International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.1 (2014), pp.255-264
http://dx.doi.org/10.14257/ijmue.2014.9.1.24

Development of a Disaster Information Extraction System based on Social Network Services

Younghak Chun[1], Hyun Suk Hwang[2] and Chang Soo Kim[3]

[1,3] *Pukyong Natinal University, Dept. of IT Convergence and Application Engineering*
[2]*Pukyong National University, Interdisciplinary Program of Information Systems,*
45, Yongso-ro, Nam-Gu, 608-737, Busan, South Korea

{chun0831[1], hhs[2], cskim[3]}@pknu.ac.kr

Abstract

The increased use of Social Network Services has led to an increase of availability of information. In addition, information found within SNS posts can be useful for victims in disaster-affected areas. However, a method for retrieving disaster-related information has not yet been created, thus making it difficult to find disaster-related information in SNS posts. Accordingly, this paper suggests a system which can provide efficient information when disasters occur. This system calculates three attributes of disasters. First, the Filtering module can search disaster related tweets. Moreover, it can extract the attributes of disasters systematically. Second, a manager can determine the attributes of a disaster manually by using the Decision module. In this module, disaster type, location and risk level attributes are determined. Once all of the attributes are determined, the Display module shows the attributes of the disaster. This system will provide citizens with more up-to-date disaster information.

Keywords: *Disaster-Related Information, SNS posts, Filtering*

1. Introduction

In the past, the only information made public regarding disasters was provided by government authorities. However, presently, citizens are able to share their information via social network services (SNS). There exists a great deal of information in SNS posts, related to such topics as fashion, the economy, the stock market, surveys, disasters and a variety of other topics . We can obtain a wealth of information via SNS. In fact, many researchers have attempted to analyze SNS to use.

However, retrieving information from SNS posts is not without fault. Not all posts are grammatically correct, nor do they contain accurate information [9]. Accordingly, if we could implement a system to recognize the disaster-related related information within SNS posts, we could systematically obtain important attributes of disasters. In particular, South Korea needs this system. There is a high amount of rainfall in South Korea in the many mountainous regions. Therefore, flooding occurs frequently. If citizens living in the disaster-affected areas were to take photographs and share these via SNS, other citizens could be alerted to the dangers in this area.

Accordingly, we designed this system to be of help to citizens in disaster situations. We will implement three attributes such as the type of disaster, the location and the risk level by using several algorithms.

ISSN: 1975-0080 IJMUE

2. Related Works

2.1. Second-order headings

Social Network Services are online services or sites that facilitate social relations among people who want to share interests, activities, or real-life stories. To analyze the usefulness of SNS in other fields, some methods have been researched as to how to filter messages. SNS messages are categorized into three types of messages; text message, text message with a photograph, and text message with a location. Uploaded data must be reliable to be utilized for analyzing disasters.

Hong *et al.*, [3] proposed the method of separating SNS data, by 'Word removal', 'Sliding windows', 'Including spaces', and a technique of 'Selection'. The methods were 70.44% accurate in filtering results.

There are four limitations when we upload messages on Twitter. First, text messages are limited to 140 characters. Next, abbreviations or internet neologisms are often used in text messages. Third, text messages can contain URL sites, such as photos or videos. Finally, there are many spelling errors in text messages [9, 4].

As Social Network Services (SNS) have continued to advance, so has the exchange of information. Several researchers are studying SNS in order to extract information in several fields.

Jung Me-Ae *et al.*, [5] researched how to activate real estate marketing using SNS and internet. They argued that existing internet marketing should focus on the trend of SNS in order to engage potential customers.

In Australia, the 'Emergency 2.0 Australia' system has been implemented to provide disaster-related information to citizens in real-time using Twitter and Facebook. This system serves a similar purpose as SNS to convey disaster-related information, but it runs only with the exclusive application [1].

Several related works have researched the collection of SNS posts for sharing disaster-related information when disasters occurred. Individuals living in the same disaster-affected region can share information to prevent damage and injury [2, 6].

To display the location of the disasters, we refer to related research. Rui Li *et al.*, [7] researched the method of creating a buffer for areas affected by disasters. They analyzed tweets and accumulated them to set the buffer that designates the disaster area. This method was used to make a pattern for disaster occurrence. Furthermore, it is also important to filter Twitter effectively. Hurlock *et al.*, [8] researched the method of creating a semantic result by analyzing statistics collected from similar tweets.

2.2. Second-order headings

In this paper, the system of determination of disaster information and attributes utilized is Open APIs. Libraries are utilized to determine efficient functions of existing contents. Likewise, we improved the system by convergence of Twitter and Google Map API. The system implemented by .NET programming language refers the libraries to adapt the APIs into the program. We will also explain the related libraries.

First, the Twitter API provides the Timeline module which updates the tweets in real-time, and the filtering module filters the results by text that the users want [10].

However, we cannot apply them into the .NET program because of the difference of language type. Therefore, we need to use the library to get the parameters into the .NET program. We use the 'Twitterizer 2.0' open library. When we get tweets with Timeline, we

can access them by Twitter API in javascript, and we can convert the values into the .NET program by 'Twitterizer 2.0' [11].

In addition, Google Map API depicts maps of disaster-affected areas. Google Map API can provide the world map to individual developers. Likewise, we used the marker to pinpoint the exact place of the disaster. However, dangerous areas are not represented by points, but by polygon types. Hence, we used the other function of Google Map API, to draw polygons. Furthermore, we can also use Google Map API to get the coordinates by geocoding API [12]. If we can find an address in a tweet, we can use it as an input parameter to be replaced by the coordinates.

Finally, the EXIF Library is used to extract values or information from the photo. The EXIF is a tag that includes a great deal of information about the photo such as the date, the name of the camera, and the GPS coordinates [13].

3. System Design

The whole construction of this system is shown in Figure 1.

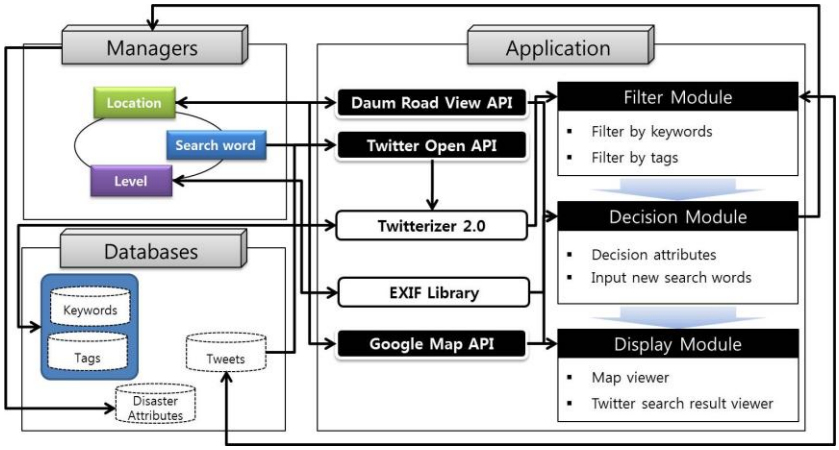


Figure 1. Components of this system

This system is implemented to extract disaster attributes from tweets by mash-up programming. The mash-up is a creation of a new service using existing open API or contents. We have implemented a new system to extract disaster attributes using Twitter and Google map API, and some open libraries to use efficient functions of existing contents.

3.1. Managers

The message in a tweet may be ambiguous. It may not include all attributes of the disasters. Most of tweets do not have GPS coordinates or a photo attached. This is why we need managers to verify and set the missing attributes.

Managers can determine the location of a tweet by using Daum Road View. Accordingly, we used Daum Road View API. If a tweet has no GPS coordinates, the manager should find the location by exploring the Road view. In the case identifying the risk level, the manager can also find this information by searching included photos or messages in the tweet.

The last role of the manager is to select a new tag. The tags are selected systematically. However, it is possible that an error could occur. For more efficient filtering result, the manager can input the new tag based on his or her observations.

3.2. Application

Users can see the application when they use this system. It has three modules to extract disaster information: filter, decision and display.

The filter module can be used when the user would like to filter the tweets according to disaster-related words. There are two ways of filtering words and two ways of filtering boundaries. By offering different ways of filtering, the user is guaranteed to receive results.

The application retrieves the keyword and tag from the database using the 'Twitterizer 2.0' library. Then, the filter module matches the tweets with the filter words. When the filtering is finished, this module can provide filtered tweets. These results are transferred to the decision module. The decision module can help the manager of this system to determine the attributes of the disasters. The manager refers to the Daum Road view API or the photo attached to the tweet. If the tweet includes a photo, the manager can verify the location or the risk level of the disaster. If the tweet includes an address of the disaster location, it could be translated into coordinates by geocoding. To perform this, the system utilized is Google Map Open API. It autonomously provides a function of geocoding. In particular, if the tweet includes a photo, the manager can extract the coordinates by using the EXIF library, so as to determine the location of the disaster.

When all of the attributes have been determined, the values are stored in the databases for tags and keywords by the manager.

Once all of the disaster attributes are determined, this system will prepare the display module. This module consists of map viewer and twitter viewer parts. For the map viewer, this application utilizes the Google Map Open API which provides a general map and tools .It also provides the exact point of location. When the location is determined by the decision module, this system reads the coordinates and displays them with a marker.

In addition, this system utilizes the Twitter Open API to display general tweets and disaster-related tweets. The data from tweets is extracted from the database. This system should display tweets three times: general tweets, filter-resulted tweets, and attributes-extracted tweets.

3.3. Databases Schema

Databases are built for storing several important things. These are shown by Figure 2. First, the Keyword table is used for filtering Twitter by keywords. Each table includes one keyword, and has two properties to separate the keywords. Likewise, the Tags table also has number and name properties but it has one more important property to count frequency of the tag. It is important when the tag is selected to filter using the filtering module. Because tags are spoken language, it can be difficult to select the proper tags to filter tweets. Therefore, we filtered fifty disaster-related tweets and counted ten words of high frequency. For this work, the searched words in the Filter_Result were read. In the tag table, there are many possible tags. The most frequently used words would be selected as the tags to filter Twitter.

The Filter_Result table includes results of the filtering module. It has several attributes of resulted tweets, and words searched in the tweets. More attributes are needed to utilize tweets: the SRnumber, the user_ID, the upload_time, the Contents etc. The 'upload_time' attribute is for filtering by date or time. The searched words from the tweets are stored in the Contents property as 'Text' data type. They can also use to record history of disaster occurrence. This history can be used to analyze and make predictions about future disasters.

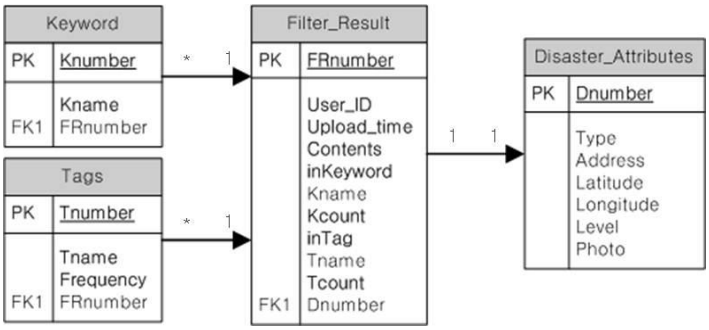


Figure 2. Entity Relation Diagram

In this system, attributes are determined after several stages. If the system identifies attributes of disasters, they are stored in the 'Disasters'_Attributes' table. These are the final results of this system. Properties of this table would have no value when the filtering module is finished. Then, the manager can fill with values in the Decision model of this system.

We can determine and store three attributes of disasters: the type of disaster, the risk level, and the location. The Address, the Latitude and the Longitude attributes are to set the location. If a tweet includes photos of the disaster, they can be stored in the Photo property as 'BLOB' data type. The Photo property is to assist the manager in finding the location or determining the risk level. The database of the disaster point is to store the output of this system. From this table, the system could extract attributes of disasters from Twitter systematically and manually. The output from this system is important information, and we will be able to use this information when making predictions regarding future disasters. In addition, we will use these databases for providing information such as detours, calculating flood risk index, etc.

4. Scenario of the System

Figure 3 shows the main dialog of this system. It consists of Google map viewer, Twitter viewer and result viewer. For convenience, Twitter viewer is shown in a separate tab.

Figure 3. The capture image of main page

The first tab contains general tweets. The second tab contains disaster-related information via SNS as the result of this system. The first time operating this system, the second tab may be empty because it does not depict the result.

This system has three menus. These are the Filter, the DecisionModel and the Preference. The Preference menu has three sub-menus: the Filtering Condition, Keyword and Tag.

4.1. Preference Menu

Figure 4 shows preferences of some functions. It is requested before running this system to get efficient results.

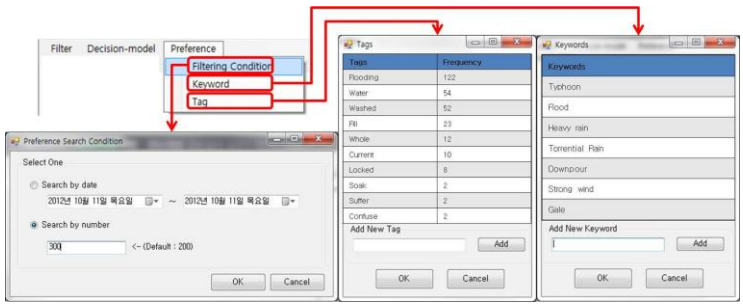


Figure 4. The configurations of this system

First of all, users are requested to set the environment of several fields. The Preference menu has three sub-menus. First, the range of the filtering module can be set for filtering in the Filtering condition menu. Users are provided with two conditions, filter by date or number. The default condition is filtering by 200 numbers.

Second, the Keyword preference menu provides the function to manage the keywords to filter. If necessary, the user can change or add the keyword. The Tag preference menu is similar with the Keyword preference.

4.2. Filtering Menu

The Filtering menu is the start point of this system. It is shown by Figure 5.

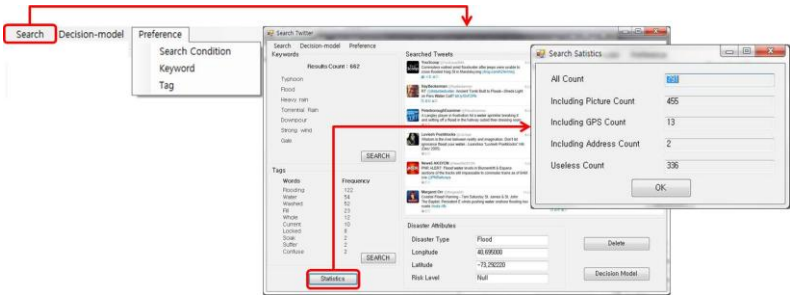


Figure 5. Filtering results and statistics

In the filtering page, the first step is keyword filter. When the system performs the keyword filter, it reads a tweet message as a text file in the database. After this step, it saves the filter results. The next step is the tag filter. In this module, it reads the filter results and stores the tags. It selects tags by their frequency.

As a result, the related tweets are extracted by a two-way filter method. When the filtering is finished, users can read the statistics of the filtering results by pressing the 'Statistics' button. If the button is clicked, details of the results are shown in a new dialog box: All count, Including Picture Count, Including GPS Count, Including Address Count and Useless Tweet Count. At the last step, the Determine Attribute Module can determine the attributes of the disasters. The module reads the database to check whether the attributes have been determined or not. If some of them are not determined, this module can extract them and store

determined attributes of the disasters. Finally, the system can filter Twitter, and extract and determine attributes from them.

4.3. DecisionModel Menu

This system needs a 'DecisionModel' menu to determine undecided attributes. Figure 6 shows the model for a manager of this system. The manager determines the unspecified values of disaster attributes by assessing photos or the Daum Road View.

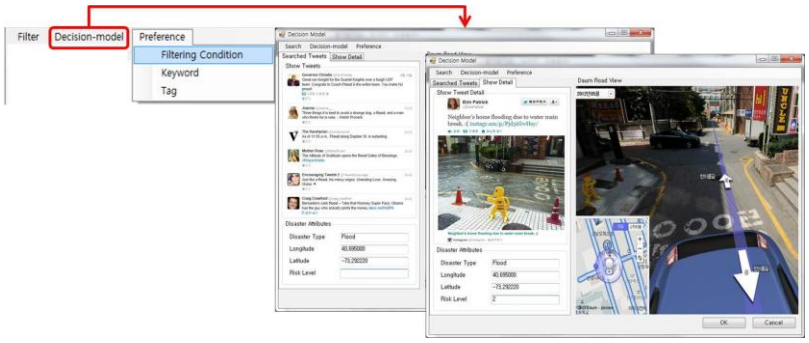


Figure 6. Determination of attributes using DecisionModel

When the filtering module is finished, it can include tweets which have no value for their attributes. The manager can find the unknown value by using the DecisionModel. In the manager mode, this system shows the details and photos attached to a tweet. Then, this module provides a Road View from Daum API. It can assist the manager in determining the location or level of a disaster.

When all of the processes are done, this system shows all of the filtered results and each attribute on the Google map.

5. Conclusion

The model for road inundation risk index decides a value which represents a risk level. When disasters occur, users can assess the situation in advance and make informed decisions regarding their actions. However, existing models lack practicality because they depend only on historical data.

When disasters occur, users need real-time data. Therefore, we have suggested methods to obtain real-time data. In this study, we suggested using SNS messages to retrieve disaster-related information. To filter Twitter for disasters, we built a system and filtered Twitter messages using keywords and tags in order to obtain information. We also suggested the method to extract disasters attributes: the type of disaster, the location and the risk level. The system utilized messages, GPS coordinates and photos from tweets to determine these attributes.

Through this system, we can determine all of the attributes of disasters, and users can see what kind of disasters it is, where the disaster is occurring, and how dangerous it is.

In the future studies, this system will be improved by introducing possible routes for detours and by providing information regarding shelters for victims in the disaster-affected areas.

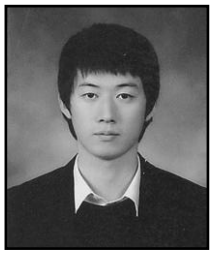
Acknowledgements

This research was supported by a grant 'A Development of the Safety Route System in Inundation Area based on Smartphone' [NEMA-NH-C-D-2012-0243] from the National Hazard Mitigation Research Group, National Emergency Management Agency of Korea.

References

[1] J. -H. Lee, "Social Media is Perfect to Disasters Information - Management of Disasters Situation in Australia, Spread of Disasters Information in Japan", Disaster Focus Magazine, **(2010)** April.
[2] K. Tatsuya, "The Development for partnership SNS System between local disasters prevention using iPhone", The Hyper Institute of Network Society, Local Information Magazine, vol. 68, **(2011)** May-June, pp. 56-63.
[3] C. H. Hong and H. S. Kim, "Effective Feature Extraction for Twits Classification", Korea Computer Congress, vol. 38, no. 1(A), **(2011)**, pp. 229-232.
[4] W. B. Yoon, W. Yu and H. S. Lim, "Text analysis of limited length SNS", Korean Society for Internet Information, vol. 12, no. 1, **(2011)**, pp. 141-142.
[5] M. -A. Jung and D. -H. Kim, "The Method to Activate Estate Marketing using Internet and SNS", Korea Institute of Electronic Communication Societies, vol. 6, no. 1, **(2012)**, pp. 121-126.
[6] K. Kireyev, L. Palen and K. M. Anderson, "Applications of Topics Models to Analysis of Disasters-Related Twitter Data", Neural Information Processing Systems Foundation Workshop, Seattle, WA, **(2009)**.
[7] R. Li, K. H. Lei, R. Khadiwala and K. C. -C. Chang, "TEDAS: a Twitter Based Event Detection and Analysis System", 2012 IEEE 28th International Conference on Data Engineering, **(2012)**.
[8] J. Hurlock and M. L. Wilson, "Searching Twitter: Separating the Tweet from Chaff", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, **(2011)**.
[9] T. W. Seo, "A Study of Real-Time Disasters Information Extraction and Display using the Mash-up based on SNS - using the Twitter API and Google map API -", Master's Thesis, Pukyong National University, Busan, Korea, **(2012)**.
[10] Twitter Developers Page, https://dev.twitter.com/.
[11] Twitterizer Development Portal, http://pm.twitterizer.net/.
[12] Google Maps API - Google Developers, https://developers.google.com/maps/?hl=ko.
[13] The libexif C EXIF Library Page, http://libexif.sourceforge.net/.

Authors



Younghak Chun

He received the B.S. degree in IT Convergence and Application Engineering from Pukyong National University in 2011. He is in master course in from Pukyong National University. His current research interests are disaster prevention, gathering disasters information, GIS and u-prevention system.

Hyun-Suk Hwang

She received the M.S degree in Computer Science and the Ph.D. degree in Management Information System from Pukyong National University of Korea in 1995 and 2001 respectively. She worked the program of Post-Doc. At Kyungname University from 2001 to 2002 and at University of Missouri, Kansas City(UMKC) from 2003 to 2004. Her current research interests are in the area of LBS/GIS, ontology, WSN, and u-prevention system.

Chang Soo Kim

He received a B.S degree in Computer Science from Ulsan University, Korea, in 1979, and a M.S. degree in Computer Engineering and Ph.D. degree in Computer Engineering from Chungang University, Korea, in 1984 and 1991 respectively. He has been a professor at the department of IT Convergence and Application Engineering, Pukyong National University, Korea, since 1992. His research interests are operation system, LBS/GIS, WSN and urban disaster prevention system.