TWTRENDS — A VISUALIZATION SYSTEM ON TOPIC MAPS EXTRACTED FROM TWITTER TRENDS

Jun Iio
Chuo University, 1-18 Ichigaya-tamachi, Shinjuku-ku, Tokyo 162-8478, Japan

ABSTRACT

Twitter provides a list of trending topics, a list of social trends or issues that users of the social networking services (SNS) will be interested in at a given time. However, they are presented as a list of key phrases making it challenging to identify or infer the message of the trending topics. In an attempt decipher the meaning and the structure hidden in the list of trending topics provided by Twitter, we developed a system to visualize topic maps according to the co-occurrence structure of typical tweets that mention the trending topics. This paper describes an overview of the system structure, the main logic to construct topic maps, and a discussion on several findings from its operation between January 1 and June 30, 2019. Also the possibilities of modification to suit the system for the other languages is discussed.

1. INTRODUCTION

Social network services (SNS) are commonly used today, and they have become fundamental tools for daily communication online. Twitter is a microblogging service with more than 333 million active users. It has 139 million daily active users worldwide, and 45 million monthly active users in Japan (@TwitterIR 2019) as of April 2019.

Twitter provides a list of trending topics in an area e.g. a city such as Tokyo, Paris, and London (Figure 1). Trending topics are highly dependent on an area as they are representative and/or are connected to social issues in the area. Additionally, the tweets are presented in the language which the users speak in their daily lives. If one is interested in the trends in Japan, trending topics are the listed in Japanese, and if one is interested in the trends in the US, they are listed in English.
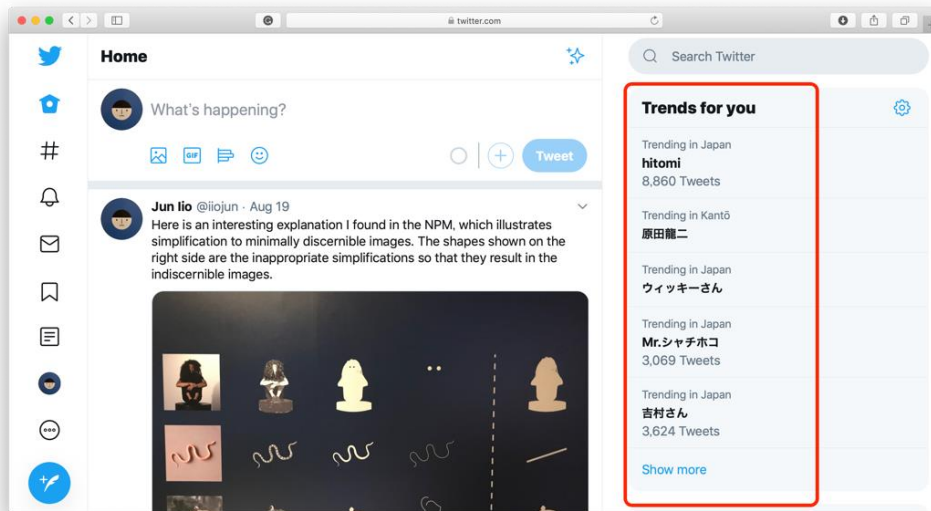
Figure 1. Twitter provides a list of the trending topics

Although the algorithm used by Twitter to select the trending topics is not known, it is easy to presume that the list of trending topics is derived from the most popular topics based on the number of users who submit tweets on the topic. This suggests that the list of trending topics shows several social phenomena and is a form of collective intelligence. The list of topics suggests to other users' topics to consider as interesting. However, they are sporadic, and this makes it difficult to select important topics and those which are not. Twitter only provides a list of key phrases, and it does not elaborate on the structure or relationship between the topics.

Before describing an overview of the system and results from the operation of the system, note there are two kinds of trending topics i.e. one is a hashtag type topic phrase, and the other is a non-hashtag type topic phrase. The hashtag type topic phrase has intentions of the user who creates it or the information of the providers. Twitter users who have interests in some issues use a specific hashtag to identify their tweets on that issue and to distinguish them from other topics. The non-hashtag type topics are spontaneously arising topics. In our study we excluded tweets with specific intentions by not including the hashtag type topics.

Analysis of the topics that are discussed on the SNS every day and how users of the SNS mention such issues in their tweets can reveal social trends. The aim of our study was to visualize topic structures and to construct trending-topic maps. Visualization would allow for us to easily capture the details and the information of the social trends generated by a large number of SNS users.

This paper describes the related work of our study (Section 2), an overview of the methods (Section 3), and its implementation (Section 4). Section 5 illustrates how to create a topic map graph based on the similarity between two topics, and several findings from the operation conducted between January 1 and June 30, 2019 are discussed (Section 6). Section 6 also explains the possibility of modifying the system to suit other languages.

2. RELATED WORK

There are several studies on Twitter trends, especially on its application on citizen's attitude surveys of specific themes that are proposed by many researchers.

Lewis *et al.* analyzed a discussion on Twitter on the pros-and-cons of electronic cigarettes (Lewis *et al.* 2015). Bovet *et al.* examined the results of applying the Twitter analysis system they developed for the US presidential election in 2016 (Bovet *et al.* 2018). Lin *et al.* analyzed Twitter data to allow journalists to obtain news topics from social media and they conducted network analyses that provided network graphs comparable to what we are proposing in our study (Lin *et al.* 2016). Thus, there have been many studies that have applied Twitter's trend analysis to various themes.

Zubiaga *et al.* reported that Twitter had become an essential platform for socio-informatics researchers (Zubiaga *et al.* 2015). They defined a typology to classify a trending topic into four types: news, ongoing events, memes, and commemoratives and performed classification experiments using support vector machine classifiers. Dawar *et al.* also proposed a data-investigation tool for Twitter trends (Dawar *et al.* 2018), and their tool drew network graphs and is comparable to our proposed tool, although their diagram is a user-base network graph. Khandekar *et al.* (2017) and Wen and Marshall (2014) have also summarized the topics discussed on Twitter by using Twitter trends. However, the novelty of our study is not only in that it is visualization method but in that we also developed a system to provide visualized topics.

There are other approaches to disclosing social trends by analyzing Twitter trends. For example, Becker *et al.* (2011) used a standard machine-learning technique for their event classifier to distinguish between messages about real-world events and non-event messages. Several trending topics were shown in clusters identified by real-world events. Benhardus and Kalita proposed an algorithm to detect and identify trending topics from streaming data. Their approach applied a calculation based on a term frequency-inverse document frequency (tf-idf) analysis. Aiello *et al.* (2013) also discussed a method of detecting trends, using a set of tweets. In their case studies, they proposed an algorithm designed to extract trending topics from a massive set of tweets. On the other hand, our approach is based on the existing set of trending topics. Their results of previous studies can be used as input in the present study.

Guille and Favre proposed a novel method to detect events and provide visualizations of trending topics (Guille and Favre 2015). They proposed the mention-anomaly-based Event Detection (MABED) method, which detects events and also visualizes them. They conducted experiments on both English and French Twitter data and showed that several visualizations could provide insightful information. Giridhar and Abdelzaher (2017) reported an interesting finding in their work, which visualizes events extracted from SNS. Their system is unique because it analyzes both Twitter and Instagram simultaneously. It is a challenge to analyze social trends using multiple sources. Their method creates fusion models from location information, drawn from the Instagram application program interface (API) and detailed event information from Twitter API. Although these visualization methods are both interesting, they seem experimental. By contrast, our system is both practical and useful for real applications, such as marketing and cultural analysis.

Other visualization studies on tweets tend to create graphical information using tempo-spatial meta-data associated with each tweet. Sechelea *et al.* (2016) proposed a system to indicate the intensity of Twitter activity at a given geographical location. Yaqub *et al.* (2018) implemented a system showing results of sentiment analysis on tweets. This system also plots

their results onto a map using geographical information from tweets. Similar studies have recently been proposed (Maia *et al.* 2016, Garg and Rani 2017). However, there is a limitation, as tweets with geographical information are very rare. Furthermore, Twitter announced that it will no longer support accurate geographical information for tweets. Therefore, further geographical analysis using Twitter data will be considered difficult hereafter.

3. THE METHOD OF OUR STUDY

The method of our study is simple and is described as follows:

1. collection of the lists of trending topics using Twitter trends API,
2. collection of tweets that mention the trending topics using Twitter standard search API, and
3. the development of a system and its use for co-occurrence network to show the structure of the trending topics.

This procedure was conducted every 20 minutes, and this was done daily. All data collected for a day was used to make a topic map from the set of all co-occurrence networks operated. The results were a visualization of the topic structure for the day, hidden in the list of topic sentences.

3.1 Preparation

Before implementing the system, we had to create a Twitter developers' program account which would give us the rights to access some of the APIs provided by Twitter. We registered ourselves as a developer for Twitter applications through their registration page[1], and we gave a short description of our information. Our request was assessed by Twitter, and when it was approved we were given access to the consumer API key (CK), the consumer API secret key (CS), the access token (AT), and the access token secret (AS), which are requested when a program accesses the server requesting for information from Twitter.

3.2 Trending Topics

Acquisition of the four essential codes allowed us to access several services from Twitter. Access to the Twitter trends API provides information on trends and also information on the region-specific trends. The "GET trends/available" command's response is a list of acceptable regions in a JavaScript Object Notation (JSON) form. The results of the trends APIs showed that there were twenty-two regions available in Japan.

We requested a list of the trending topics for the regions by sending the area-identifying information, in the form of "where on earth ID (WOEID)" as defined by Yahoo! and placed it in the parameters of the query in Twitter. The prototype system we developed in this study had one limitation on the system, which was for it to only collect the trends for the Twitter users living in the Tokyo metropolitan area (Figure 2). Note that the personal information (CK, CS, AT, and AS) is masked.

---

[1] https://developer.twitter.com

Figure 2. The python script used to obtain the list of trending topics (left), and its result (right)

The system had other restrictions and options. The list of trending topics was updated every five minutes and accesses was restricted to 75 times every 15 minutes. The hashtag type trending topics were filtered out by adding the phrase "`exclude=hashtag`" to the query parameter.

3.3 Tweets Mentioning the Trending Topics

After obtaining the list of trending topics, the system further filtered the tweets to sets of the listed issues. The tweets were obtained using the Twitter search API. There are three types of search APIs i.e. the standard, premium, and enterprise. The standard search API is available for free; however, the premium and enterprise search APIs are not free, and payment must be made for their use. In this study, we used the Standard search API as it is a free service, although it has some restrictions on retrieving tweets. The most significant limitation was that we could only access the tweets posted within a week. However, this did not impact this study because we were interested in the current topics. The standard search API also had a limited access and could only be accessed 180 times (in the case of user authentication base) or 450 times (in the case of app-authentication) every 15 minutes.

Under such restrictions, we collected a maximum of 100 tweets for each trending topic excluding retweets. We excluded retweets because they have identical sentences and were not suitable for word frequency or co-occurrence calculations. The retweets were filtered from the results by adding the phrase "`exclude: retweets exclude:nativeretweets`" to the query parameter.

3.4 Visualization of Trending Topics

The procedure of visualizing trending topics hides the calculation of the word frequency and the co-occurrence network. A morphological analysis was conducted on every tweet to allow for the calculation of word frequency and co-occurrence probabilities because all the words in sentence in Japanese are connected without any spacing characters except punctuations. The system filtered proper nouns from the set of words of each tweet. The word frequency and the co-occurrence network were calculated based on the proper noun sets.

The co-occurrence network illustrates the topic structure and the scope of co-occurrence in one tweet. The size and color of the node and the thickness of the link is dependent on the value of the word frequency and the co-occurrence probability, respectively. Note that only one-fourth of the links are shown in the graph to simplify the network graph.

An example of a co-occurrence network graph is given in Figure 3. The node "(Tabi (journey) salada)" has the largest word frequency and several links to other nodes. The link between "(Tabi (journey) salada)" and "(Milano)" has the largest value of co-occurrence probability, and the link is defined by the thickest line.
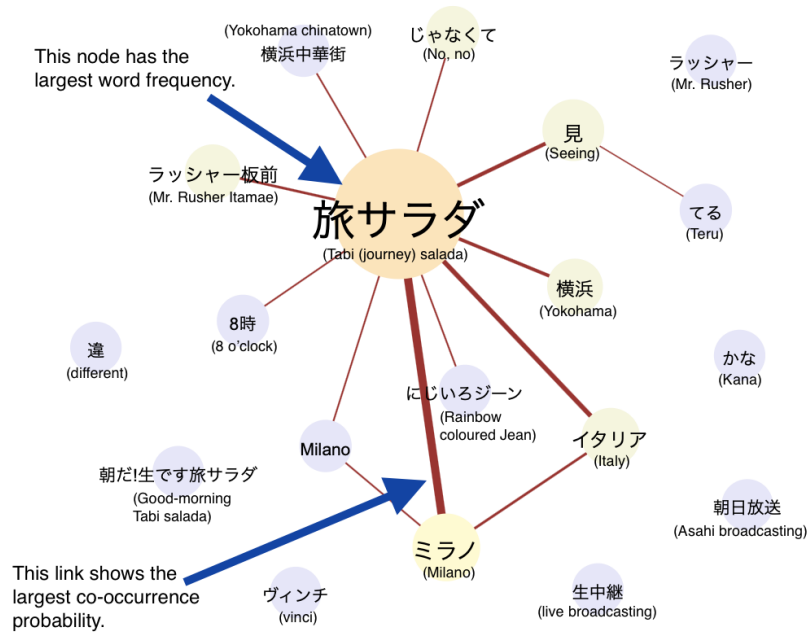

Figure 3. An example of the co-occurrence network graph

4. SYSTEM OVERVIEW

Based on the algorithm mentioned previously, the results show that the system we developed to retrieve trending topics on tweets was able to calculate word frequencies and co-occurrence networks, and the topic structure was visualization.

4.1 System Architecture

An overview of the system is given in Figure 4. The python script `get_trends.py` collected trending topics and tweets. It produced a Ruby script `db/seeds.rb`, which had the seeds to the database entries in the Ruby on Rails application. The system retrieves trending topics and tweets every 20 minutes. It analyzes the topic structure and stores the results in a database. The database stores the information on nodes and links as the co-occurrence network is not created on-demand dynamically.
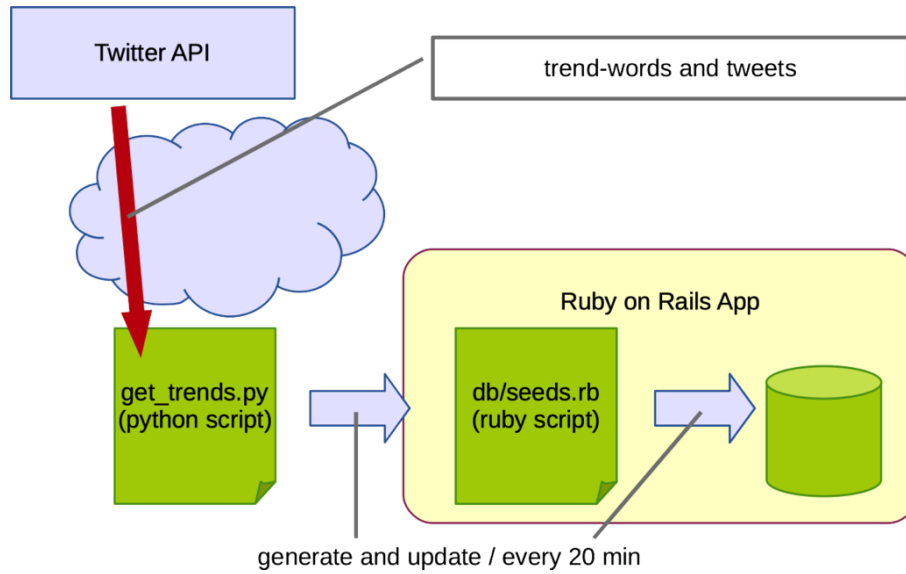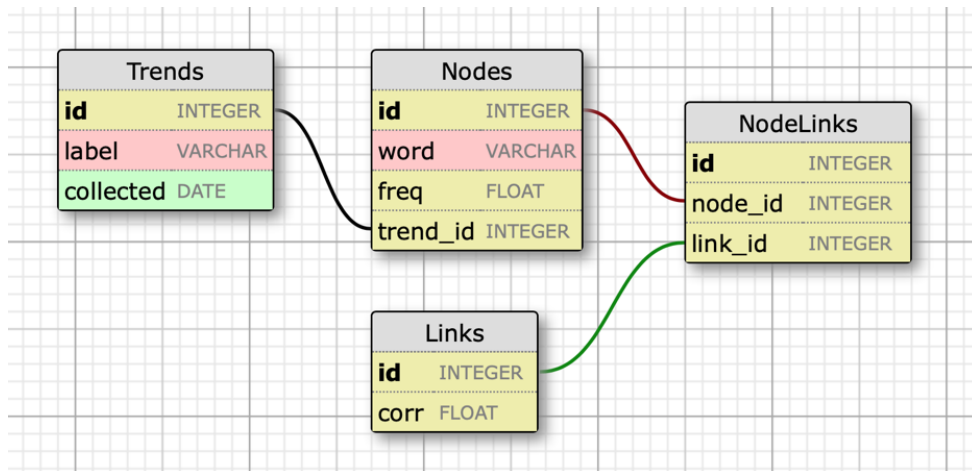
Figure 4. An overview of the system

Figure 5. The schema of the database (which was drawn by WWW SQL designer[2])

The database schema had four tables and three references as shown in Figure 5. The database has four tables to represent the co-occurrence network graph. The table titled *Trends* had the list of trending topics, the tables titled *Nodes* and *Links* had the list of nodes and links shown in every co-occurrence network graph, and the table titled *NodeLinks* had the list of relations between the nodes and the links.

---

[2] http://ondras.zarovi.cz/sql/demo/

110

Presenting the network graphs was implemented using a web application framework called Ruby on Rails (RoR). The python script `get_trends.py` collecting trending topics and tweets did not access the database directly and it generated a Ruby script `db/seeds.rb` which stored the entries of the nodes and links. After the Ruby script was generated, the command "`rails db:seed`" was used and the data entries were stored in the database.

4.2 TWtrends: The Application for Topic Visualization

The implemented web system called TWtrends[3], was created based on Ruby as described previously. The application was deployed to the server and ran in the Apache web server. The application was run through the Phusion Passenger module (Phusion 2019).

The typical use-cases of the system are presented in Figure 6 and the index page that contains a list of the trending topics collected in a day is given on the left. All topics were clickable. When the users clicked any of the topic phrases, page transition occurred, and the co-occurrence network graph was presented (the middle of Figure 6). The network graph was drawn using D3.js (Zhu 2013) so as for them to be clickable and operable allowing for the modification of the shape of the figure. Additionally, the system had a keyword search function. The right part of Figure 6 shows the results of a search.
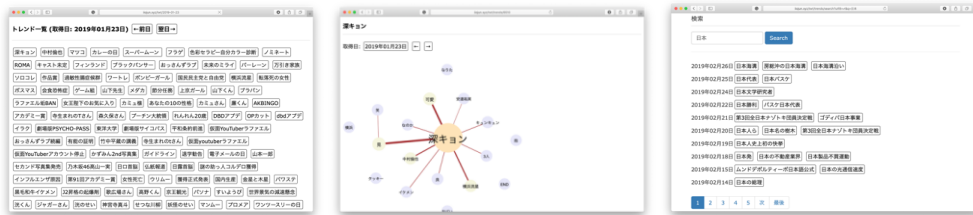

Figure 6. The screenshots of TWtrends: (Left) an index page, (Middle) showing the co-occurrence network graph, and (Right) a result of the keyword search

5. TOPIC MAP GRAPH

This section explains the second visualization based on the co-occurrence network graphs.

5.1 The Similarity of Topics

The system collects over 200 trending topics a day and a review of the topics showed that there were several relationships among some of the topics. For example, the co-occurrence network graph of three topics is given in Figure 7. The graph on the left is for the topic *Aladdin*, the graph in middle id for the topic *Jasmine*, and the graph on the right is for the topic *Jafar*. Aladdin, Jasmine, and Jafar are names of characters of a Walt Disney film called *Aladdin*. These

---

characters as topics were very popular among Twitter users on June 15, 2019, and they came up on Twitter's timeline as the trending topics in the Tokyo metropolitan area.

The three graphs are topologically comparable, and their similarities imply that there are several clusters hidden in the list of the trending topics. This suggests that these three topics arose from tweets mentioning the movie. Therefore, these three topics should be categorized into one cluster.

Figure 7. The co-occurrence network graphs on 'Aladdin' (Left), 'Jasmine' (Middle) and 'Jafar' (Right)

5.2 Topic Map Graph using Cosine Similarity

An example of the topic map graph, which was calculated on January 26, 2019 is given in Figure 8. On that day, the main topic was Ms. Naomi Osaka, a famous Japanese tennis player who won the women's trophy for the Australian Open, a tennis tournament. There was a big cluster with several topics on her win as shown in Figure 8.

Figure 8. An example of the topic map graph calculated on January 26, 2019 (left). The magnified part of the topic-map graph with some English translations is illustrated (right)

Trending topics are associated with the co-occurrence network graphs. Using this relationship, we can build a massive word space which consists of all keywords appearing in the co-occurrence graphs in a day. A topic is considered as a point in the space. In such word spaces, the distance of the two topics can be calculated so that their similarity can be determined.

In this study, we used the cosine similarity as the measure of distance, e.g. if vector a is for topic A and the vector b is for the topic B, the cosine similarity between a and b, which is denoted by sim(A, B), is calculated using the following formula:

$$\text{sim}(A, B) = \frac{a \cdot b}{|a||b|}$$

We defined the threshold value as 0.75 and the nodes that had a similarity value more than the threshold were aggregated in one cluster. A topic map graph was drawn using GraphViz utilities (Ellison 2001) based on the results of the calculation. Figure 9 shows interim results of the calculation of similarities between each node, and a dot script for GraphVis created by the system based on the result of similarities calculations. After the generation of the dot script, an image of the topic map graph is rendered according to the code.
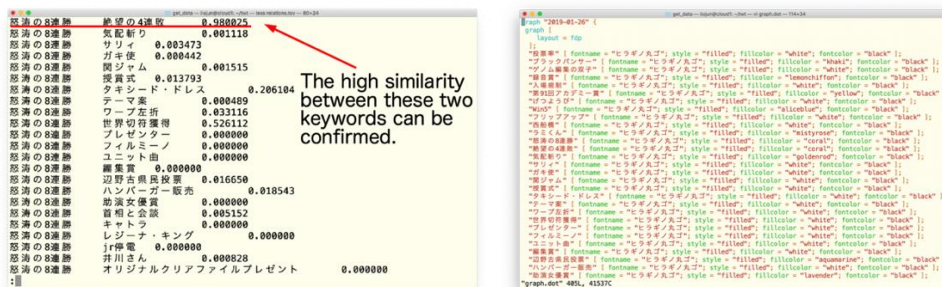
Figure 9. An example of the interim result of the calculation of similarities between each node (left), and a script ('dot script') for GraphViz created by the system (right)

6. DISCUSSION

The system is able to show the most popular topic on a specific day. This section describes several findings from the approximate half-year operation in 2019, and a possibility to modify the system to suit other languages.

6.1 A Distinguishing Social Trend Revealed by the System

Figure 10 is a sequence of topic map graph, and it is considered as the distinguishing example. The transition of the topic map between April 28 and May 4 shown in Figure 10, and it shows that the topic maps on April 30 and May 1 had an unusual shape in comparison to the others.
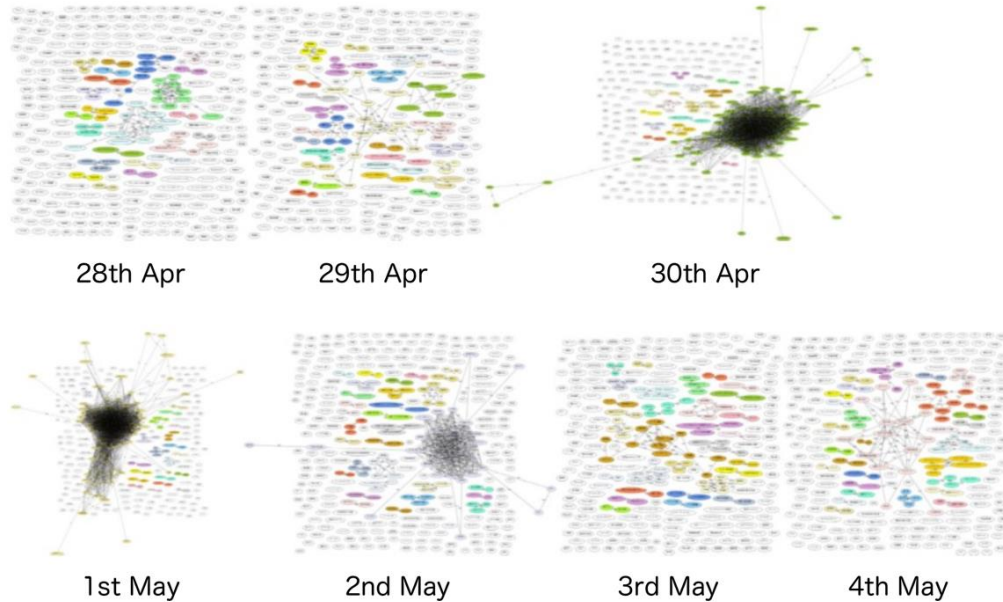
113

Figure 10. A typical example of the sequence of topic map graphs. These graphs were calculated every day between April 28 and May 4, 2019

We investigated the contents of the trending topics carefully, and the item "*the last things* in Heisei era" and similar issues were popular on April 30. Similarly, we found that the topic "*the first things* in Reiwa era" and related matters were popular on May 1. The results presented in Figure 10 showed that the changes in era from Heisei era to Reiwa era was a matter of considerable interest to the Japanese people.

6.2 Clusters within the Topic Map Graph and their Characteristics

According to the discussion in the previous section, we can assume a hypothesis considering a careful investigation of topic maps provided by TWtrends discloses the type of topics Twitter users frequently mention and which tend to be listed as "trending."

TWtrends has an API for third parties to deliver the data stored in the database. Therefore, trending maps were re-calculated by using the API to extract the clusters in the graphs. From the list consisting of date, the number of clusters, and the nodes of the cluster, we extracted a certain level of large clusters. The period of re-calculation was from the January 1 to May 31, 2019.

Average values and variances of the number of nodes, number of clusters, and the size of clusters are summarized in Table 1. We do not have any interests in the clusters containing only one node. Hence, the similar data in clusters containing only one node were excluded, as shown in Table 2.

Table 1. Fundamental data about re-calculated clusters. The number of nodes, number of clusters, and size of the clusters in one day

	Average	Variance	Standard Deviation
Number of nodes	279.6	786.9	28.1
Number of clusters	211.9	638.5	25.3
Size of clusters	1.326	0.009	0.096

Table 2. Fundamental data about re-calculated clusters in cases in which containing only one node were excluded

	Average	Variance	Standard Deviation
Number of clusters	35.1	45.3	6.73
Size of clusters	2.963	0.328	0.573

The cluster size of the largest and the second largest is 94 and 83, respectively; others were less than 35. The largest cluster represented the topic, "*the first things* in Reiwa era," observed on May 1, and the second largest showed the topic, "*the last things* in Heisei era," observed on April 30.
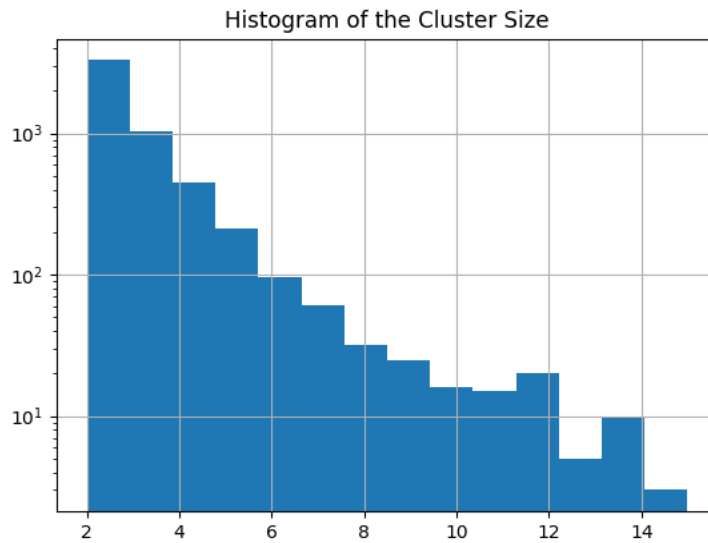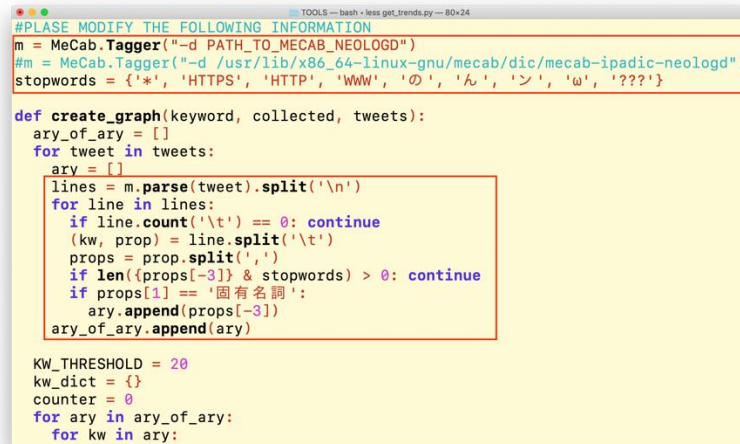

Figure 11. Distribution of the cluster size (from the size of 2 to that of 15)

Figure 11 shows the distribution of the cluster size observed during the re-calculating period. Because many clusters have only one node, and there are few numbers of large-scale clusters, the summations of the counts of clusters whose size is 2 to 15 are plotted in Figure 11. Note that the vertical axis is in log scale due to the large number of small-scale clusters.

Figure 11 implies that the cluster size depends upon exponential distribution. Regarding this phenomenon, a theoretical corroboration will be expected.

6.3 Modification to Process Tweets Written in Other Languages

Current implementation of TWtrends is suitable for analyzing Japanese tweets. However, it would be easily modified to handle tweets written in other languages.

```
#PLASE MODIFY THE FOLLOWING INFORMATION
m = MeCab.Tagger("-d PATH_TO_MECAB_NEOLOGD")
#m = MeCab.Tagger("-d /usr/lib/x86_64-linux-gnu/mecab/dic/mecab-ipadic-neologd")
stopwords = {'*', 'HTTPS', 'HTTP', 'WWW', 'の', 'ん', 'ン', 'ω', '???'}

def create_graph(keyword, collected, tweets):
    ary_of_ary = []
    for tweet in tweets:
        ary = []
        lines = m.parse(tweet).split('\n')
        for line in lines:
            if line.count('\t') == 0: continue
            (kw, prop) = line.split('\t')
            props = prop.split(',')
            if len({props[-3]} & stopwords) > 0: continue
            if props[1] == '固有名詞':
                ary.append(props[-3])
        ary_of_ary.append(ary)

    KW_THRESHOLD = 20
    kw_dict = {}
    counter = 0
    for ary in ary_of_ary:
        for kw in ary:
```

Figure 12. Two red rectangles indicate the code fragments which are dependent on parsing Japanese sentences

Figure 12 shows the code fragment of get_trends.py, which gathers tweets by topics and extracts proper nouns from the tweets. Only the part of the code indicated by the red rectangles depends upon parsing Japanese sentences.

As described in Section 3.4, every word in Japanese sentences is concatenated, without punctuations, which is in contrast to Western languages. Hence, a morphological analysis is needed to separate words and to extract proper nouns. There are several morphological analyzers implemented as the open-source software (Takaoka 2018). Among them, this study chose the famous software *MeCab* (Kudo 2004), the improved dictionary *mecab-ipadic-neologd*, and the python libraries to use them from the python script.

The whole lines of the code get_trends.py are 159, and only 12 lines depends upon Japanese language processing. It suggests that modification to handle other languages is quite easy. If you want to modify the code to analyze the tweets written in English, you can write a simple code to extract proper nouns using the Natural Language Toolkit (NLTK) (Loper 2002), and replace the code shown in Figure 12 with your code.

7. CONCLUSIONS AND FUTURE WORK

In this study, we developed a system to visualize the structure of trending topics by calculating the word frequency and co-occurrence network. Trending topics were collected every 20 minutes, and the results of the calculation were stored in a database. The Visualization system

was built based on the Ruby on Rails framework and the application was built using D3.js and Ajax.

The topic map graphs were constructed based on the word frequencies of the trending topics after calculating co-occurrence network graphs. The topic map graphs have several topic clusters, which were hidden in the list of topics. The analyses of the similarities in topics can show us relationships that allows us to know what topics are popular on SNSs.

The present paper also reports on a distinguishing example observed from the end of April to the beginning of May 2019. A historic event occurred in Japan that week: the change of an era, from Heisei to Reiwa. The magnitude of interest in that event among SNS users was shown in a sequence of topic map graphs, drawn by our system.

Converting the system to enable it to process tweets in other languages will not be a difficult modification, although the current system is set up to process tweets written in Japanese. The script code depends upon language only to extract from the collected tweets a set of keywords, which constitute the candidate of the nodes in the co-occurrence network. As a small part of the process, this step can be easily modified. Our future work will involve making the system multilingual and evaluating it in other cultures overseas.

ACKNOWLEDGEMENT

REFERENCES

@TwitterIR, 2019, Q2 2019 Letter to Shareholders.

Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Goker, A., Kompatsiaris, Y., and Jaimes, A., 2013. Sensing trending topics in Twitter, *IEEE Transactions on Multimedia,* Vol. 15, No. 6, pp. 1268-1282. doi: 10.1109/AICCSA.2014.7073242

Becker, H., Naaman, M., and Gravano, L., 2011. Beyond Trending Topics: Real-World Event Identification on Twitter, *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media,* pp. 438-441.

Benhardus, J. and Kalita, J., 2013. Streaming trend detection in Twitter, *International Journal on Web Based Communities*, Vol. 9, No. 1, pp. 122-139.

Bovet, A., Morone, F., and Makse, H. A., 2018. Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump, *Scientific Reports*, No. 8, Article Number: 8673.

Dawar, A., Purwar, A., Anand, N., and Singla. C., 2018. TweetRush: A tool for analysis of twitter data. *International Journal on Education and Management Engineering*, Vol. 8, No. 2, pp. 31-40, 2018. doi: 10.5815/ijeme.2018 .02.04

Ellson, J., Gansner, E., Koutsofios, L., North, S. C., and Woodhull, G., 2002. Graphviz— Open Source Graph Drawing Tools, *In: Mutzel P., Jünger M., Leipert S. (eds) Graph Drawing. GD 2001. Lecture Notes in Computer Science*, vol 2265. Springer, Berlin, Heidelberg.

Garg, N. and Rani, R., 2017, Analysis and Visualization of Twitter Data using k-means Clustering, *Proceedings on International Conference on Intelligent Computing and Control Systems, ICICCS 2017,* pp. 670-675.

Giridhar, P. and Abdelzaher, T., 2017. Visualization of Events Using Twitter and Instagram, *IEEE International Conference on Pervasive Computing and Communications Demonstrations,* Kona, HI, USA, pp. 82-84.

Khandekar, C., Daiya, R., Parekh, R., and Kelkar, K., 2017. TweetSum: Automated News Summarization of Twitter Trends, *International Journal of Computer Applications*, Vol. 165, No. 8, pp. 5-8.

Kudo, T., Yamamoto, K. and Matsumoto, Y., 2004. Applying conditional random fields to Japanese morpho- logical analysis. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230–237, Barcelona, Spain.

Lewis, H. C., *et al.,* 2015. Social Listening: A Content Analysis of E-Cigarette Discussions on Twitter, *Journal of Medical Internet Research*, Vol. 17, No. 10, e243, doi:10.2196/jmir.4969

Lin, C.-Y., Li, T.-Y., and Chen. P., 2016. An information visualization system to assist news topics exploration with social media. *Proceedings of the 7th 2016 International Conference on Social Media & Society*, SMSociety '16, pp. 23:1-23:9. ACM, New York, NY, USA, doi: 10.1145/2930971.2930995

Loper, E. and Bird, S., 2002, NLTK: The Natural Language ToolKit, Cornell University, arXiv:cs/0205028 [cs.CL].

Maia, A., Cunha, T., Soares, C., and Abreu, P. H., 2016, TweeProfiles3: visualization of spatio-temporal patterns on Twitter, *Rocha Á., Correia A., Adeli H., Reis L., Mendonça Teixeira M. (eds) New Advances in Information Systems and Technologies. Advances in Intelligent Systems and Computing*, vol 444. Springer, Cham, pp. 869-878, doi:10.1007/978-3-319-31232-3_82

Phusion, 2019. *Passenger - Enterprise grade web app server for Ruby, Node.js, Python*, viewed 25 August 2019, https://www.phusionpassenger.com

Sechelea, A., Huu, T. D., Zimos, E., and Deligiannis, N., 2016, Twitter Data Clustering and Visualization, *Proceedings of the 23rd International Conference on Telecommunications (ICT2016)*, Thessaloniki, Greece.

Takaoka, K., Hisamoto, S., Kawahara, N., Sakamoto, M., Uchida, Y., and Matsumoto, Y., 2018. Sudachi: a Japanese Tokenizer for Business, *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC2018),* Miyazaki, Japan.

Wen, D., and Marshall, G., 2014. Automatic Twitter Topic Sum-marization, *The 17th Intternational Conference on Computer Science and Engineering*, Chengdu, China. doi:10.1109/ CSE.2014.69

Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., and Vaidya, J., 2018. Analysis and visualization of subjectivity and polarity of Twitter location data, *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, Article No. 67, Delft, The Netherlands, doi: 10.1145/3209281.3209313

Zhu N. Q., 2013. *Data Visualization with D3.js Cookbock, over 70 recipes to create dynamic data-driven visuzalization with D3.js*, Packt Publishing, Birmingham, UK.

Zubiaga, A., Spina, D., Martınez, R., and Fresno. V., 2015. Real-time classification of twitter trends. *Journal of the Association for Information Science and Technology*, Vol. 66, No. 3, pp. 462-473, 2015. doi: 10.1002/asi.23186