

Over the last few decades, the importance of effective management of information is being increasingly recognized in disaster management [4, 5]. A number of countries have set up disaster management information systems according to their own specific needs. The applications of these systems range from emergency response planning to short-range early warning to long-range mitigation and prevention planning [5]. A variety of information and communication technologies (ICT) have been proposed for managing national, regional, or local natural disasters and man-made crises [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. ICT can be applied during different stages of a disaster, including disaster prevention, mitigation, preparedness, disaster response, and disaster recovery. Basically, IT technologies can be used for: (a) effective warning of disasters using different communication channels, (b) integrating information on necessary supplies and other sources; (c) coordinating disaster relief work; (d) encouraging social, institutional, and public responses; (e) evaluating the damages caused by a disaster and the need for disaster relief.

A. Design Overview

There are two major features that the system offers: The first is search feature. This feature searches the queried information against the real-time information from social websites and the local database of reported incidents. In our current implementation, we have initially considered incorporating the Twitter social media platform as the source of user generated content. These messages are commonly referred to as the "tweets" and forwarding these messages are referred to as "retweets" according to the Twitter jargon. Our search engine fetches the tweet updates based on the key word search and selected geographic location. Also it helps the users to retrieve the updates within a certain radius of the selected geographic location. These updates are topographically depicted on the Google maps based on the latitudinal and longitudinal information of the source of tweets.

The second feature is report Feature. This feature allows the user to record an incident in geographic vicinity that they have witnessed. This further pops up in the search query results if another user searches for information in the particular geographic location where the incident was reported. Then the engine fetches the results that match the query, from both social sites and local database and displays it to the user.

B. Search

The first and foremost step is to accept the users' input through a web user interface. The user input may include the following parameters: (1) Emergency Category, (2) Search Keyword, (3) Location, and (4) Radius, as shown in Fig 3. Emergency Category is defined based on the disaster ontology which we defined according to the International Disaster Database [24]. The category is illustrated as a multilevel dropdown menu that lists the possible types of emergencies/disasters. If a user does not specify the category, the system will try to extend the category information based on the keywords user input. The search is further narrowed down using either the zip code or city entered by the user. With this the user is providing the system with a location of interest. The user can select a value for radius, from the distance drop down menu. This distance parameter can further help the user to obtain tweets within a particular radius around the selected location

Fig. 3. Screenshot of the search interface

Once user submits the input parameters, they need to be converted into a valid format required for calling Twitter Search API, which acts as an interconnection between the eSMS tool and Twitter. The emergency category and search keyword(s) are concatenated together into a single string called as search term. This search term is arranged in such a way that it can be passed as a single input entity to the Twitter Search API. Also the location information is converted into geo-coordinate (latitude/longitude) values as Twitter Search API cannot understand the user entered location information as is.

As shown in Fig. 2, based on the input parameters received from the user, the location information is channelized to the Google Maps Geocoding API. This is further transformed by the Google Maps Geocoding API into geo-coordinate (latitude/longitude) values. The Geocoding API and Reverse Geocoding API can be accessed directly over HTTP by sending a request to Google Map. In Geo-Coordinate Generation process, the Geocoding API is called by sending an HTTP request to http://maps.googleapis.com/maps/api/geocode/xml?$addr ess, where '$address' is the location information received from the user. The API returns an XML response for this request, which is wrapped inside the root element and its sub-elements. Each sub-element contains a single set of geocoded address information and geometry information. This XML response is then parsed with the XML parser function within PHP to obtain the latitude and longitude coordinate values for a location parameter keyed-in by the user.

A call is made to Twitter Search API by sending a URL request over HTTP in order to retrieve the relevant tweets from Twitter. These tweets contain specific keywords and have been generated by users of a particular geographic area. Such URL is constructed using the 'search term', 'geo-coordinate values', and the radius parameter input from the eSMS user. When combined together, these key-value pairs form a URL request that is sent to Twitter Search API, which in turn returns a JSON object that encapsulates a set of real-time tweets having specific key words, and are generated by the users in a particular geographic area.

The response from Twitter Search API is parsed into multi-dimensional string array using PHP Json_decode() function. Each element in the array stores a single Twitter feed data. Each element of the array is iterated over the key-value pairs, and the values that the eSMS user is interested in are sent out for display, while the geographical coordinates information in the tweets are sent to the Google Maps Loading Component of the eSMS system. The eSMS tool uses the Google Maps feature to topographically depict the location information present in the tweet results, for the end user search. Google Maps JavaScript API provides a free service to embed Google Maps in the web pages and also provides a number of utilities for customizing and adding content to the map through a variety of services.

An example of the result page display is shown in Fig. 4. In this example the tweets for the Red River Flood are shown on the left hand side while the location based information through the Google maps interface is shown on the right. Specific locations of the tweets on the map are indicated by a green flag icon. It has to be mentioned here that Twitter might not have location information for all the tweets generated through the search. In such a situation the flag icons are shown only for the available location information.

Fig. 4. Screenshot of the search result

C. Report

Report incident is an important feature of eSMS, which enables the user to submit emergency related information. Once an incident is reported, it is stored in the local database and can be retrieved by searching.

The reported incidents are shown as special information by separating them in results section. Report feature was designed keeping both people as well as Emergency Administrative offices in mind. Emergency Administrative offices dealing with emergencies in a

particular location can find vital information on eSMS that can better equip them to deal with the situation when

incidents are reported by people of that area/location. Fig. 5 shows a screenshot of the Report Incident page.

Fig. 5. Screenshot of the report incidents interface

IV. SYSTEM EVALUATION

After a successful completion of the emergency information system, the system has been deployed to a website targeting the residents of the Fargo-Moorhead area. The website has been advertised in local forums and volunteers' Facebook walls. Although the website has not yet been widely used, we have collected some usage data.

Fig.6 plots the number of requests our system received in the period from September 2012 to September 2013. We received most of the quires in the cold winter (December – February) and in the flood season (March and April). Fig. 7 shows the records (mostly tweets) we retrieved in the same period.
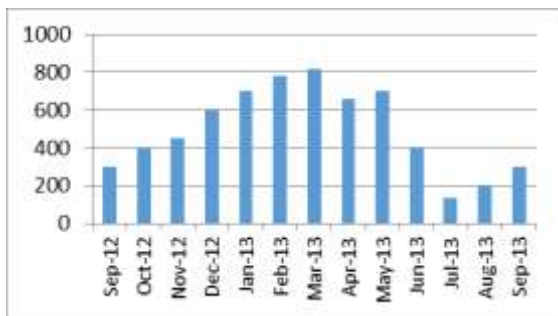

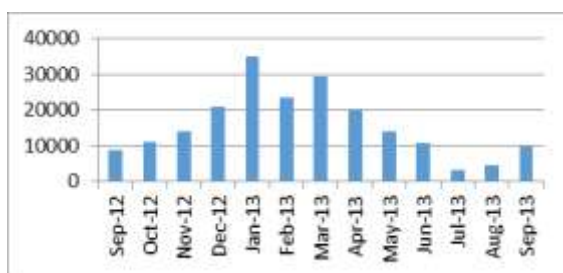Fig. 6. Number of requests received from Sep. 2012 to Sep.2013

Fig. 7. Number of records (tweets and internal reports) retrieved from Sep. 2012 to Sep. 2013.

Table 1. Top 3 queries

Query	# of appearance
storm (winter storm)*	709
blizzard	467
flood	388

* We count these two together as the same query

Table 1 lists the top three queries the system received during the one year period. We can see that the most popular query is about storms. Blizzard and flood are two other important queries.

V. CONCLUSIONS

We have designed and developed an effective social search engine for emergency management. This tool will assist people to collect, visualize, and propagate real-time information. The web user interface is extremely user friendly and can be easily understood and operable by a layman. This tool will enable the community to respond to emergency or event in a fast and effective manner, by provision of information concerning the "ground-zero" situation.

The current tool has limitations and lot of enhancements can be made to the existing tool. Apart from twitter, a number of other social networking sites can be integrated into the tool to press for a larger informational database. Ranking the results based on the trusted sources of information and by tracking the "current most trending" words related to the emergency in a particular geographic location. Additionally the types of emergencies and disasters can be further classified at a micro level. Validate the user reported emergency information to check the format before storing it in the eSMS local database tables.

REFERENCES

[1] Kaplan, A. M., Haenlein, M. "Users of the world, unite! The challenges and opportunities of Social Media". Business Horizons 53 (1): 59–68, 2010.

[2] Hancock, C, Johnson, CW, "Thinking the Unthinkable: Exposing the Vulnerabilities in the NHS Response to Coordinated Terrorist Actions.", In the International Systems Safety Conference. 2007.

[3] Starbird, K., Palen, L., Hughes, A. L., Vieweg, S., "Chatter on the red: what hazards threat reveals about the social life of microblogged information", Proceedings of the 2010 ACM

[4] Li, J., Li, Q., Khan S., Ghani, N. "Community-Based Cloud for Emergency Management", 6th IEEE International Conference on System of Systems Engineering (SoSE), Albuquerque, NM, USA, June 2011.

[5] Li, J., Li, Q., Liu, C., Khan S., Ghani, N. "Community-Based Collaborative Information System for Emergency Management", Computers & Operations Research – Elsevier, forth coming.

[6] O'Brien, G., Read, P., "Future UK emergency management: new wine, old skin?", Disaster Prevention and Management, Vol. 14 Iss: 3, pp.353 – 361, 2005.

[7] Wattegama, C., "ICT for Disaster Management, Asia-Pacific Development Information Programme" e-Primers for the Information Economy, Society and Polity. 2007.

[8] Wybo, J-L.and Lonka, H., Emergency management and information society: How to improve synergy?, Int. J. Emergency Management, 1, 2, 183-190. 2002.

[9] Mendonca, D., Jefferson, T. and Harrald, J. "Collaborative Adhocracies and Mix-and-Match Technologies in Emergency Management", Communications of the ACM, 50, 3, 45-49. 2007.

[10] Fiedrich, F. and Burghardt, P. , "Agent-based systems for disaster management," Commun. ACM, vol. 50, pp. 41- 42, 2007.

[11] Palen, L. and Liu, S. Citizen communications in disaster: Anticipating a future of ICT-supported public participation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI `07). ACM Press, NY, 2007.

[12] Fabbri, K., and Weets,G., "ICT for Environmental Risk Management in the EU Research Context", GEO-INFORMATION FOR DISASTER MANAGEMENT, Springer Berlin Heidelberg, 2005.

[13] Carver, L. and Turoff, M. , "Human-computer interaction: the human and computer as a team in emergency management information systems," Commun. ACM, vol.50, pp. 33-38, 2007.

[14] Becker, S. "Emergency communication and information issues in terrorist events involving radioactive materials". Biosecurity Bioterrorism: Biodefense Strategy Pract. Sci. 2(3):195–20, 2004.

[15] Hagar, C. "The information and social needs of Cumbrian farmers during the UK 2001 foot and mouth disease outbreak and the role of information and communication technologies." The socio-cultural impact of foot and mouth disease in the UK in (2001).

[16] Nerlich, B. eds. The Socio-Cultural Impact of Foot and Mouth Disease in the UK in 2001: Experiences and Analyses., Manchester University Press. 2009.

[17] Kreps, GL, Alibek, K, Neuhauser, L, Rowan, KE, Sparks, L. "Emergency/risk communication to promote public health and respond to biological threats", pp. 349–62. 2005.

[18] Currion, P., Silva, C.D., Walle,B.V., "Open source software for disaster management", Communications of the ACM - Emergency response information systems: emerging trends and technologies CACM Homepage table of contents archive, Volume 50 Issue 3, March 2007.

[19] Okolloh, O., Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information, Participatory Learning and Action, 2009 .

[20] Careem, M., Silva, C. D., Silva, R. D., Raschid, L., and Weerawarana, S. (2006) "Sahana: Overview of a disaster management system. In IEEE International Conference on Information and Automation," 2006.

[21] SwiftRiver Website: http://swift.ushahidi.com/

[22] Whostalkin: http://www.whostalkin.com/

[23] Socialmention: http://socialmention.com/

[24] EM-DAT: http://www.emdat.be/

Authors' Profiles

Juan Li received a B.S. degree from Beijing Jiaotong University, Beijing, China, in July 1997, and a Ph.D. degree from the University of British Columbia, Vancouver, Canada, in May 2008. Currently, she is an Associate Professor of Computer Science Department at the North Dakota State University, Fargo, ND, USA. Dr. Li's major research interest lies in distributed systems, including P2P networks, grid and cloud computing, mobile ad hoc network, social networking, and semantic web technologies.

Qingrui Li received a B.S. degree from Zhejiang University, Hangzhou, China, in 2002, and a M.S. degree from the National University of Singapore, Singapore, in 2004. Currently, he is a PhD student of Computer Science Department at the North Dakota State University, Fargo, ND, USA. His current research focuses on intelligent systems.

Nazia Zanman received a B.S. degree from the University of Dhaka, Dhaka, Bangladesh, in 2007, and a M.S. degree from the same university, in 2009. Currently, she is a PhD student of Computer Science Department at the North Dakota State University, Fargo, ND, USA. Her current research focuses on intelligent systems, social networking, and natural language processing.

Ashok Ginjala received his Master degree from the North Dakota State University, Fargo, ND, USA in 2013. His research focused on social media search.