■ 376

Temporal Exploration in 2D Visualization of Emotions on Twitter Stream

Mochamad Nizar Palefi Ma'ady*[1], Chuan-Kai Yang[2], Renny Pradina Kusumawardani[3], Hatma Suryotrisongko[4]
[1,3,4]Department of Information Systems, Sepuluh Nopember Institute of Technology, Jl. Raya ITS No. 1, Surabaya, +62 31-5964965, Indonesia
[1,2]Department of Information Management, National Taiwan University of Science and Technology, Keelung Road No. 43, Taipei City, Taiwan
*Corresponding author, e-mail: mnizarpm@gmail.com

Abstract

As people freely express their opinions toward a product on Twitter streams without being bound by time, visualizing time pattern of customers emotional behavior can play a crucial role in decision-making. We analyze how emotions are fluctuated in pattern and demonstrate how we can explore it into useful visualizations with an appropriate framework. We manually customized the current framework in order to improve a state-of-the-art of crawling and visualizing Twitter data. The data, post or update on status on the Twitter website about iPhone, was collected from U.S.A, Japan, Indonesia, and Taiwan by using geographical bounding-box and visualized it into two-dimensional heat map, interactive stream graph, and context focus via brushing visualization. The results show that our proposed system can explore uniqueness of temporal pattern of customers emotional behavior.

Keywords: 2D visualization, twitter crawling, emotions, temporal data, text classification

1. Introduction

With the rapid growth of social media information and vast amounts of information, how to deal with and to understand information becomes more and more difficult [1]. This information presents a bag of unstructured text data, which usually make people unable to find the knowledge, although prior to the presentation, text classification has already been done. To face this challenge, many recent studies already perform follow-up research by combining traditional text analysis with visualization technology or also known as text visualization [2, 3, 4, 5]. Visualization was also used to obtain pattern results indicated that the spread of Dengue disease occurred on periods [6]. Furthermore, [7] provided a summary research about text visual basic framework in big data analysis with three main parts: producing visualization of the process of text analysis, visual presentation, and interaction between user and text information. One of their most referenced research is from [8], which described the basic process visualization system for text analysis and understanding of the text as shown Figure 1. This framework has been applied in some related researches with various studies.
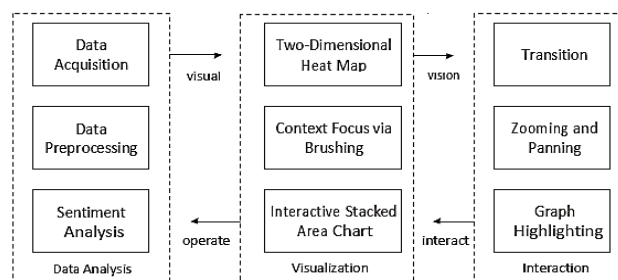

Figure 1. Basic Process Visualization System

Text visual basic framework is applied in OpinionSeer [5], an interactive visualization system that could analyze a large collection of online hotel customer reviews from TripAdvisor site, and OpinionBlocks [2], a visualization system to emphasize on enhancing users' experience of reading unorganized texts from survey data. OpinionSeer adopt the same framework as [9] and [4] that contain three major components: opinion mining, subjective logic, and opinion visualization component included a set of user interactions. While OpinionBlocks exposed more intuitive way of visualization technique by providing interactive tools and utilize smooth animations. Both of them contain parts of text analysis, visualization presentation, and interaction as basic framework in their major component's researches.

A VIsual Sentiment Analysis (VISA) system [9] analyzed the user's sentiments in the social media data, which consists of two main parts: data analysis backend covers data retrieval, feature extraction, and opinion mapping; and mash-up visualization frontend covers structure facet chart, sentiment trend overview, and mash-up interactions. VISA crawled tweets on the topic of National Football League (NFL), regardless geographic location filter. The topic of dataset can be set to be a company name like Apple or Samsung [4]. The analysis of Twitter streams evaluates consumers' reactions to the launch of a new product, which dates back to 72 hours ago, for the content to reflect their collective judgment on Apple and Samsung [4]. It can support human analysts in the collection and interpretation through a variety of analytical tools including semantic and topological metrics as well as visualization of concept and social maps [4]. However, the topic of dataset is not specified on a certain location and collected from all around the world. Analyzing consumers' reactions to a kind of product at certain locations may result more interesting information to be visualized. Therefore, we proposed a system that able to collect the data from desirable locations.

[10] implemented Naïve Bayes method for Chinese text by using Java. They classified text data to be some categories such as IT, finance, health, education, military, travel, etc. They compared English or most western languages that the text representing units are usually natural language units, such as words or phrases, with Chinese and some other oriental languages such as Japanese, Korean, and Thai. Our research is similar to huang's work that classified tweet data in different languages and encodings with the same method. [11] with their experiments about Chinese text categorization found that Naïve Bayes will be useful to consider the case of elimination of meaningless words, the way to deal with numbers and symbols, and dealing with spelling mistakes. In widely application, [12] performed Bayesian filtering for classification the spam emails in Japanese text.

Many previous related studies [13]-[16] introduced visualization for exploring the sentiments of time-varying Twitter data. SentiCompass [16] combined the circumflex model of effect on with the time tunnel representation. A combination for visual analysis of Twitter time-series is also presented in sentiment and stream analysis with geo and time-based interactive visualizations for the exploration of real- world Twitter data streams [13]. Existing system can analyze societal behaviors by visualizing time-varying topics in multiple media [15]. Topicflow shows the frequency of tweets related to a particular topic over time in shape time slices [15]. We proposed visualization for exploring the users' emotion of Twitter time stream with particular geo-topic spacial.

Nowadays the application of visualization technology is very widely used [7]. Available data visualizing tools in a free and open source system, such as D3.js and infoVis [17], motivate us to develop a visualization system that serves as a convenient tool for such a purpose. To meet the above-mentioned challenges, we propose the exploration of temporal data from Twitter stream by using two-dimensional heat map visualization, interactive stacked area chart, and context focus via brushing with its graphical interaction as well.

2. Architecture of the Proposed System
The general architecture of the proposed system is shown in Figure 2. The input data is collected from Twitter in real-time and stored directly in our tweet database. Next, we perform data preprocessing, sentiment analysis, and visualization presentation as output of the system. The different components of the proposed system are described in detail below.
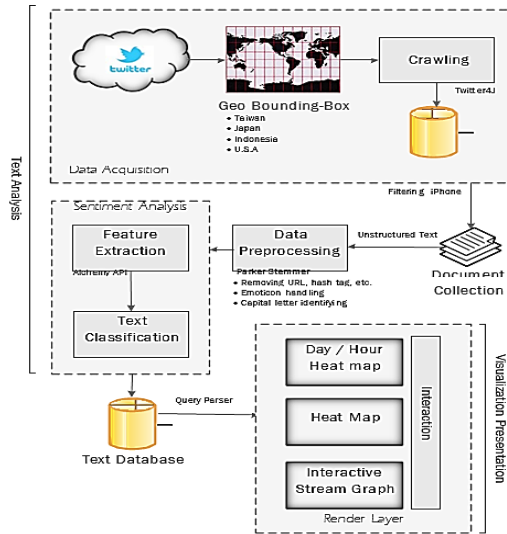
Figure 2. The architecture of our proposed system

2.1. Text Analysis

There are many accessible libraries in several programming languages making it easier for developers to access the full fire hose of public tweets [18]. In this system, we collected tweets in real-time using Twitter's streaming API crawler, which is made from the Twitter4J library for Java applications. We use Twitter4J library to work together with our proposed text classification phase and visualization tools in Java platform, since Twitter4J is the most up-to-date Java binding library for Twitter Streaming API [20]-[21]. Simple interface of our crawling system can be seen as Figure 3.

Figure 1. Automatic twitter crawling system

Due to the restriction in Twitter Streaming APIs, we are not able to set keyword filter along with geographical filter in the same time. Thus, we separated acquired data into two steps. We developed a data collection system to extract live tweets, and to refine it by a combined filtering: geographical bounding-box and keyword setting. First, we retrieved all tweets from several certain countries by using a geographical bounding-box and stored in our tweet database, then we set up a keyword filter to get the dataset that is ready to be processed. Twitter data used in this research was collected from the entire U.S.A., Japan, Indonesia, and Taiwan.

In our crawling system, we define the coordinate values of state borderline that cover all exactly the regions of a country. Geographical bounding-box given the red line on Japan and Taiwan. It becomes more complex when defining state borderline for U.S.A. and Indonesia, since they have many neighboring countries around them. Therefore, we rather extend the borderline of some region spots for the sake of getting the dataset that covers all the regions of the country.

The aim of collecting data from these chosen countries is to gather the data with different Latin languages and character encodings, and so to consider about the adequacy of the data, where these countries include in top-20 countries according to the number posted tweets [1]. Multi-languages are represented by Indonesia and English language as mother tongue of U.S.A., while Chinese character and Japanese kanji represent different character encodings. In addition, with the issue of online social media, Twitter users may post unique characters (e.g., '❤', '♪', '✈', etc.) besides in general character set. Our proposed system performed utf8mb4_unicode_ci, by using the Unicode Collation Algorithm as defined in the Unicode standards, as a customizable collation to recognize unique character including different character encoding in language and emoticons state.

Our dataset contains about 11.338 tweets related to iPhone between October 19, 2015 and November 30, 2015 in four specific locations: U.S.A., Indonesia, Japan, and Taiwan. Each comment is a targeted data as the nature of user tweet. Due to the nature of language used in micro-blogging posts, preprocessing on the messages is very necessary. In the next step, we preprocessed tweets, which are as only relevant to the keyword of iPhone as follows.

Tweet message contains meta-data information with typical format that is with the following order: the user name and ID of the sender, the user name of the receiver preceded by @ symbol, the hyperlink to other web contents which is optional, the actual tweet message, the hash tag, location information, creation date, time and year [16]. Our primary goal in this step is to extract temporal information and its actual content of the tweets. While all tweets contain timestamps (i.e., at the moment when the tweet is posted), this data provides arbitrary information relevant to the present timeframe [18]. Besides we labelled appropriate attributes, creation date and tweet message, we also used tweet ID attribute related to the topic, which may contain useful information considering the use of retweet. A retweet allows a user to repeat a tweet created by another user, usually indicating their support for or interest in that tweet's content [20]. Tweet ID played as a distinguisher role for every posted tweet regardless repeated tweet. Moreover, due to the need of text classification, we performed text operations such as parsing, stemming, stop word removal, and part-of-speech tagging by using third-party analysis called Porter Stemmer. Its main use is as part of a term normalization process that is usually done when setting up information retrieval systems

This idea, which may be regarded as an easily computable representation of a syllable. A typical rule is thus as follows:

$$(m > 0) * fullness \qquad * ful \tag{1}$$

this means that the suffix "fullness" should be replaced by the suffix "ful" if, and only if, the resulting stem has a non-zero measure (m).

Furthermore, we used emoticon sentiment lexicon dataset created by research of [21]. Each line in the lexicon dataset contained an emoticon and its associated sentiment according to human annotators, that is positive and negative class. Its mode is in sensitive case, so it has some very similar smileys, of capital or lower case of the alphabet. The emoticon dataset collected from several lists of smiley faces from other researches. Here, we show some of emoticons dataset along with its classified emotion in Table 1. So that we have emoticons state in additional feature dataset and then we start to process the prepared data in sentiment analysis systems that explained in the next sub-section.

Table 1. Emotions Dataset

No	Emoticon	Class	No	Emoticon	Class
1	#)	Positive	6	>;)	Negative
2	#-)	Positive	7	>[	Negative
3	%)	Positive	8	>\	Negative
4	%-)	Positive	9	>+(	Negative
5	(^)o(^)	Positive	10	>+-(	Negative

We developed a text classification to judge customer's emotion from opinions in micro-blog. Process flow of the system begins by extracting features from tweets. In feature extraction

phase, we used another third-party analysis, AlchemyAPI, to get word features of each document in two sets: positive and negative. Then, we build an application for text classification based on Naïve Bayes method. We developed a sentiment analysis system or also called opinion mining to judge customer's emotion from opinions in micro-blog. Process flow of the system begins by extracting features from tweets. In feature extraction phase, we used another third-party analysis, AlchemyAPI, to get word features of each document in two sets: positive and negative. Then, weExtracting a bag of word features is needed for classifying text data in the algorithm of Naïve Bayes. This classifier is often used in text classification due to its speed and simplicity. It provides a way of calculating the posterior probability. The formula of Naïve Bayes classifier is defined as:

$$P(c|x)=(P(x|c)\ P(c))/(P(x)) \tag{2}$$

where P(c|x) is the posterior probability of classes of positive, neutral, and negative, while P(x|c) is the likelihood which is the probability of word features in a document of the given class in training data. P(c) and P(x) is the prior probability of class and predictor. In our research, this formula, then, becomes a classifier application with uploading system like Figure 4.
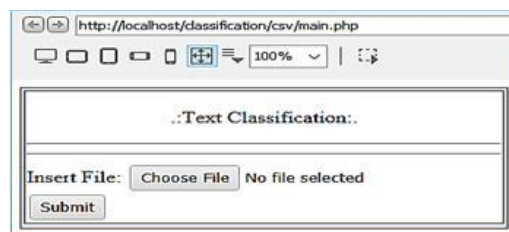

Figure 4. Text classification application

Twitter data that has been through data preprocessing phase is uploaded to the system. Uploading a Comma Separated Values (.csv) extension file into database is to imply Naïve Bayes algorithm computation and to save its classification result into the empty field in the sentiment column of database. Before it is stored, the calculation result is displayed in the preview window to get to know the comparison between positive and negative value in each record. As example, our text classification process can be seen in Table 2.

Table 2. Text Classification Process

ID	Date	Name	Text	Sentiment
2146037760	2015-11-06 21:19:33	danielaames sere	who has an iphone 5 charger ?	neutral
2141949952	2015-11-09 03:04:19	ahmet_bekta s23	rip iphone charger	negative
2108252160	2015-10-21 23:36:29	loncrinibrian na	sooo excited ❤ #rosegold #iphone6s #glassscreenprotector #clearcase #sopumped #thanksmom	positive

2.2. Visualization Presentation

Since we have a dataset in four different locations, then we separate the visualization presentation in four windows by linking each other. We create a home page, which allows a user to go back to the main page, and go to another visualization results' window as well. We create an automatic color scale based on the number of data. The scale is representing gradation color that should be put on the cell. We apply this formula to the color scale legend:

$$colorScale=(0,\ bucket-1,\ d3.max(data,\ function\ (d)) \tag{3}$$

where bucket is a threshold that we can set approximately by looking at the highest value of the data and d is referring to the related data. So by determining the threshold, we can get color scale legend that implies on finding the best coloring of the heat map visualization. The scale can be automatically set when applied in several displays. Color scale legend can be seen on Figure 5.
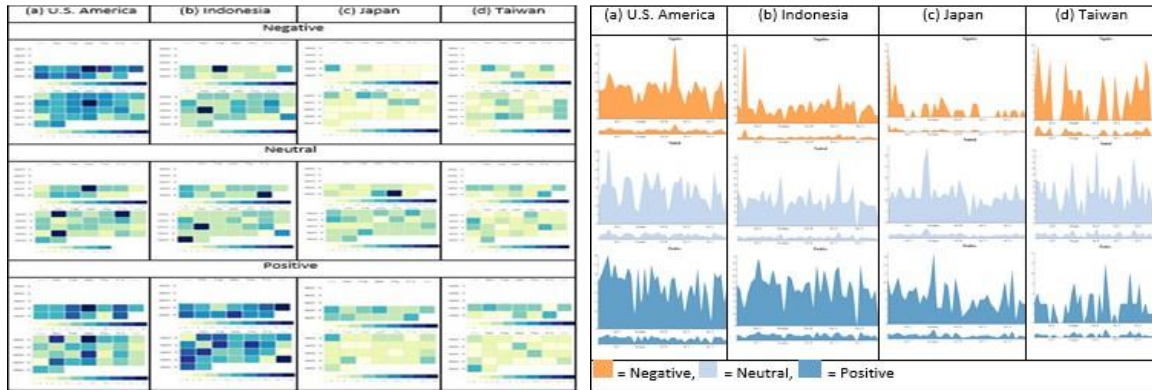

Figure 5. Color scale legend on week/day heat map

Figure 6. Context focus chart after adjusted

Selecting period in context chart can affect the changing of focus chart, but it is not vice-versa. Focus chart shows time period adjustment more specifically. Context and focus chart is provided by utilizing D3's component. Context chart is an addition small chart that enables us to pan or zoom by clicking and dragging specific period, where focus chart is like its enlarged size. For example, Figure 6 shows focus chart during one week as adjusted, which can be dragged to the right or left as well.

To utilize the use of chart, we employ conventional margins and a number of D3 features such as d3.tsv, d3.time.format, d3.time.scale, d3.scale.linear, d3.extent, and d3.max. These features used to load and parse time data. So that x-axis can display the label in parse date based on how the date format in loaded data. Furthermore, y-axis is also able to set the maximum value and parses it into several value labels based on the given data in a csv file. To describe the shape precisely, we use the following notation. We model our time series as a set of n real-valued functions, $f_1, \ldots, f_n$. We refer to the baseline function that defines the bottom of the stacked graph as g_0. The top of the layer corresponding to the ith time series f_i is therefore given by the function g_i, where:

$$g_i = g_0 + \sum_{j=1}^{i} f_j \tag{4}$$

Which is illustrated for these definitions for n=2 by Figure 7. There are variety of possible criteria to choose the baseline function g_0. The simplest is the traditional stacked graph, which has $g_0=0$.
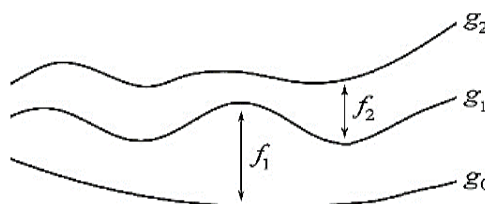

Figure 7. A Visual description of stacked graph functions

Like general statistical chart, stacked area chart shows date time and number of data as x-axis and y-axis, respectively, but the difference is that a stacked area chart having a set of layers grouped in an area chart that determined by two factors: the shape of the "baseline," or bottom of the lowest layer, and the order of the layers. Stacked area chart shown as Figure 8.

The stacking effect was achieved through d3.layout.stack. the only customization we have done on stack layout is setting its offset to zero. The stack layout automatically calculates a baseline y0 for each datum in our three different data series. The D3 stack layout has built-in support for this kind of stacking algorithm therefore changing a zero based stacked area chart to streamgraph is trivial. The key difference is that stream graph uses wiggle as its layout offset mode, since there are a few different offset modes, which determine what stacking algorithm to use.
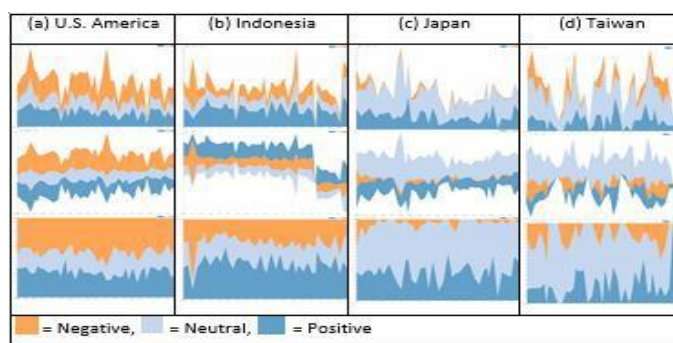

Figure 8. Stacked area chart

We employ D3 human visualization interaction techniques to our visualization. Moreover to get interactive visualization, we implement some animations into it based on the needs of temporal exploration and transitions as well, consist of: in the two-dimensional heat map, we enable element transition into each cell, which represented the number of data in color scale, with the duration to 2500 milliseconds using the duration() function; context focus via brushing provides zooming and panning in the small chart to work with the big chart one; zooming and panning are common and useful techniques, which work particularly well with Scalable Vector Graphics (SVG) based visualization; we develop three types of interactive stacked area chart visualization with the same animation.

3. Results and Discussion

To achieve the research objective, several stages were done, including: collecting dataset, preprocessing the data, classifying the processed data using sentiment analysis, presenting the data into appropriate visualizations and implementing its interactions, evaluating the model, and analyzing the presentation in order to explore the temporal data of Twitter stream.

3.1. Dataset

The data from three different categories of positive, negative, neutral or also labelled happy, unhappy, and normal are collected from Twitter stream about iPhone product along with related issues from October 19, 2015 to November 30, 2015. After post-processing the data, we got 5.545 tweets from U.S. America, 4.012 tweets from Indonesia, 1.181 tweets from Japan, and 600 tweets from Taiwan, for 11.338 tweets. Then we took 129 sample tweets that we got by pick up three tweets randomly in each day for being the testing data, and the other tweets for being the training data. With the remaining training tweets, we extract features as many as 752 words and emoticons within.

3.2. Model Evaluation

Experimental results for positive or happy class using information taken from confusion matrix table give precision, recall, and accuracy 66%, 87.5%, and 80%, respectively. Comparing

with existing research, it is almost same with the accuracy of another research [22], which also used Naïve Bayes for Twitter data that is 81.3%. However, they did the text classification for dataset in English language only. We also compared with another research using the same data language [13] that they have precision and recall result as follows 76.3% and 77.9%. It shows that our proposed sentiment analysis system combined with the appropriate third-party analyses achieves higher result for recall, despite having lower precision. The high value of recall may result the good enough performance of classifier to recognize the existence of all positive values.

3.3. Visualization Analysis

We presented all of our proposed visualization techniques based on classification results by comparing the entire datasets of U.S. America, Indonesia, Japan, and Taiwan to make ease of temporal pattern analysis. We joint all visualization results among interactive stacked area chart, context focus via brushing, and two-dimensional heat map into several tables.

Based on the results show that we noticed unusual shape on streamgraph for Indonesia dataset. On November 11, 2015, there is a stream lane fracture followed by different direction after, because it reverses the number of positive and negative data after that day. It may be occurred due to technical issue of the use of Twitter API that cannot to retrieve the data for hours. Meanwhile, a peculiar analysis is found on Japan dataset. By looking at tweets distribution on day/hour heat map for positive set, we observe that there are much more users talking nicely about iPhone at night than daylight, where Japan lapse of time is one hour ahead from Taiwan time zone. At the local time itself, it is difficult to figure out any analytical finding, because Taiwan dataset has a little number of tweets, where it is at the least of data among these four countries. However, in November, we spot that people are rarely talking about iPhone product in negative way on Friday.

With the accuracy of 80%, these visualization models may have contributed to acquire temporal pattern of iPhone customer behavior. That table provides whole visualization views in different location datasets. We strive to figure out analysis outcome from each aggregation dataset. According to the U.S. America dataset, the most common happy expression is fluctuated significantly on every Thursday, which happened in the morning at local time or about late evening in U.S. America.

Based on the visualization presentations, we highlight some data representative results from collected country dataset by analyzing its temporal pattern. Firstly, most of iPhone customers alternately give positive comments on every Thursday evening in U.S. America, while it is often occurred on the night than daylight in Japan. Meanwhile, Indonesia is having a strange event on November 11, 2015 letting no one comment to iPhone on that day. Last, lacking data suffered on Taiwan causing it hard to figure out that most customers almost never comment negative opinions on every Friday in November.

By using appropriate third-party analyses and Naïve Bayes as text classification method, we achieved 80% in performance measurement of accuracy, while precision and recall is 66% and 87.5%, respectively. Then we developed JavaScript libraries in such a way to visualize its result for the sake of temporal exploration and figured out some analysis results during the observation process in the visualization presentations. Comparing with similar study, we show that our proposed classification method achieved higher value of recall than previous research. However, we would like to point out that we have obtained the precision and accuracy with not satisfied enough result and sustained shortage of Taiwan data causing it to be hard to analyze its visualization.

4. Conclusion

In this research, the proposed method can acquire the data in a desirable location. In this case, we set Twitter data collection in U.S. America, Indonesia, Japan, and Taiwan. The results show that the recall achieved higher value than the existing study with 87.5% when the accuracy and precision achieved as well as the value of 80% and 66% in consecutive. It should be noted that this study has been primarily concerned with performing a visualization for the sake of temporal exploration. We would like to develop a better visualization presentation by using 3D visualization for providing more representative visualization presentation. In further

study, we also interest to build a visualization system that able to present not only iPhone product but also along its related topics. So that we may explore visualization presentations in deeper to attain more valuable analysis results. It would be more useful for companies, thereby to trace the overall opinion trend in the social network as well as to identify key issues. To reach out this work, we would like to combine our system with the method of general social graph.

References

[1] Semiocast. (2012). Retrieved from Semiocast Top 20 Countries: https://semiocast.com/en/publications/2012_01_31_Brazil_becomes_2n d_country_on_Twitter_superseds_Japan

[2] Alper B, Yang H, Haber E, Kandogan E (n.d.). OpinionBlocks: VIsualizing Consumer Reviews.

[3] Duan D, Qian W, Shi L, Pan S, Lin C. VISA: A VIsual Sentiment Analysis System. VINCI, Hangzhou, China. 2012.

[4] Lipizzi C, Iandoli L, Marquez JE. Extracting and Evaluating Conversational Patterns in Social Media: A Socio-Semantic Analysis of Customers' Reactions to the Launch of New Products Using Twitter Streams. *International Journal of Information Management 35, Elsevier Ltd.* 2015: 490-503.

[5] Wu Y, Wei F, Liu S, Au N, Cui W, Zhou H, Qu H. OpinionSeer: Interactive Visualization of Hotel Customer Feedback. Transactions on Visualization and Computer Graphics. 2010: 1109-1118.

[6] Eosina P, Djatna T, Khusun H. A Cellular Automata Modeling for Visualizing and Predicting Spreading Patterns of Dengue Fever. *TELKOMNIKA (Telecommunication Computing Electronics and Control).* 2016; 14(1): 228-237.

[7] Qi Y, Shi G, Yu X, Li, Y. Visualization in Media Big Data Analysis. *ICIS IEEE.* 2015

[8] Tang J, Liu Z, Sun M. The Review of Text Visualization. *Journal of Computer-Aided Desing & Computer Graphics.* 2013: 273-285.

[9] DC Manning, Rahavan P, Schutze H. Introduction to Information Retrieval. Cambridge University Press. .

[10] Huang J, Cen Z, Zheng Q. *Research and Implement of Chinese Text Classifier Based on Naive Bayes Method.* IEEE 6th International Conference on Semantics, Knowledge, and Grids. 2010: 426-428.

[11] Yu XR, Xiang ZL, Kang DK. Classification of Chinese-To-English Translated Social Network Timeline using Naive Bayes. *ICACT.* 2015: 296-299

[12] Iwanaga M, Tabata T, Sakurai K. (n.d.). Some Fitting of Naïve Bayesian Spam Filtering for Japanase Environment.

[13] Hao MC, Rohrdantz C, Janetzko H, Keim DA, Dayal U, Haug Le, Stoffel F. Visual Sentiment Analysis of Customer Feedback Streams Using Geo-Temporal Term Associations. Information Visualization 12 SAGE. 2013: 273-290.

[14] Itoh M, Toyoda M, Kitsuregawa M. *Visualizing Time- Varying Topics via Images and Texts for Inter-Media Analysis.* 17th International Conference on Information Visualization. 2013: 568-576.

[15] Malik S, Smith A, Hawes T, Papadatos P, Li J, Dunne C, Shneiderman B. *TopicFlow: Visualizing Topic Alignment of Twitter Data over Time.* IEEE/ACM International Conference on Advances in Soical Networks Analysis and Mining. 2013: 720-726.

[16] Wang AH. *Don't Follow Me: Spam Detection in Twitter.* Proceedings of the 2010 International Conference on Security and Cryptography (SECRYPT), Piraeus. 26-27 July 2010.

[17] Nadirman F, Ridha A, Annisa. Searching and Visualization of References in Research Documents. *TELKOMNIKA (Telecommunication Computing Electronics and Control).* 2014; 12(2):447-454.

[18] Twitter. (2006, July 6). Retrieved from Twitter Libraries: https://dev.twitter.com/docs/twitter-libraries

[19] Yamamoto Y. (2013, Septembser 9). http://twitter4j.org/ko/. Retrieved from http://twitter4j.org/ko/

[20] V Pandey, a. C. (n.d.). Sentiment Analysis of Micro blogs.

[21] Hogenboom A, Bal D, Frasincar F. Exploiting Emoticons in Sentiment Analysis. SAC ACM. 2013.

[22] Go A. Bhayani R, Huang L. Twitter Sentiment Classification using Distant Supervision. CS224N Stanford. 2009: 1-12.