Detection of School Foundation Day Tweets That Can Be Used to Distinguish Senders' Schools

Yasuhiko Watanabe, Hiroaki Onishi, Ryo Nishimura, and Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, Japan
Email: watanabe@rins.ryukoku.ac.jp, t150458@mail.ryukoku.ac.jp, r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp

***Abstract*—These days, many people use a Social Networking Service (SNS). When we use SNSs, we carefully protect the privacy of personal information: name, age, gender, address, telephone number, birthday, etc. However, we sometimes submit online messages that can threaten our privacy and security when combined with other information. In this study, we investigated tweets that can disclose senders' affiliations, especially, high schools to other people, including unwanted audiences, when combined with other public information. We collected 1,000 tweets including word "school foundation day" and found 46% of the collected tweets were ones disclosing foundation days of senders' schools. Furthermore, we found tweets including word "school foundation day" can be used to distinguish senders' schools when combined with event calendars in school web sites. Finally, we obtained 74% accuracy when we applied Support Vector Machine (SVM) to classify tweets including word "school foundation day" into ones disclosing the foundation days of senders' schools.**

Keywords–school foundation day; personal information; Twitter; SNS; privacy risk.

I. INTRODUCTION

These days, many people use a Social Networking Service (SNS) to communicate with each other and try to enlarge their circle of friends. SNS users are generally concerned about potential privacy risks. To be specific, they are afraid that unwanted audiences will obtain information about them or their families, such as where they live, work, and play. As a result, SNS users are generally careful in disclosing their personal information. They disclose their personal information only when they think the benefits of doing it is greater than the potential privacy risks. For example, students are generally careful in disclosing which schools they go to. They are concerned that once unwanted audiences know which schools they go to, the unwanted audiences can obtain several kinds of their personal information: how old they are, where they live and study, who their friends and families are, and when they are at home or away. However, they often submit online messages that threaten their privacy and security when combined with other information. In this paper, we focus on school event messages submitted to Twitter. Student often submit tweets concerning school events, such as sports festival, culture festival, entrance ceremony, graduation ceremony, and foundation day. School event tweets often give students opportunities to start new communications on Twitter. However, they also give anyone, including unwanted audiences, opportunities to distinguish which schools they go to. Take a tweet concerning a school foundation day for example. Figure 1 shows a school foundation day tweet submitted by a high school student, *momone*. We retouched her photos in Figure 1 for protecting students' privacy. (exp 1) is the text of her tweet.

Figure 1. A school foundation day tweet submitted by a high school student, *momone*, on 6th October 2018.

(exp 1) *Souritsu kinenbi ha 6 nin. Nodo ga itaku naru kurai shabette waratta –*
(Six of us on foundation day. We talked so much and our throats were raspy –)

This tweet does not show which school *momone* went to. It only shows when she enjoyed the foundation day of her school. *Momone* might think that this tweet was not enough to distinguish which school she went to. In other words, there were too many schools and it was difficult to find schools whose foundation days were the day or just before she submitted this tweet. However, this tweet can threaten her privacy and security more than she expected. In this paper, we show that this tweet gave a chance to other people, including unwanted audiences, to distinguish which school she went to. In order to discuss the privacy risks caused by school event tweets, we show how school event tweets are used to distinguish which schools students go to. Furthermore, we discuss whether unwanted audiences can collect school event tweets by using machine learning techniques.

The rest of this paper is organized as follows: In Section II, we survey the related works. In Section III, we report school events and how school event tweets, especially school foundation day tweets, are used to distinguish which schools students go to. In Section IV, we discuss whether unwanted audiences can collect school foundation day tweets by using machine learning techniques. Finally, in Section V, we present our conclusions.

 ISBN: 978-1-61208-721-4

II. Related work

Personally identifiable information is defined as information which can be used to distinguish or trace an individual's identity such as social security number, biometric records, etc. alone, or when combined with other information that is linkable to a specific individual, such as date and place of birth, mother's maiden name, etc. [1] [2]. Internet users are generally concerned about unwanted audiences obtaining personal information. Fox et al. reported that 86% of Internet users are concerned that unwanted audiences will obtain information about them or their families [3]. Also, Acquisti and Gross reported that students expressed high levels of concern for general privacy issues on Facebook, such as a stranger finding out where they live and the location and schedule of their classes, and a stranger learning their sexual orientation, name of their current partner, and their political affiliations [4]. However, Internet users, especially young users, tend to disclose personal information on their profiles, for example, real full name, gender, hometown and full date of birth, which can potentially be used to identify details of their real life, such as their social security numbers. As a result, many researchers discussed the reasons why young users willingly disclose personal information on their SNS profiles. Dwyer concluded in her research that privacy is often not expected or undefined in SNSs [5]. Barnes argues that Internet users, especially teenagers, are not aware of the nature of the Internet and SNSs [6]. Hirai reported that many users had troubles in SNSs because they did not mind that strangers observed their communication with their friends [7]. Viseu et al. reported that many online users believe the benefits of disclosing personal information in order to use an Internet site is greater than the potential privacy risks [8]. On the other hand, Acquisti and Gross explain this phenomenon as a disconnection between the users' desire to protect their privacy and their actual behavior [4]. Also, Livingstone points out that teenagers' conception of privacy does not match the privacy settings of most SNSs [9]. Joinson et al. reported that trust and perceived privacy had a strong affect on individuals' willingness to disclose personal information to a website [10]. Also, Tufekci found that concern about unwanted audiences had an impact on whether or not students revealed their real names and religious affiliation on MySpace and Facebook [11]. The authors also think that most of students are seriously concerned about their privacy and security. However, they often underestimate the risk of their online messages and submit them. For example, many students submit online messages concerning school events. Most students do not mind that these messages can threaten their privacy and security. To be specific, these messages may give a chance to other people, including unwanted audiences, to distinguish which schools students go to. However, no studies have been made on the risk of online messages concerning school events.

III. Risk of school event tweets

SNS users want to start new communication and enjoy it. As a result, they want good topics for starting new communication in SNS. For example, school events are good topics for starting new communication in SNS. Actually, many students submit tweets concerning school events. However, school event tweets may threaten students' privacy and security because most of them are submitted during or just after school events. Specifically, unwanted audiences can

- imagine when students in the target school submit school event tweets and
- use the dates of school event tweets as clues to distinguish which schools students go to.

Furthermore, it is easy to obtain school event calendars because many schools show their school event calendars in their web sites. In order to clarify the risks caused by school event tweets, in this paper, we discuss

- school events that are held every year and good topics for tweets, and
- how to distinguish which schools senders go to by using school event tweets and event calendars.

A. Annual school events frequently reported in tweets

Several kinds of events are held in schools. In this section, we discuss

- school festivals,
- school ceremonies, and
- school memorial days.

This is because these events are held every year in most schools in Japan and frequently reported in tweets by students. Furthermore, many schools show the dates of these school events in their web sites.

1) School festivals: We first discuss two major school festivals in Japan: sports festivals and culture festivals. In Japan, school sports festivals are mainly held in spring (May and June) or autumn (September and October). Many students submit tweets concerning sports festivals in their schools. Figure 2 shows a tweet concerning a sports festival submitted by a high school student, *ayane*, on 29th September 2018. On the other hand, in Japan, culture festivals are mainly held from September to December. Culture festivals are chances for students to show what they have learned in the year and create a performance for their parents, teachers, and in some cases, the public. As a result, many students submit tweets concerning culture festivals in their schools. Figure 3 shows a culture festival tweet. It was also submitted by *ayane* on 22nd September 2018.

Students often attach pictures to their tweets concerning sports festivals and culture festivals. As shown in Figure 2 and Figure 3, *ayane* attached many pictures to her sports festival tweet and culture festival tweet. Figure 4 shows pictures attached to *ayane*'s sports festival tweet. In the pictures, there were many students and we found a clue to distinguish which school these students and *ayane* went to: a girl student in the lower left picture wore an athletic uniform with the name of their school. The web site of their school showed that the sports festival and culture festival were held on 28th and 21st October 2018, respectively. As a result, *ayane* submitted her sports festival tweet and culture festival tweet on the next day of the sports festival and culture festival, respectively.

Sports festivals and culture festivals are held every year, however, the dates of them may change every year. For example, culture festival in *ayane*'s school was held on 21st October 2018 while it was held on 22nd October 2017.

 ISBN: 978-1-61208-721-4

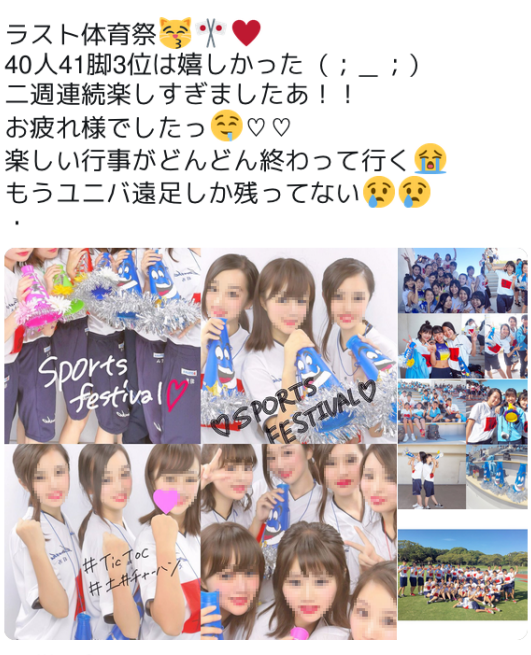

Figure 2. A sports festival tweet submitted by a high school student, *ayane*, on 29th September 2018.

Figure 3. A culture festival tweet submitted by a high school student, *ayane*, on 22nd September 2018.

2) School ceremonies: We discuss two major school ceremonies in Japan: entrance ceremony and graduation ceremony. In Japan, entrance ceremonies are generally held in the first half of April. For example, in 2018, all public high schools in Kyoto held the entrance ceremonies on 9th or 10th April. On the other hand, graduation ceremonies are generally held in March. For example, in 2018, all public high schools in Kyoto held the graduation ceremonies on 1st March. The dates of entrance ceremonies and graduation ceremonies may change every year.

Figure 4. Pictures attached to *ayane*'s sports festival tweet (Figure 2).

Figure 5. A graduation ceremony tweet submitted by a high school student, *nanae*, on 7th March 2019.

Many students and their family members submit tweets with pictures concerning these ceremonies. Pictures attached to school ceremony tweets, just like those attached to school festival tweets, often give chances for readers to obtain several kinds of senders' personal information. Figure 5 shows a graduation ceremony tweet submitted by a high school student, *nanae*, on 7th March 2019. *Nanae* took pictures concerning her graduation ceremony and attached them to her tweet. In the pictures, we found a clue to distinguish which schools *nanae* went to: students were holding a signboard with the name of her school. It shows *nanae* went to the same school as *ayane*.

Entrance ceremonies and graduation ceremonies are held in short periods. Many schools hold these ceremonies on the same day. For example, in 2018, all public high schools in Kyoto held the graduation ceremonies on the same day. As a result, it is difficult to distinguish which schools senders go

 ISBN: 978-1-61208-721-4

Figure 6. A foundation day tweet submitted by a high school student, *ayane*, on 6th October 2018.

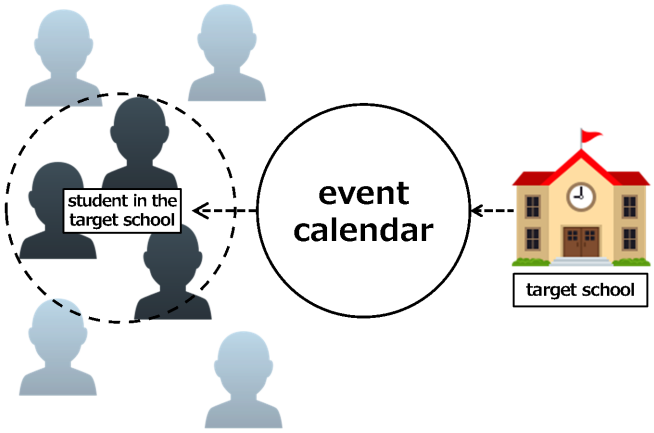

Figure 7. It is easy to detect school event tweets submitted by students in the target school on the day or just after the event was held.

to by using the submission date of tweets concerning entrance ceremonies and graduation ceremonies.

3) School memorial days: We discuss one major school memorial day in Japan, school foundation day. School foundation days are held all year around. Furthermore, school foundation days are fixed while the dates of school festivals and ceremonies may change every year. The important point is that many schools are closed on their foundation days. As a result, many students submit school foundation day tweets in order to show where they go, what they do, and who they are with in the special day. Figure 6 shows a foundation day tweet submitted by *ayane* on 6th October 2018. (exp 2) is a text message in her school foundation day tweet.

(exp 2) *Souritsu kinenbi ha suggoi shiawase na ichinichi yatta. chotto hayame no tanjyoubi oiwai shite kurete konna kawaii purezento moracchatta (;_;) daiji ni tsukawasete itadaki masu.*
(The foundation day was a very happy day. Thank you for an early birthday party and beautiful presents. I'll treasure them.)

B. How to use school event tweets for distinguishing which schools students go to

Many students submit their school event tweets. However, most of them may think that their school event tweets are not enough to distinguish which schools they go to. The reasons are as follows:

- There are many similar online messages. Their messages are not special. No one pays attention to them.
- There are many schools. It is difficult to find schools whose school events were held on the day or just before they submitted their school event tweets.

Take the tweet shown in Figure 3 for example. This culture festival tweet was submitted by *ayane* on 22nd September 2018. Most of students may think that it is hard to visit many school web sites and check whether the culture festival was held on 22nd September or just before. However, it is not difficult to detect school event tweets submitted by students in a particular school. To be specific, it is easy to obtain the event calendar from the web site of the target school and collect school event tweets that were submitted on the day or just after the target school held it. Figure 7 shows the overview of it.

In order to evaluate this method, take a high school in Kobe for example. We visited the web site of the target school and obtained the event calendar. According to the event calendar, in 2018, this school held

- the foundation day on 5th October,
- the sports festival on 28th September, and
- the culture festival on 21st September.

We think, the date of foundation day is more useful than those of sports festival and culture festival to collect tweets submitted by students of the target school. This is because the number of schools that have the same foundation day is less than those that have the same sport festival day or the same culture festival day. School foundation days are held all year round while sport festivals and culture festivals are not. As a result, we planned to collect foundation day tweets submitted on 5th October 2018, the foundation day of the target school, or just after it. In this paper, we used Twitter search and obtained 45 tweets that were submitted on 5th October or 6th October 2018 and included *kaiko kinenbi* (school foundation day) or *souritsu kinenbi* (foundation day). 15 tweets of them had pictures and two of them had pictures of young persons. We focused on these two tweets with pictures of young persons. One was shown in Figure 1 and submitted by *momone* on 6th October. The other was shown in Figure 6 and submitted by *ayane* on 6th October. Furthermore, *momone* submitted

- her sports festival tweet on 29th September (one day after the sports festival was held in the target school)
- her culture festival tweet on 21st September (the day when the culture festival was held in the target school)

Also, *ayane* submitted

 ISBN: 978-1-61208-721-4

- her sports festival tweet (Figure 2) on 29th September (one day after the sports festival was held in the target school)
- her culture festival tweet (Figure 3) on 22nd September (one day after the culture festival was held in the target school)

We think that it is not coincidence that the event dates of the target school were matched with the submission dates of these tweets. As a result, we concluded that *ayane* and *momone* went to the target school. Furthermore, we found a girl student wearing an athletic uniform with the name of the target school in the pictures attached to *ayane*'s sports festival tweet (Figure 2), submitted on 29th September.

We found that *ayane* sent birthday tweets to *momone* on her birthday. *Ayane* was *momone*'s friend. Furthermore, in 2018, not only *ayane* but three other users sent birthday tweets to *momone*. We could not detect these three users because they did not submit foundation day tweets. However, they submitted school festival tweets the dates of which were matched with those of the school festivals in the target school. As a result, we also concluded that they went to the same school as *momone* and *ayane*. Actually, one of them was *nanae*. The pictures attended to *nanae*'s graduation ceremony tweet (Figure 5) and *ayane*'s sports festival tweet (Figure 2) show that *nanae* and *ayane* went to the same school.

IV. Detection of foundation day tweets

In order to collect school event tweets efficiently, it is important to detect them automatically. If school event tweets, especially foundation day tweets, are detected automatically, the detected tweets can be used for the following conflicting purposes.

- We can give warnings to users before they submit their school event tweets that threaten their privacy and security.
- Unwanted audiences can collect school event tweets and obtain several kinds of senders' personal information.

In this section, we discuss whether we can automatically detect foundation day tweets by using machine learning techniques.

In this study, we used the Support Vector Machine (SVM) for data training and classifying. Table I shows feature $s1 \sim s16$ used in machine learning on experimental data. $s1 \sim s7$ were obtained by using the results of morphological analysis on experimental data. In the experiments, we used a Japanese morphological analyzer, JUMAN, for word segmentation of Japanese tweets [12]. $s8 \sim s10$ and $s12 \sim s14$ were obtained by extracting character N-gram from experimental data. Odaka et al. reported that character 3-gram is good for Japanese processing [13]. $s4 \sim s7$ and $s12 \sim s15$ were obtained from first sentences of tweets. This is because, we thought, clue expressions of school events are often found at first sentences of tweets.

In this study, we used 1,000 Japanese tweets including "*kaiko kinenbi* (school foundation day)" for the experimental data. We collected these tweets by Twitter search from July to November 2018. These 1,000 tweets can be classified into two types:

TABLE I. Features used in SVM method for data training and classifying Japanese tweets including word "*kaiko kinenbi* (school foundation day)"

$s1$	word unigrams of the tweet
$s2$	word bigrams of the tweet
$s3$	the number of words in the tweet
$s4$	word unigrams of the first sentence of the tweet
$s5$	word bigrams of the first sentence of the tweet
$s6$	the number of words in the first sentence of the tweet
$s7$	the last word of the first sentence of the tweet
$s8$	character unigrams of the tweet
$s9$	character bigrams of the tweet
$s10$	character 3-grams of the tweet
$s11$	the length of the tweet
$s12$	character unigrams of the first sentence of the tweet
$s13$	character bigrams of the first sentence of the tweet
$s14$	character 3-grams of the first sentence of the tweet
$s15$	the length of the first sentence of the tweet
$s16$	whether the tweet is a reply

TABLE II. The details of the 1,000 Japanese tweets including "*kaiko kinenbi* (school foundation day)" (from July to November 2018).

	FD tweet	others	total
normal tweet	393	309	702
reply	66	232	298
retweet	0	0	0
total	459	541	1,000

- foundation day tweet
  foundation day tweets show when the foundation days of senders' schools are.
- others
  others do not show when the foundation days of senders' schools are although they include word "*kaiko kinenbi* (school foundation day)".

Furthermore, tweets can be classified into three types [14]:

- reply
  A reply is submitted to a particular person. It contains "@username" in the body of the tweet.
- retweet
  A retweet is a reply to a tweet that includes the original tweet.
- normal tweet
  A normal tweet is neither reply nor retweet. Normal tweets are generally submitted to general public.

Table II shows the numbers of normal tweets, replies, and retweets in the 1,000 tweets. Also, it shows the numbers of foundation day (FD) tweets and others in the 1,000 tweets. As shown in Figure II, there were no retweets in the 1,000 tweets. We conducted this experiment using TinySVM [15]. Table III shows the experimental result of the 1,000 Japanese tweets. The experimental result was obtained with 10-fold cross-validation. In order to discuss our method in more detail, we divided the experimental result of the 1,000 Japanese tweets

 ISBN: 978-1-61208-721-4

TABLE III. THE SVM CLASSIFICATION RESULT OF THE 1,000 JAPANESE TWEETS INCLUDING WORD "*kaiko kinenbi* (SCHOOL FOUNDATION DAY)".

human expert result	SVM result		
	FD tweet	others	recall
FD tweet	332	127	0.72
others	132	409	0.76
precision	0.72	0.76	

TABLE IV. THE SVM CLASSIFICATION RESULT OF THE 702 JAPANESE NORMAL TWEETS INCLUDING WORD "*kaiko kinenbi* (SCHOOL FOUNDATION DAY)".

human expert result	SVM result		
	FD tweet	others	recall
FD tweet	308	85	0.78
others	123	186	0.60
precision	0.71	0.69	

TABLE V. THE SVM CLASSIFICATION RESULT OF THE 298 JAPANESE REPLIES INCLUDING WORD "*kaiko kinenbi* (SCHOOL FOUNDATION DAY)".

human expert result	SVM result		
	FD tweet	others	recall
FD tweet	24	42	0.36
others	9	223	0.96
precision	0.73	0.84	

(Table III) into those of 702 normal tweets (Table IV) and 298 replies (Table V).

As shown in Table III, 741 Japanese tweets were classified correctly and 259 tweets incorrectly in this experiment. 332 tweets out of the 741 correctly classified tweets were ones where the foundation days of senders' schools were disclosed. As shown in Table III, both the recall and precision of tweets disclosing the foundation days of senders' schools were 72%. Furthermore, as shown in Table IV and Table V, the precision of normal tweets and replies disclosing the foundation days of senders' schools were 71% and 73%, respectively. Our method is useful for collecting foundation day tweets precisely. As a result, it is easy for unwanted audiences to collect tweets disclosing the foundation days of senders' schools by using our method. On the other hand, the recall of replies disclosing the foundation days of senders' schools was 36%. Our method could not detect many replies disclosing the foundation days of senders' schools. As a result, in order to give warnings to users before they submit their school event tweets that threaten their privacy and security, it is necessary to improve our method.

V. CONCLUSION

In this paper, we investigated school event tweets, especially foundation day tweets and showed that they should be treated carefully. This is because school event tweets can threaten students' privacy and security when combined with event calendars. We showed that foundation day tweets can be collected precisely by using machine learning techniques. On the other hand, anyone, including unwanted audiences, can obtain event calendars easily because they are often available on schools' web sites. In order to discuss how easy is it to obtain school event calendars, we are investigating the percentage of schools that let anyone obtain event calendars through their web sites. Finally, we should note that we retouched all the photos in this paper for protecting students' privacy.

REFERENCES

[1] C. Johnson III, Safeguarding against and responding to the breach of personally identifiable information, Office of Management and Budget Memorandum, 2007. [Online]. Available: http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf [accessed: 2016-10-04]

[2] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," Computer Communication Review, vol. 40, no. 1, 2010, pp. 112–117. [Online]. Available: https://doi.org/10.1145/1672308.1672328 [accessed: 2019-05-10]

[3] S. Fox et al., Trust and Privacy Online: Why Americans Want to Rewrite the Rules, The Pew Internet & American Life Project, 2000. [Online]. Available: http://www.pewinternet.org/2000/08/20/trust-and-privacy-online/ [accessed: 2019-05-10]

[4] A. Acquisti and R. Gross, Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 36–58.

[5] C. Dwyer, "Digital relationships in the "myspace" generation: Results from a qualitative study," in Proceedings of the 40th Annual Hawaii International Conference on System Sciences, ser. HICSS '07. Washington, DC, USA: IEEE Computer Society, 2007, p. 19.

[6] S. B. Barnes, "A privacy paradox: Social networking in the united states." First Monday, vol. 11, no. 9, 2006. [Online]. Available: http://firstmonday.org/article/view/1394/1312 [accessed: 2019-05-10]

[7] T. Hirai, "Why does "Enjyo" happen on the Web? : An Examination based on Japanese Web Culture," Journal of Information and Communication Research, vol. 29, no. 4, mar 2012, pp. 61–71. [Online]. Available: http://doi.org/10.11430/jsicr.29.4_61 [accessed: 2019-05-10]

[8] A. Viseu, A. Clement, and J. Aspinall, "Situating privacy online: Complex perception and everyday practices," Information, Communication & Society, 2004, pp. 92–114.

[9] S. Livingstone, "Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression." New Media & Society, vol. 10, no. 3, 2008, pp. 393–411.

[10] A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield, "Privacy, trust, and self-disclosure online." Human-Computer Interaction, vol. 25, no. 1, 2010, pp. 1–24. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/07370020903586662 [accessed: 2019-05-10]

[11] Z. Tufekci, "Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites," Bulletin of Science, Technology & Society, vol. 28, no. 1, 2008, pp. 20–36.

[12] S. Kurohashi and D. Kawahara, JUMAN Manual version 5.1 (in Japanese). Kyoto University, 2005.

[13] T. Odaka et al., "A proposal on student report scoring system using n-gram text analysis method," The transactions of the Institute of Electronics, Information and Communication Engineers. D-I, vol. 86, no. 9, sep 2003, pp. 702–705.

[14] Y. Watanabe, K. Nakajima, H. Morimoto, R. Nishimura, and Y. Okada, "An investigation of a factor that affects the usage of unsounded code strings at the end of japanese and english tweets," in Proceedings of the Seventh International Conference on Evolving Internet (INTERNET 2015), Oct 2015, pp. 50–55. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=internet_2015_2_40_40038 [accessed: 2019-05-10]

[15] T. Kudoh. TinySVM: Support Vector Machines. [Online]. Available: http://chasen.org/taku/software/TinySVM/index.html [accessed: 2019-05-10]

 ISBN: 978-1-61208-721-4