The Method to Extract Disaster Information Factors from SNS for the Flood Risk Index Model

Younghak Chun', Hyun Suk Hwang[2] ,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. Social network services (SNS) are widely used to share information for several purposes. Specifically, as the disaster information from SNS posts circulates rapidly, people can prepare for the coming damage. We propose a method to extract the disaster information from SNS posts in real-time. This study aims to systematically classify related keywords regarding disasters, and identify the locations and the danger levels from SNS posts. The transferred information will be used to calculate a flood risk index, which identifies dangerous levels of flooded roads.

Keywords: Flood Risk Index, SNS posts, Keyword and Tag Search

1 Introduction

Geographically, South Korea is a mountainous region, and therefore receives a great deal of rain every year. Expansive urban floods caused by downpours have damaged people's lives and estates due to flooded roads and houses in which lack damage-prevention facilities.

In Seoul in July of 2011, a record-breaking rainfall caused flooded roads in several places and drivers who were unaware of the situation were isolated. If the isolated people had known that there was an urgent real-time situation, they would have been able to avoid the disasters safely. The Social Network Services (SNS) can be an extremely important type of media for utilizing and sharing real-time content. SNS can be used as a resource for providing real-time messages via on-line communication [1]. In particular, with increasing use of smart phones and tablet PCs, the number of uploaded posts has increased as well. According to the Gruter Institute in March 2012, an average of 453 posts are uploaded a day on Twitter in Korea. Therefore, we propose a method for extracting SNS factors as input parameters to predict risk levels of flooded areas.

[3] Corresponding Author

2 Relative Works

2.1 SNS data

Social Network Services are online services or sites that facilitate social relations among people who want to share interests, activities, or real-life stories. Some methods have researched how to filter interesting messages on SNS, which are categorized into three types of messages, such as only a text message, pictures added to a text message, and locations added to a text message. Uploaded data must be reliable to be utilized for analyzing disasters.

Seo [1] studied a method which is six times more effective than general methods. Hong et al. [5] proposed the method of separating SNS data, such as the removal of stop words, sliding windows, inclusion of spaces, and a technique of selection. The methods were accurate up to at least 70.44% of search 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 [4]. Finally, there are many spelling errors in text messages [6][7].

2.2 Flood Risk Index

The Flood Risk Index (FRI) is defined as a quantified value to measure the risk of each area damaged by floods. Most of the researchers presented models including factors of rainfall, base water levels, altitudes and disaster history data. However, these models cannot provide results extracted from real-time situations. In an urgent situation, the real-time information considering social factors will help civilians to avoid damage.

3 An Extraction Method for SNS Factors 3.1

Keyword and Tag Search

We use the keyword and tag search to assign posts to a disaster code. The keyword search means that users execute the general search with keywords on Internet. The National Emergency Management Agency (NEMA) categorized related keywords with a disaster code using classified words regarding natural disasters, which were defined in South Korea. The tag words are the same as related words with similar meanings to keywords which are generally used a lot on Twitter. To find tag words, we searched 50 SNS posts and obtained several words related to floods, heavy rains and typhoons. Tag words are refreshed before a search is processed. Several words regarding the keyword and tag words are shown in Table 1.

Table 1. Keywords and tag words

Disaster Code	Keyword	Tag word
Flood	River flood, Unexpected flood, Coast flood, Urban flood	Water, Whole place, Flooding, locked, Currents, Water fills, Soak, Suffer, Confusion
Heavy Rain	Severe rain storm, General Heavy Rain	Storm, Noisy, Crash, Tsunami, wave, Sea wave
Typhoon	Mega typhoon, Large typhoon, Windstorm, Storm wind, Wild Middle typhoon, Small typhoon wind, Gale, Tempest, Warning	

3.2 Process Flow

This study aims to make a module to extract values parameters utilized as the Flood Risk Index from SNS data. The SNS factors consist of three variables: the disaster code, the location, and the danger level. Therefore, SNS posts may maintain a pattern according to disaster codes in a FRI model.

This system is divided into three parts. The first part is to identity disaster codes by classifying SNS posts into defined disaster codes. The second part is to decide a location of each post which originated from SNS. The last part requires government authorities to identify a location with attached photos and identify a danger level by verifying filtered posts.

Figure 1 shows the entire process flow to extract the above-mentioned three parameters. Specifically, the process begins when an individual searches for a keyword with a search module. The search results are parsed and stored in temporary storage. Next, the parsed words are stored in a database including keywords and tag words. , Following this, a code related to the keyword is set as a value of the code parameter. Otherwise, the disaster code will be assigned by means of the tag search method. If the keyword is in tag the tag candidate database, this system adds the value **1** to a tag count variable. The tag words are stored into the database including the tag words if the count value is more than the threshold value, which is defined by government authorities.

In addition, the system examines the GPS data or photos to find locations. If posts include locations from GPS, the coordinates can be assigned. Otherwise, the system scans posts and finds an address. It will then obtain the x and y coordinates by geocoding. If posts do not have GPS and address data, government authorities can decide the location with attached photos in a verification step. It assumes that a location cannot be assigned without photos. In the case of the danger level, government authorities can determine a level according to how dangerous the situation is in the verification step. The street view matching model is to help government authorities to determine a location with photos and street views. Moreover, a decision model helps government authorities to determine the danger level using rainfalls and tag codes.

For example, if the amount of rainfall is more than 70 mm in 6 hours or 110 mm in 12 hours, a flood warning is issued. Specifically, if rainfall is more than 110 mm in 6 hours or 180 mm in 12 hours, a flood watch is issued.

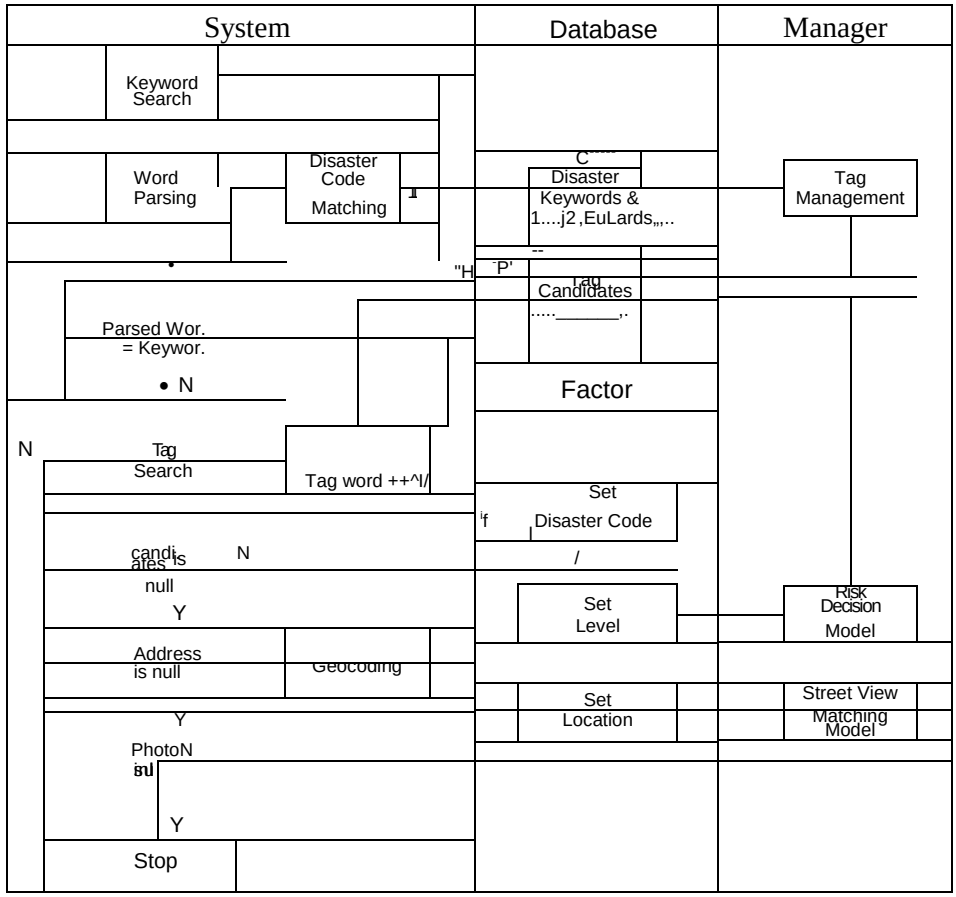

Fig. 1. Process flow

4 Conclusion

The flood risk index represents a risk level using a model with five factors, which are SNS, CCTV, USN, rainfall amounts and history data. We propose a method to extract disaster information from SNS posts in real-time. The model attains three parameters: the disaster code, location and danger level as determined by SNS posts. The keyword and tag search results are extracted by comparing defined disaster keywords and tag words to extract these parameters. The location is assigned using attached photos, and government authorities can determine a danger level using a risk-decision model.

Our work will be utilized as the social factor of the FRI model to measure risks of damaged areas. We will implement the proposed method after analyzing and

categorizing SNS posts. In addition, we will make an advanced engine by using the ontology and semantic-based search algorithm.

Acknowledgments. This research was supported by R&D program through the Natural Hazard Mitigation Research Group (NHMRG) funded by the National Emergency Management Agency (NEMA) (No. C-D-2012-0243).

References

1. Tae Woong Seo : A Study of Real-Time Disaster 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)
2. Ki-Uk Kim, Chang-Soo Kim : Design of a Data Model for the Rainfall-Runoff Simulation Based on Spatial Database : Korean Association of Geographic Information Studies, Vol. 13 No. 4, pp.1--11 (2010)
3. Ki-Uk Kim, Sung-Woo Park, Hyun-Suk Hwang, Chang-Soo Kim : The Modeling of Real-time Forecasting System for Urban Flood connected with UIS : Korea Multimedia Society, No.13 Vol. 2, pp. 223--226 (2006)
4. Dae-Yeon Won, Ki-Jung Park, Young-Jun Park, Gyu-Bae Shim, Jae-Woong Lee, Yong-Hyuk Kim : Spam Twit Filtering using Naïve Bayesian Algorithm and URL Analysis : The Korean Institute of Information Scientists and Engineers, Vol. 38 No.2(B), pp.375--378 (2011)
5. ChoHee Hong, HarkSoo Kim : Effective Feature Extraction for Twits Classification : Korea Computer Congress, Vol.38, No.1(A), pp.229--232 (2011)
6. Myung-Gwon Hwang, Dong-Jin Choi, Pan-Koo Kim : A Context Information Extraction Method according to Subject for Semantic Text Processing : Korean Institute of Information Technology, Vol.8, No.11, pp.197--204 (2010)
7. Bo Hyun Yun : Automatic Information Extraction for Structured Web Documents : Korean Society for Internet Information, Vol.6 No.3, pp.129--145 (2005)
8. WonBeom Yoon, Wonhee Yu, HeuiSeok Lim : Text analysis of limited length SNS : Korean Society for Internet Information, Vol.12, No.1, pp.141--142 (2011)
9. Donsu Lee, Eunhye Kim, Jongyeon Park, Sangjun Lee : Design and Implementation of Social Network Service based on Location using Open API: Korea Computer Congress, Vol.38, No.1(C) pp.60--63 (2011)
10. Soo-Young Park, Yong-Ho Ha, Yong-Hyuk Kim : Recent Studies on Twitter in the Field of Information Retrieval: The Korean Institute of Information Scientists and Engineers, Vol.37 No.2(C), pp.25--29 (2010)
11. Kiruchi Tatsuya : The Development for partnership SNS System between local disaster prevention using iPhone: The Hyper Institute of Network Society, Local Information Magazine, May+June Vol.68, pp. 56--63 (2011)
12. K. Kireyev, L. Palen, K. M. Anderson : Applications of Topics Models to Analysis of Disaster-Related Twitter Data : Neural Information Processing Systems Foundation Workshop, Seattle. WA. (2009)