ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 1, pp: (291-295), Month: January - March 2016, Available at: www.researchpublish.com

The Trend Analysis Service Using Twitter

Pyung Kim

Dept. of Computer Education, Jeonju National University of Education, Jeonju, Korea

Abstract: **While Twitter, micro blog service based on short messages, is important online communication media, it is increasingly important to detect and track issues through Twitter analysis. For Twitter analysis, various features including contents, time, retweets and follower relations of Twitter are used, and Twitter analysis results are visualized by using tag clouds, tree maps or time series graphs. This study aims to suggest the modelling of a trend analysis system and service for Twitter analysis results related to a specific subject to the user. The components of system architecture include data collection, data streamer, trend analysis, topic keyword extractor, language pre-processor. The proposed service is designed for presenting Twitter analysis results effectively by using selected information when a user selects the period and keywords of interest after presenting a keyword list related to a specific subject to the user.**

Keywords: **Trend Analysis, Twitter Monitoring, Trend Tracking, Twitter.**

I. INTRODUCTION

The SNS (Social Network Service) is an online platform for creating and enhancing social relations and through free communication, information sharing and human connection between users. The most important point of SNS is to create, hold, strengthen and extend a social relation network through the service. While the SNS, for example, Twitter, Facebook, LinkedIn, or Instagram, based on the mobile environment creates and distributes big data fast, various SNS analysis techniques by using big data technology have been studied.

Twitter is a micro blog service based on short messages, and enables registered users around the world to make friends each other without limitations in areas and time to send and receive short messages for communication. As the number of registered Twitter users and tweets explosively increases, Twitter is an important means to express social issues now. A registered Twitter user can extend social relations by using the relation of following-follower and retweets, and more and more people use Twitter because the messages can be delivered fast without limitations in time and space, and the registered users can respond instantly to and participate in tweets.

This study suggests a method of constructing a system for analysing trends related to specific subjects more accurately by using various features of Twitter suggested in prior trend analysis studies on SNS and Twitter. In addition, a user interface is designed as a method of effectively suggesting trend analysis results to users, which includes the visualization method suggested in prior studies. Chapter 2 describes related trend analysis studies by using Twitter; Chapter 3 describes the functions and process of components of the Twitter analysis system and service; and Chapter 4 describes conclusion and a future study plan.

II. RELATED WORKS

Studies on Twitter have been carried out in various fields including studies on statistics analysis of tweet types and users thereof [1], new information discovery or trend analysis [2, 3, 4, 5, 6, 7], crisis analysis [8, 9, 10] or opinion mining [11, 12].

Twitter plays an important role of propagating various social issues and news as a micro blog, and studies on various trend detection and analysis in relation to the role have been carried out [2, 3, 4, 5, 6, 7]. Becker et el. identify each event using an online clustering technique that groups together topically similar tweets. They suggest a general online clustering framework, suitable for large-scale social media and identify revealing cluster features, to learn event classification models [2]. They develop the TITS (Twitter Issue Tracking System) to extract issues from Twitter and visualize them on the web. The TITS can track similarity between extracted topics and resulting topic change on time series by extracting and expressing word collections, not expressing issues just as keywords [3]. Lau et al. present a novel topic modelling-based methodology to track emerging events in Twitter using time slices and dynamic vocabulary. They use LDA topic model and the model has the important properties that it does not grow over time, and can cope with dynamic changes in vocabulary [4]. Lu et al. improve the original MACD indicator according to the characteristics of news topics on Twitter, define a new concept as trend momentum and use it to predict the trend of news topics. The trends momentum is related to two moving averages and different sized moving average can track different period of trends [5]. Mathioudakis et el. suggest TwitterMonitor, a system that performs trend detection over the Twitter stream. TwitterMonitor performs trend detection with identifying 'bursty' keywords and grouping bursty keywords into trends based on their co-occurrences. After a trend is identified, TwitterMonitor attempts to compose a more accurate description of it by identifying more keywords associated with a trend using context extraction algorithms, Grapevine's entity extractor and so on [6]. Lee et al. applied the method of removing repeating contents of posts with the same hash tags by calculating the significance of posts included in the hash tags to develop an automatic Twitter summarizing system. To develop the system, they used the Twitter API to save the collected streaming data in HBase and use TF-IDF(Term Frequency-Inverse Document Frequency), Timestamp and length to develop an extraction-type summarizing system [7].

Studies on crisis analysis by using Twitter include analysis of crisis situation by analysing Twitter architecture and contents, and networks [8, 9, 10]. Klein et al. use social network analysis to identify reliable tweets and content analysis techniques to summarize key emergency facts [8]. Cameron et al. suggest Emergency Situation Awareness - Automated Web Text Mining (ESA-AWTM) system to detect, assess, summarise, and report messages of interest for crisis coordination published by Twitter [9]. Sakaki et al. concentrate changes in the frequency of tweets and analyze them to observe what happened to network users around the Great Eastern Japan Earthquake [10].

Studies on public opinion analysis by using Twitter include construction of a corpus for deciding sensitivity and then analysis of public opinions [11, 12]. Pak et al. collect a corpus for sentiment analysis and opinion mining purposes, and then perform linguistic analysis of the collected corpus and explain discovered phenomena. Using the corpus, they build a sentiment classifier that is able to determine positive, negative and neutral sentiments for a document using the multinomial naive bayes classifier [11]. Khan et al. suggest an algorithm for twitter feeds classification based on a hybrid approach. They propose a hybrid approach for determining the sentiment of each tweet and resolve the data sparsity issue using domain independent techniques [12].

III. SYSTEM AND SERVICE DESIGN FOR TREND ANALYSIS OF TWITTER

This study suggests a method of limiting trend analysis to a specific subject area to improve trend analysis accuracy, analyzing trends and suggesting analysis results. A subject keyword collection is created by suing keywords of a journal that belong to the specific subject area and used in trend analysis in order to suggest keywords in the specific subject area to users. Various twitter features are used to improve the accuracy of Twitter trend analysis. This study also suggests a trend analysis system and a method of modeling services by using various features including subject-related keywords, Twitter keywords, time, network, and relation to subjects.

1. System Components Modelling for Trend analysis of Twitter:

For providing Twitter analysis service focusing on keywords related to a specific subject by a user, the Twitter analysis system suggested in this study is composed of various units as shown in Fig.1. The Twitter analysis system includes a unit for extracting keywords related to the specific subject, a unit for determining and visualizing subject keyword weight in a concerned period when the user selects the specific period, a unit for collecting and indexing Twitter data, and a unit for analyzing and visualizing Twitter trends by using keywords delivered from the user interface.

Extracting keywords related to a specific subject is carried out in four steps of: 1) collecting journal data related to the specific subject by using a Data Collector; 2) extracting keywords and date of publishing from the journal metadata collected by using the Language Preprocessor; 3) collecting articles for the keywords from Wikipedia to extract related keywords of the extracted keywords by using the Data Collector; and 4) extracting keywords and subject information for the Wikipedia articles collected by using the Language Preprocessor. The Language Preprocessor includes the process of tokenization, word cleansing, stop-word elimination and stemming, is a system component used jointly for text processing, and has a substantial effect on the entire system accuracy.

The final query for trend analysis is made by the user and the Topic Keyword Extractor module. When the user selects a period of interest through the user interface, 1) the Calculating Weight module calculates a keyword weight depending on keyword co-occurrence and subject relation in the specific period selected by the user; 2) and the Visualizing Keyword module visualizes the keyword weight of tag cloud type to send it to the user interface. For the weight for the subject keyword, co-occurrence of journal keywords, date of journal publishing and keyword TF-IDF values shown in the article for the concerned keywords collected from Wikipedia are used.

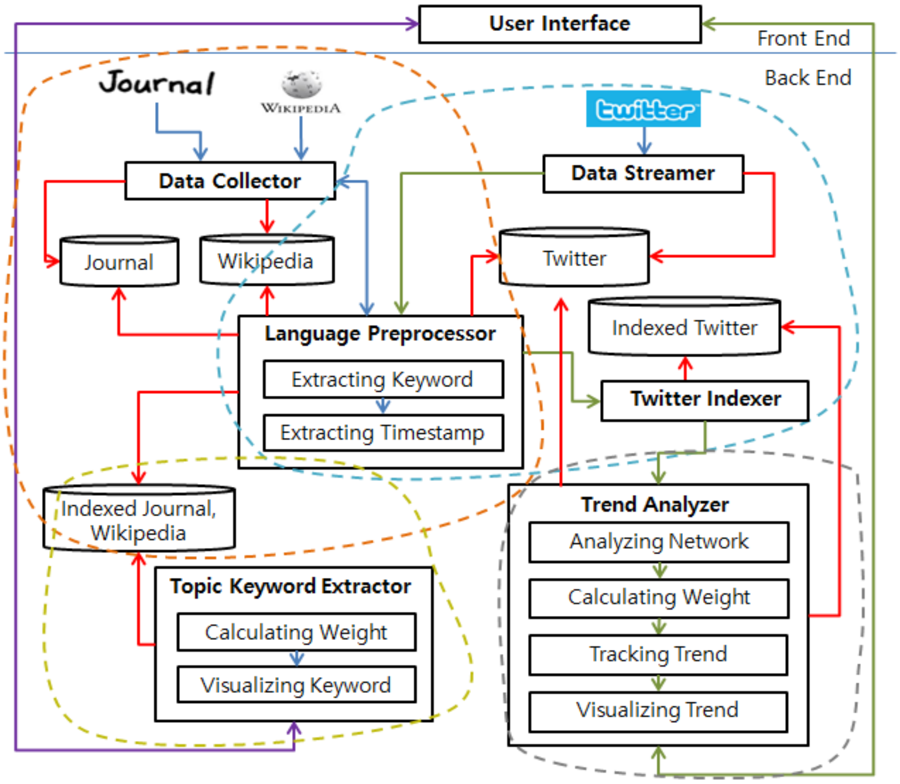


Fig. I: System Components Modelling for Trend Analysis of Twitter

Collecting and indexing Twitter data is carried out in three steps of: 1) the Data Streamer collects Twitter data through the Twitter API; 2) the Language Preprocessor extracts keywords, hash keywords and the date of Twitter by analyzing Twitter contents; 3) the Twitter Indexer converts the date mentioned in Twitter, and recognizes the following-follower relation, retweets, and named entities about place names or institution names.

When the user examines the keyword weight suggested in the selected period and selects at least one keyword, related keywords greater than the threshold are sent to the Trend Analyzer in consideration of relation among keywords. Twitter trends are analyzed by the Trend Analyzer, which includes 1) the Analyzing Network module for analyzing a network reflecting the following-follower relation among Twitters, the number of retweets and time of a tweet including a concerned subject keyword; 2) the Calculating Weight module for calculating tweet weight reflecting hash tags and keywords included in Twitter, and a relation to the subject; 3) the Tracking Trend module for tracking time and persons for the tweet of highest weight; and 4) the Visualizing Trend module for interworking to deliver trend analysis results to the user interface through visualization carried out in prior studies.

TABLE I: Features and Data for Each Module in Twitter Analysis System

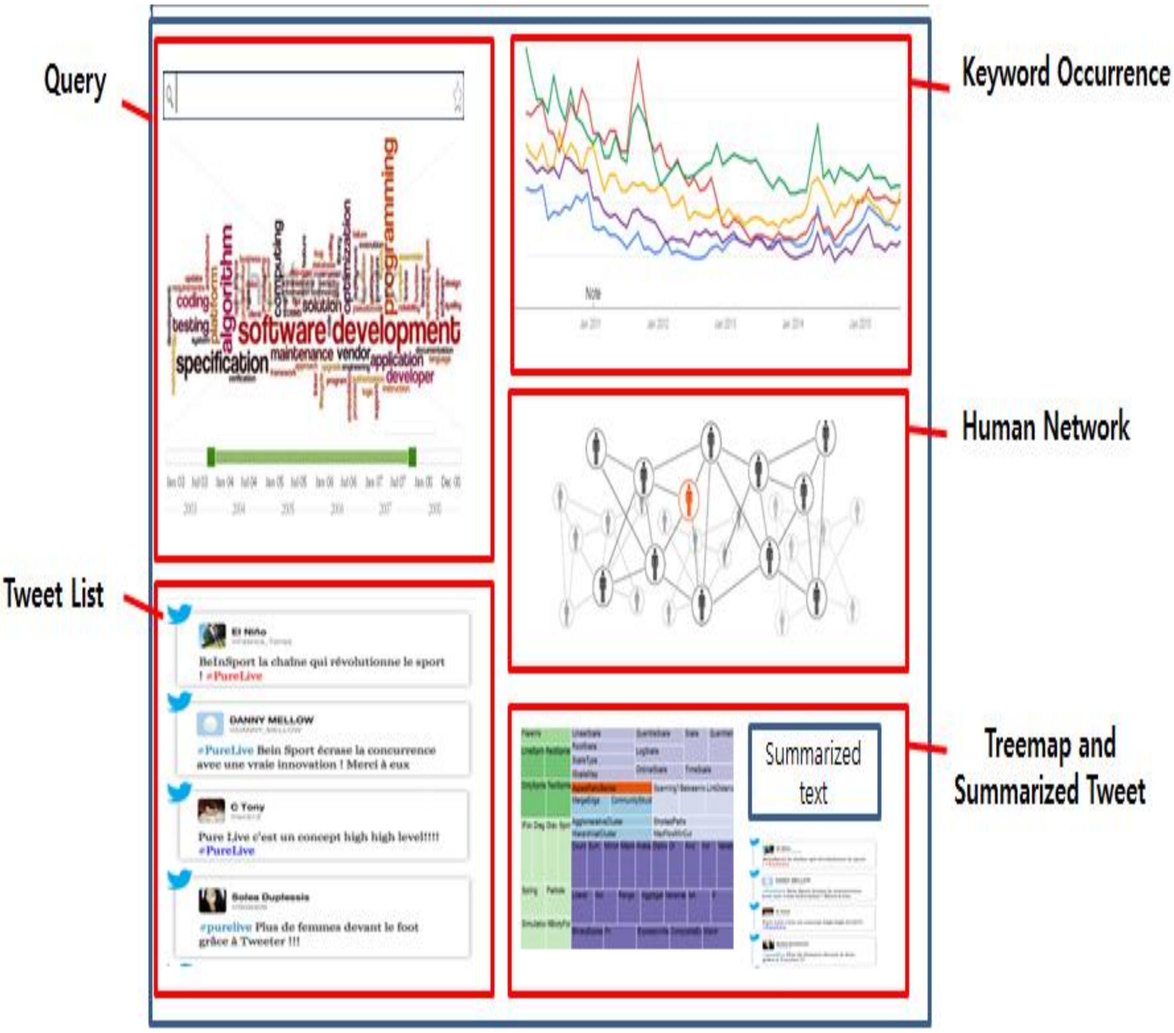
Module	Features and Data used by Each Module
Language Preprocessor	stop-word, time expression patterns, journal article metadata and abstract, Wikipedia article, indexed time
Topic Keyword Extractor	keyword list selected by user, time period selected by user, co-occurrence of keyword, TF-IDF of keyword
Twitter Indexer	tweet metadata, tweet content
Trend Analyzer	follower-followee network, retweet network, number of tweet, number of retweet, TF-IDF of topic-specific tweet, TF-IDF of topic-specific retweet, tweet date, retweet date

Table. I shows features used by each module in Twitter analysis system. Language Preprocessor analyzes the content of Twitter, metadata and abstract of journal article, Wikipedia article, and then extract keyword and time expression. To present topic-related keywords with weight to user, Topic Keyword Extractor extracts keyword list from articles of journal and Wikipedia which are in topic-specific category. Twitter Indexer analyze the structural information of tweet such as tweet time, tweeter, retweet. To detect and track topic issue Trend Analyzer use many useful feature including human network, TF-IDF of keyword, time and so on.

2. Service Components Design for Trend Analysis of Twitter:

The trend analysis service suggests analysis results by using various visualization techniques including time series graphs, tag cloud, networks and tree maps. The user interface suggested in this study is designed to 1) allow users to select time areas of interest by using a time slide bar or suggest related keywords as a tag cloud in the Query area; 2) suggest time series graphs depending on occurrence of a specific keyword selected by the user and related keywords in the Keyword Occurrence area; 3) suggest treemaps and summary information for the trend analysis results in the Treemap and Summarized Tweet area; and 4) suggest keywords or persons selected in other areas, and an actual tweet list related to the treemap in the Tweet List area.

Because Twitter is a service based on short messages, it is not easy for a user to know trend analysis results for the trend shown on Twitter through a tweet list or occurrence of words on the time series. Therefore, there is a need for visualization techniques for effectively suggesting Twitter trend analysis results. To this end, the service is provided with the topic modelling technique [3] carried out in prior studies and the function [7] for suggesting automatically summarized results by using the extraction-type summarization technique.



El Niño

BeInSport la chaîne qui révolutionne le sport ! #PureLive

DANNY MELLOW

#PureLive Bein Sport écrase la concurrence avec une vraie innovation ! Merci à eux

C Tony

Pure Live c'est un concept high high level!!!! #PureLive

Solea Duplessis

#purelive Plus de femmes devant le foot grâce à Tweeter !!!

Fig. II: Service Components Design for Trend Analysis of Twitter

IV. CONCLUSION

While data spread quickly through SNS in the mobile environment, the SNS have been studied with big data analysis technology, and Twitter based on short messages is used positively in various fields of new information discovery or tracking by using various statistics analysis techniques, crisis analysis and opinion mining. This study suggests a service model for designing a system that uses various elements of Twitter architecture, contents, time, retweets and human networks, and effectively providing trend analysis results based on the design in order to provide a trend analysis service related to a specific subject. In future studies, more accurate trend analysis models will be suggested by using the suggested model to develop an actual system and services, and testing the weight equation for various Twitter features. Another plan is to evaluate the trend analysis system performance and service usability through usability test.

ACKNOWLEDGEMENTS

This work was supported by Jeonju National University of Education Research Grant.

REFERENCES

[1] Akshay Java, Xiaodan Song, Tim Finin, Belle Tseng. “Why we Twitter: understanding microblogging usage and communities.” In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, 2007. pp. 56-65.

[2] Hila Becker, Mor Naaman, Luis Gravano. "Beyond trending topics: Real-world event identification on Twitter." Proceedings of AAAI, 2011.

[3] Jung-hwan Bae, Nam-gi Han, Min Song, “Twitter Issue Tracking System by Topic Modeling Techniques.” Journal of Korea Intelligent Information Systems Society 20(2), 2014. pp. 109-122

[4] JeyHan Lau, Nigel Collier, Timothy Baldwin. “On-line Trend Analysis with Topic Models: # Twitter Trends Detection Topic Model Online.” COLING, 2012. pp. 1519-1534.

[5] Rong Lu, Qing Yang. “Trend analysis of news topics on Twitter.” International Journal of Machine Learning and Computing 2(3), 2012.

[6] Michael Mathioudakis, Nick Koudas. “Twittermonitor: trend detection over the Twitter stream.” Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010. pp. 1155-1158.

[7] Sanghoon Lee, Seung-jin Moon. “HBase-based Automatic Summary System using Twitter Trending Topics.” Journal of Internet Computing and Services 15(5), 2014. pp. 63-72.

[8] Bernhard Klein, Xabier Laiseca, Diego Casado-Mansilla, Diego López-de-Ipiña, Alejandro Prada Nespral. "Detection and extracting of emergency knowledge from Twitter streams." Ubiquitous Computing and Ambient Intelligence, Springer Berlin Heidelberg, 2012. pp. 462-469.

[9] Mark A. Cameron, Robert Power, Bella Robinson, Jie Yin. "Emergency situation awareness from Twitter for crisis management." Proceedings of the 21st international conference companion on World Wide Web, ACM, 2012.

[10] Takeshi Sakaki, Fujio Toriumi, Yutaka Matsuo. "Tweet trend analysis in an emergency situation." Proceedings of the Special Workshop on Internet and Crisiss. ACM, 2011.

[11] Alexander Pak, Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." LREc. 10, 2010.

[12] Farhan Hassan Khan, Saba Bashir, Usman Qamar. "TOM: Twitter opinion mining framework using hybrid classification scheme." Decision Support Systems 57, 2014. pp. 245-257.