SNS-based Issue Detection and Related News Summarization Scheme

Daeyong Kim
School of EE
Korea University
Seoul, Korea
+82-2-3290-3784
ritgd05@korea.ac.kr

Daehoon Kim
School of EE
Korea University
Seoul, Korea
+82-2-3290-3784
kdh812@korea.ac.kr

Siwan Kim
School of EE
Korea University
Seoul, Korea
+82-2-3290-3784
duo0701@korea.ac.kr

Minho Jo
Dept. of Comp.& Comm.
Korea University
Seoul, Korea
+82-2-3290-4938
minhojo@korea.ac.kr

Eenjun Hwang
School of EE
Korea University
Seoul, Korea
+82-2-3290-3784
ehwang04@korea.ac.kr

ABSTRACT
Due to the unprecedented popularity of social network services (SNSs), such as Twitter and Facebook, means that a huge number of user documents are created and shared constantly via SNSs. Given the volume of user documents, browsing documents in a selective manner based on personal interests is a time-consuming and laborious task. Therefore, in the case of Twitter, trend keyword lists are provided for the user's convenience. However, it is still not easy to determine the details based on a few simple keywords. The keywords usually relate to the hot issues at any time so many documents will contain pertinent details, such as news on the Internet. Thus, to provide detailed information about an issue, it is necessary to identify relationships among them. In this study, we developed a SNS-based issue detection and related news summarization scheme. To evaluate the effectiveness of our scheme, we implemented a prototype system and performed various experiments. We present some of the results.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing –*Abstracting methods*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Information filtering, Search process*

General Terms
Algorithms, Design, Experimentation

Keywords
Issue summarization, News summarization, News trend, SNS analysis, Trending issue

IMCOM (ICUIMC)'14, January 9–11, 2014, Siem Reap, Cambodia.

1. INTRODUCTION
Recently, the world has witnessed many dramatic changes in term of IT technology. In particular, the emergence of various social network services (SNSs) such as Facebook and Twitter has been rapid. Twitter is an online social networking and service, which allows its users to write and share text messages known as "tweets." It has gained worldwide popularity, with more than 500 million active users (as of 2012) who generate over 340 million tweets daily, and it handles in excess of 1.6 billion search queries per day. Owing to this popularity, it has been described as "the SMS of the Internet." Unregistered users can read tweets, but registered users can post tweets via the website interface or via a range of apps for mobile devices [1]. In Twitter, a user can acquire trend keywords for a particular location via the trends delivery service. This service provides a list of the top ten trend keywords based on an analysis of recently posted tweets. However, using the keyword list alone, it is not easy to grasp the details or subjects indicated by the keywords. Moreover, a keyword could be related to multiple subjects or issues and users might not know the issues that have been mentioned in tweets at the time. For example, two hot issues related to a recent trend keyword, "USA," were "Obamacare" and "sending troops to Syria." However, based on this keyword, it is almost impossible for users to understand why it became a keyword or to identify the related issues. Thus, to provide detailed information about trending keywords, we propose a SNS-based issue detection and related news summarization scheme. In this scheme, we extract trend keywords from the Twitter data stream and collect tweets containing these keywords. Next, we extract trending issues from these tweets. In this case, the issues for a trend keyword represent subjects that are closed related to the keyword. For each issue, we produce a summary of related tweets and select the single most representative tweet. Finally, for each issue, we collect related news from the Google site and summarize them. To evaluate the effectiveness of our scheme, we implemented a prototype system and performed various experiments. We present some of the results.

The remainder of the paper is organized as follows. In Section 2, we discuss the background of this study and some related work. In

Section 3, we provide a detailed description of our scheme for detecting issues and summarizing related news. Section 4 presents our experimental results. Our brief conclusions are given in Section 5.

2. RELATED WORKS

Trend detection needs to be performed accurately and effectively to develop trend-based applications or services. A typical service could be a trending news summary service, which provides a summarized version of specific issue-related news.

2.1 SNS Trend Analysis

SNS messages have different characteristics compared with traditional documents, such as books or essays. For example, typical SNS messages such as tweets are usually short but the total volume of messages generated in real-time is enormous. Thus, trend keyword detection should be applied to use SNS messages in an effective manner. Many previous studies have addressed trend detection.

Latent Dirichlet allocation (LDA) [2], a generative probabilistic model of collections of discrete data such as text corpora, has been shown to deliver good performance during general text mining. Many studies have used LDA to handle documents. Yokomoto et al. applied an LDA-based document model to the task of labeling blog posts with the Wikipedia entries they collected [3]. Their LDA parameters depended on the distribution of keywords among all of the search results for blog posts. Ramage et al. presented a scalable implementation of a partially supervised learning model, which mapped the content of Twitter feeds into dimensions [4]. They characterized users and tweets using this model, and presented their results from two information consumption-oriented tasks. Several studies have also treated SNS message as streaming data. Allan et al. proposed an online news event detection and tracking method [5], and Yang et al. investigated the use and extension of text retrieval and clustering techniques for event detection [6].

Recently, several methods have been proposed for analyzing trends in real time. Mathioudakis et al. proposed TwitterMonitor for detecting trends in Twitter streams [7]. TwitterMonitor identifies emerging topics on Twitter in real time and then provides meaningful analytics to facilitate the accurate description of each topic. This method discovers topic trends by detecting bursty single tags. However, it is difficult to obtain information about various events using only single tags. Thus, Alvanaki et al. presented the enBlogue system, which is an approach that automatically detects emergent topics based on shifts in tag correlations as they arise dynamically [8][9]. Other studies have addressed the effective detection of trends. Kumaran et al. proposed a text classification method for detecting new events [10]. To improve the detection performance, they used text classification techniques and named entities, a single pass clustering algorithm, and a threshold model, which incorporated the properties of events as a major component. Sayyadi et al. built a network of keywords based on their co-occurrence in documents [11] and they proposed a new event detection algorithm, which creates a keyword graph and uses community detection methods to discover and describe events.

However, these methods usually require considerable amounts of time for trend analysis, which means that they cannot be used as real-time applications. Another problem is that the result is simply a list of keywords. Thus, it is very difficult for the user to know why the keywords have become an issue or what the hot issues are.

2.2 News Analysis

News analysis is an essential component of the field of news recommendation or news summaries, and many studies have been conducted in this area. For example, Liu et al. proposed a hybrid system based on a user's Web history to provide personalized news recommendations [12]. They conducted a large-scale analysis of anonymous user click logs and predicted the current news interests of users with a Bayesian framework for log analysis. They also combined a content-based recommendation mechanism that employed user profiles with collaborative filtering to generate personalized news recommendation. Francisci et al. proposed a Web-based personalized news recommendation system [13], which used the learning-to-rank approach and support vector machines to rank news interests using Twitter logs and to generate user profiles. Their system could also predict the degree of interest of news based on the profiles generated and the social neighborhood of users. Phelan et al. developed a social news service called Buzzer, which ranked personal RSS subscriptions based on Twitter conversations [14] by using a content-based approach to mine trending terms from the Twitter timeline and friend subscriptions. Barzilay et al. proposed a sentence fusion method for summarizing news in multiple documents [15]. They used a novel text-to-text generation technique, which involved bottom-up local multi-sequence alignment to identify phrases that convey similar information and statistical generation to combine common phrases into a sentence to synthesize common information across documents. Huang et al. proposed a method for summarizing news article by highlighting the commonalities and differences in comparable topics [16]. They focused on the topic-related concept and calculated the weight of evidence in the summary of a cross-topic article to formalize the summarization task as an optimization problem. Gong et al. applied two generic text summarization methods, i.e., the standard IR method and latent semantic analysis, to rank the sentence relevance and to generate a summary [17], where both methods selected sentences that were highly ranked and different from each other to create a summary with a wider coverage of the document's main content and less redundancy.

3. SYSTEM ARCHITECTURE

Figure 1 shows the overall architecture of our proposed system. The system has four major components: trend analyzer, issue detector, issue summarizer, and news summarizer. The trend analyzer extracts trend keywords and semantically related keywords. The issue detector collects trend-related tweets and extracts issues based on the co-occurrence of candidate nouns. Next, the issue summarizer collects candidate words for issue summarization and selects the most representative tweet. Finally, the news summarizer finds issue-related news articles and extracts their core sentences for news summarization.

3.1 Trend Analyzer

In our previous study [18][19], we developed a trend analyzer to detect trend keywords in near real-time using the characteristics of twitter stream data. First, we select candidate keywords from tweets by performing simple syntactic feature–based filtering to select proper nouns based on the fact that most trend keywords either start with a capital letter or are enclosed by a pair of quotation marks. Each trend keyword corresponds to a proper noun and represents a unique entity, which is distinct from a common noun. Next, we merge various keyword variants using several heuristics to handle common input events such as acronyms and typos. The trend analyzer also finds related keyword groups based on the assumption that, if two or more

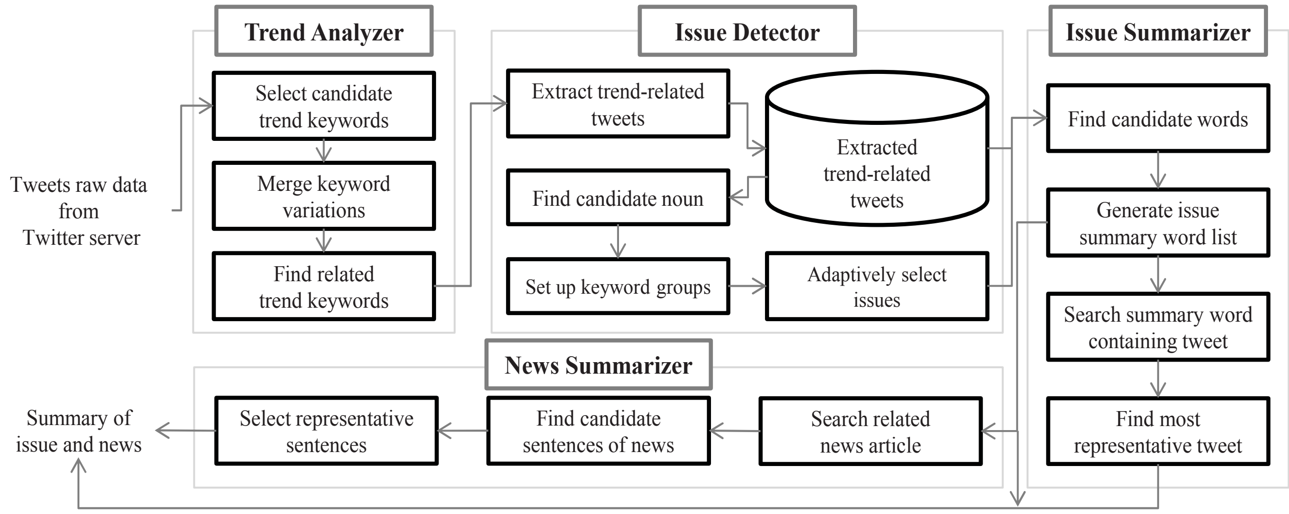

Figure 1. System architecture

keywords appear in the same tweet, then those keywords are probably related to each other. This can be checked easily by generating the tweet ID. Finally, we also expand the trend-related keywords semantically by referring to portal sites such as Wikipedia and Google.

3.2 Issue Detector

This component extracts the trend-related tweets from the tweets collected by the trend analyzer. These tweets are used during issue detection and issue summarization. There are two types of tweets among the trend-related tweets: all the tweets that contain trend keywords and all the tweets that contain variations of the trend keywords.

After collecting the trend-related tweets, the issue detector groups candidate nouns based on their co-occurrence rate in the extracted tweets. We consider each of these groups as an issue. Informally, an issue is a unique, trending subject, which is mentioned frequently by Twitter users. However, a keyword might appear in multiple distinct issues.

An issue can be found using the extracted, trend-related tweets. First, we extract candidate nouns from these tweets. To determine whether a word is noun, we used "Merriam-Webster's Collegiate® Dictionary with Audio," which is provided by Merriam-Webster Developer Center [20]. For example, for the trend keyword "Liam," we extracted its candidate nouns from the tweets between 8:00 AM to 9:00 AM on 13/08/28. Figure 2 shows the results. By chance, this was Liam Payne's birthday and most of the tweets contained congratulations on his birthday. Therefore, there were many birthday-related nouns among the tweets, such as "party" and "teenager."

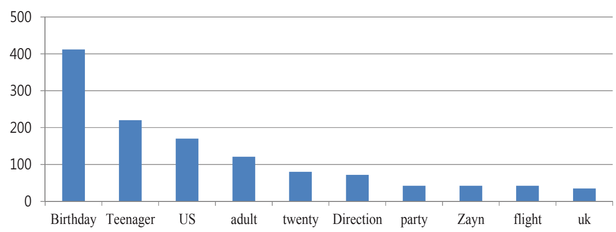

Figure 2. Example of candidate nouns

During candidate word extraction, we do not consider verbs, because verbs are usually common in tweets related to different topics. For example, "Senator Ted Cruz votes no for Obamacare" and "Senators vote for sending troops to Syria" are different topics but they are related to the same trend keyword, "Senator." If we used the co-occurrence of candidate words for grouping, two different topics could be merged as one issue because they both have a word "vote" in the tweets. This type of problem may occur frequently for commonly used verbs, such as "love," "like," and "visit." Therefore, we decided to consider nouns only during issue detection.

After extracting the candidate nouns, we group them based on the co-occurrence rate of the candidate nouns in the extracted tweets. This is based on our assumption that, if nouns appear in the same tweet several times, those nouns are likely to describe the same topic. There are many diverse issues, so it is not beneficial to fix the co-occurrence rate. Fixing the rate would result in the grouping of candidate nouns related to different issues in one issue or separating tweets related to the same issue into different issues. Therefore, we adaptively adjust the threshold of the co-occurrence rates for grouping candidate nouns, as follows. First, we start grouping the candidate nouns with a co-occurrence rate of 10% and increase the rate by 10% until it reaches 100%. If a candidate noun and a noun in the group appear in the same tweet above a certain rate, the candidate noun is considered to be in the same group or issue. We repeat this procedure until all of the candidate nouns are grouped. Eventually, this produces a set of different groups, according to the co-occurrence threshold. For the set of groupings, we determine the best grouping based on the relationship score of each co-occurrence rate's threshold. The relationship score R is calculated using the following equation below.

$$R_\theta = \sum cnt(N_k, T) - \sum cnt((n \in N_k, T) \cap (n \notin N_k, T))$$

In this equation, n indicates a candidate noun and N represents all of the nouns in a specific group. In addition, T indicates the extracted trend-related tweets. Thus, $cnt(W_k, T)$ indicates the number of tweets that contain all the candidate nouns in the k-th group and $\sum cnt((n \in N_k, T) \cap (n \notin N_k, T))$ indicates the number of tweets that contain both the candidate noun in the k-th group and the candidate nouns of different groups. We select the

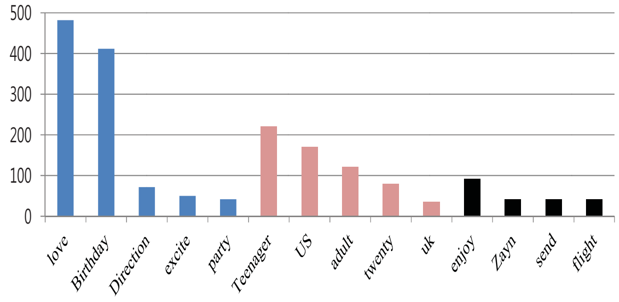

Figure 3. Example of summary word list

threshold with the highest relationship score and use the group as an issue.

3.3 Issue Summarizer

The issue summarizer constructs a list of summary words and a representative tweet for each issue, which were determined in the previous step. The provision of a summary words list was introduced in our previous study [21], where we used nouns for issue detection only. However, we also consider verbs to be candidate words, because it is very difficult to understand the subject of an issue simply by using a list of nouns. To determine whether a word in a tweet was a verb, we used the same dictionary that we used for nouns. Next, we extract verbs as additional candidate keywords based on this determination. For each issue, we construct a summary word list by adding the extracted verbs to the word list related to the issue. If a verb co-occurs in multiple issues, the issue where the verb appears most frequently is selected. These summary word lists are much more effective for understanding the issues in tweets compared with traditional word lists of nouns. Figure 3 shows summary word lists for three issues. In the figure, each group or issue is displayed in a different color, in some cases with verbs. This figure shows that the issues are displayed more clearly compared with the previous example.

In addition to the summary word list, the most representative tweet is selected among the extracted tweets as a supplement. This tweet will help people to understand the subject of an issue more clearly. The most representative tweet of an issue is selected by the score Rt, which is calculated using the following equation:

$$\mathrm{Rt}_k = \sum (cnt(w_i, t_k) * I_{w_i}) + tr_{t_k} + freq(t_k, T)$$

where w indicates a summary word, t indicates a specific tweet in the extracted tweets, tr indicates the number of times the message has been re-tweeted, and I_{w_i} indicates the weight scaled by the frequency of the candidate word w_i in the issue. Thus, $cnt(w_i, t_k)$ indicates the number of summary words that appear in the tweet t_k and $freq(t_k, T)$ indicates how many times the tweet t_k appears in the extracted tweets. The tweet with the highest Rt is selected as the representative tweet. If several tweets have the same Rt value, the most recent tweet is selected as the representative tweet. Figure 4 shows the representative tweets selected for the three issues shown in Figure 3. In contrast to the issue summary word list where some human reasoning is required to infer the subject, users can understand the subject more clearly simply by reading the representative tweet. However, there is a high likelihood that the representative tweet does not cover all of the keywords for the issue. Thus, we provide two forms of issue summary.

3.4 News Summarizer

At this point, we have calculated a summary word list and a representative tweet for each identified issue. Based on this information, we can retrieve more detailed information, such as news articles on the Internet. The news summarizer extracts the core parts of news article, which are considered to be the closest to the issue. First, we search for issue-related news on Google News [22] using the summary word list and we select the most recent. We are now ready for the last step, which is the generation of the news article summary.

For simplicity, we summarize a news article using up to three consecutive sentences from the article. Thus, an article summary could be a single sentence in the article. If this is insufficient, we could use two or three consecutive sentences as its summary. This means that there could be many possibilities, depending on the number of sentences in the article. To determine all of the valid summary candidates, we divide the article into segments, where all of the sentences in each segment contain any summary word or no summary word. Next, we simply consider the former type of segments. To select the best summary among the candidate summaries, we need a metric to evaluate each candidate summary.

Figure 4. Example of representative tweets

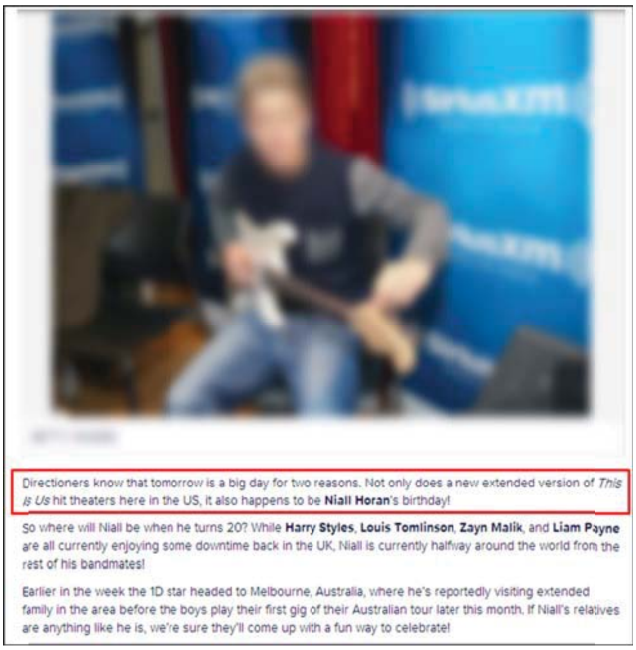

Figure 5. News summary

Table 1. Issue summary result

Trend Keywords by rank	Issue Summary: *Issues*	Issue Summary: *Issue Summary Word List*	Issue Summary: *Representative Tweet*	Trend Keywords by rank	Issue Summary: *Issues*	Issue Summary: *Issue Summary Word List*	Issue Summary: *Representative Tweet*
MARIANO RIVERA	home, Yankee, career	Home, Yankee, Career	'Mariano Rivera at home during Yankee career'	CHRIS PEREZ	Indians, mound, closer	Indians, ruin, mound, leave, closer	'Chris Perez is out to ruin the Indians. No one can …'
	Derek Jetter , Pettitte, game	Derek Jetter, take , Pettitte, game, come, goodbye	'Jeter and Pettitte come to mound to take Mariano...'	MILEY CYRUS	Rolling Stone, Robin Thicke	Rolling Stone, grab, RobinThicke	'Miley Cyrus grabs Robin Thicke at …'
	legend, player, baseball	legend, player, baseball pick, hate	'I hate Yankee, but I would pick Mariano … '		video, song, ratchet, artist	video, song, laugh, ratchet, artist, hate	'Just laughed at Miley Cyrus 23 video but …'
OBAMA CARE	Ted Cruz, Texas, Tea Party	Obamacare, Ted Cruz, stand, defund	'Texas senator Ted Cruz stand against …'		figure, goodness	hot, figure, goodness, feel	'In some pictures Miley Cyrus looks like a …'
	Hillary Clinton, health	Hillary Clinton, Defend, worry, respect, health	'Hillary Clinton defends Obamacare …'	KAYNE WEST	Twitter, Jimmy Kimmel	Twitter, blast, Jimmy Kimmel, mad	'Kanye West Blasts Jimmy Kimmel …'
	spending, money, family	spending, boost, money, family	'Obamacare will boost national health spend…'		person, job, interview, period	person, job, interview, care, period	'Kanye west, the perfect person for job interview…'

This metric, which we call a news summary score, is obtained using the following equation:

$$\mathrm{Rs} = \sum(cnt(w_i, S_k) * I_{w_i}) - E(S_k)$$

where S indicates a candidate summary, $cnt(w_i, S_k)$ indicates the number of summary words in the candidate summary, and $E(S_k)$ indicates the average number of summary words in the summary candidate S_k , which depends on the candidate summary size. We determine the candidate summary with the highest *Rs* as the summary of the news article. Based on the summary, the user can easily obtain further information related to the corresponding issue from a professional journalist. Figure 5 shows an example of a news summary extracted from a news article.

4. RESULTS AND DISCUSSION

To evaluate the performance of the proposed scheme, we implemented a prototype system and conducted various experiments. All of the tweets used in the experiment were collected using Twitter Streaming API.

In the first experiment, we performed issue detection and summarization for 494,203 tweets between 11:00 AM and 12:00 AM on 2013/09/27. We extracted five trend keywords from the tweets, detected their issues, and generated their summaries, as shown in Table 1. The results show that the issues detected were closely related to the trend keywords and the summaries represent the subjects of the issues extensively and effectively.

In the second experiment, we performed news article summarization based on the results in Table 1. For each issue, we retrieved the most recent news article from Google News and identified the core part of the news article. Table 2 shows the results. In the table, most of the news summaries are reasonable and acceptable. However, in the case of "KANYE WEST," the news summary had low relevance to the issue. This is because there were few news articles about "KAYNE WEST," although the phrase appeared frequently on Twitter. Thus, users may be provided with partially related news content.

In the third experiment, we measured the accuracy of our issue

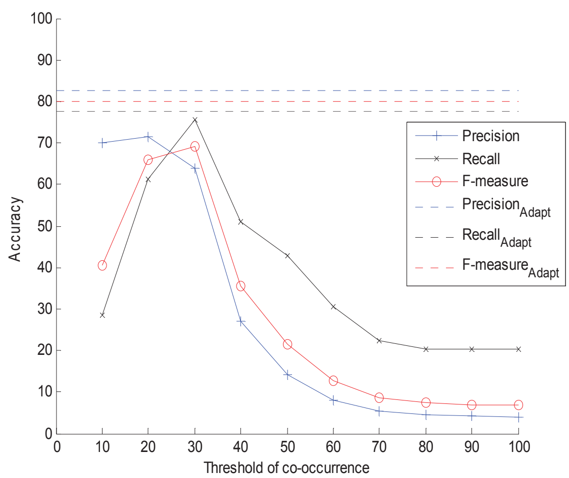

Figure 6. Detection accuracy vs thresholds

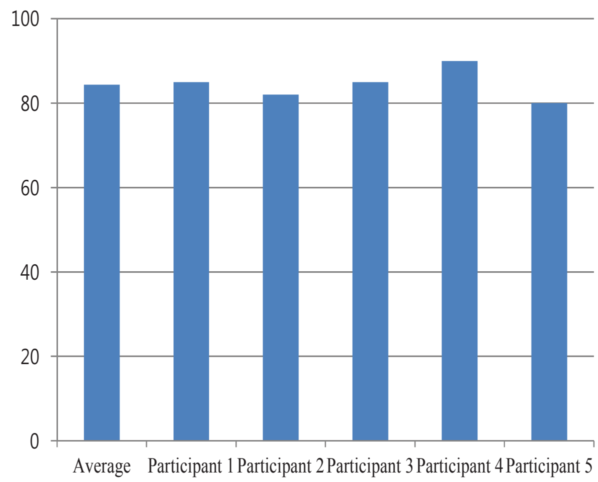

Figure 7. Satisfaction survey

Table 2. News summarization

Trend Keywords by rank	News Summary		Trend Keywords by rank	News Summary	
	Issue Summary Word List	***News Summarization***		***Issue Summary Word List***	***News Summarization***
MARIANO RIVERA	Home, Yankee, Career	And Rivera jogged onto the field for the last Yankee Stadium appearance of his career	CHRIS PEREZ	Indians, ruin, mound, leave, closer	After Indians closer Chris Perez blew a one-run save opportunity by giving up solo home run
	Derek Jetter, come , Pettitte, game, goodbye	Derek Jeter and Andy Pettitte took the ball from Mariano Rivera in his final appearance at Yankee Stad.	MILEY CYRUS	Rolling Stone, grab, Robin Thicke	The "Wrecking Ball" singer grab Robin Thicke's ass during her MTV VMA performance.
	legend, player, baseball pick, hate	From A Humble Start, Mariano Rivera Closes As A Baseball Legend Player		video, song, laugh, ratchet, artist, hate	Miley Cyrus is sure to kick up even more attention with the launch of her anticipated music video
OBAMA CARE	Obamacare, Ted Cruz, stand, defund	Cruz began speaking out against Obamacare at 2.40pm on Tuesday and vowed to keep going		hot, figure, goodness, feel	Miley Cyrus Called A "Strategic Hot Mess" By Miley Cyrus!
	Hillary Clinton, Defend, worry, respect, health	Hillary Clinton defends Obamacare, slams defunding efforts	KAYNE WEST	Twitter, blast, Jimmy Kimmel, mad	Kanye West deletes Twitter rant, insults hurled at Jimmy Kimmel after radio interview spoof
	speding, boost, money, family	President Obama likes to stress that health care spending has fallen in recent years in part to Obamacare		person, job, interview, care, period	The Perfect way to quit your job: film yourself dancing to Kanye West's "Gone"

detection scheme in terms of the precision, recall, and F-measure. Figure 6 shows the issues detected from 4,256,734 tweets during 2013/8/25 to 2013/08/29 and 2013/09/23 to 2013/09/27 using fixed and adaptive thresholds. The graph shows that the F-measures were all below 80% when using fixed thresholds. However, with adaptive thresholds, the precision was 78.4%, the recall was 81.6%, and the F-measure was 80.0%. Therefore, the adaptive threshold method provided much better results in all cases.

Finally, we performed a satisfaction survey using the summary results generated by the proposed scheme. We surveyed five subjects aged in their 20s and 30s who were familiar with SNS. Figure 7 shows the results. All of the subjects reported satisfaction levels above 80% and their average was 84.4%.

5. CONCLUSIONS

In this study, we developed a SNS-based issue detection and related news summarization scheme to generate detailed information about trending issues in an effective manner. We used Twitter stream data to detect trend keywords and issues. Each issue was represented by a list of summary words and a representative tweet. Based on this information, we retrieved a recent and issue-related news article from Google News and determined its core component by calculating the relevance score for each sentence. To evaluate the effectiveness of our scheme, we implemented a prototype system and performed various experiments. The experimental results showed that our scheme was very effective at generating news summaries and the summary results were considered satisfactory.

6. ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2013R1A1A2012627) and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the IT/SW Creative research program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2013-H0502-13-1119)

7. REFERENCES

[1] Twitter: http://en.wikipedia.org/wiki/Twitter

[2] Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022.

[3] Yokomoto, D., Makita, K., Suzuki, H., Koike, D., Utsuro, T., Kawada, Y., and Fukuhara, T. 2012. LDA-Based Topic Modeling in Labeling Blog Posts with Wikipedia Entries. *Web Technologies and Applications* 7234, 114–124.

[4] Ramage, D., Dumais, S., and Liebling, D. 2010. Characterizing microblogs with topic models. In *Proceedings of the ICWSM,* 130–137.

[5] Allan, J., Papka, R., and Lavrenko, V. 1998. On-line new event detection and tracking. In *Proceedings of the ACM SIGIR*, 37–45.

[6] Yang, Y., Pierce, T., and Carbonell, J. 1998. A study of retrospective and on-line event detection. In *Proceedings of the ACM SIGIR*, 28–36.

[7] Mathioudakis, M., and Koudas, N. 2010. TwitterMonitor: trend detection over the twitter stream. In *Proceedings of the ACM SIGMOD*, 1155–1158.

[8] Alvanaki, F., Sebastian, M., Ramamritham, K., and Weikum, G. 2011. EnBlogue: emergent topic detection in web 2.0 streams. In *Proceedings of the ACM SIGMOD*, 1271–1274.

[9] Alvanaki, F., Michel, S., Ramamritham, K., and Weikum, G. 2012. See what's enBlogue: real-time emergent topic identification in social media. In *Proceedings of the 15th International Conference on Extending Database Technology*, 336–347.

[10] Kumaran, G., and Allan, J. 2004. Text classification and named entities for new event detection. In *Proceedings of the ACM SIGIR*, 297–304.

[11] Sayyadi, H., Hurst, M., and Maykov, A. 2009. Event detection and tracking in social streams. In *Proceedings of the ICWSM*, 311–314

[12] Liu, J., Dolan, P., and Pedersen, E. R. 2010. Personalized news recommendation based on click behavior. In *Proceedings of the International Conference on Intelligent User Interfaces*, 31–40.

[13] De Francisci Morale, G., Gionis, A., and Lucchese, C. 2012. From chatter to headlines: harnessing the real-time web for personalized news recommendation. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, 153–162.

[14] Phelan, O., McCarthy, K., Bennett, M., and Smyth, B. 2011. Terms of a feather: content-based news recommendation and discovery using twitter. In *Proceedings of the European Conference on Advances in Information Retrieval*, 448–459.

[15] Barzilay, R., and McKeown K. R. 2005. Sentence fusion for multidocument news summarization. *Comput. Linguist.* 31(3), 297–328.

[16] Huang X., Wan, X., and Xiao, J. 2013. Comparative news summarization using concept-based optimization. *Knowl. Inf. Syst.*, 1–26

[17] Gong, Y., and Liu, X. 2001. Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 19–25.

[18] Kim, D., Kim, D., Rho, S., and Hwang, E. 2013. TwitterTrends: Spatio-temporal trend detection and related keywords recommendation scheme. *Multimedia Systems*, 1–14.

[19] Kim, D., Kim, D., Rho, S., and Hwang, E. 2013. Detecting Trend and Bursty Keywords Using Characteristics of Twitter Stream Data. *International Journal of Smart Home* 7(1), 209–220.

[20] Merriam-Webster's Collegiate® Dictionary with Audio: http://www.dictionaryapi.com/products/api-collegiate-dictionary.htm

[21] Kim, D., Kim, D., Jun, S., Rho S., and Hwang, E. 2013. TrendsSummary: a platform for retrieving and summarizing trendy multimedia contents. *Multimed. Tools Appl.*, 1–14.

[22] Google news: https://news.google.com/