Journal of Universal Computer Science, vol. 19, no. 16 (2013), 2472-2485
submitted: 15/7/13, accepted: 27/9/13, appeared: 1/10/13 © *J.UCS*

Text Analysis for Monitoring Personal Information Leakage on Twitter

Dongjin Choi
(Chosun University, Gwangju, Republic of Korea
Dongjin.Choi84@gmail.com)

Jeongin Kim
(Chosun University, Gwangju, Republic of Korea
jungingim@gmail.com)

Xeufeng Piao
(School of Computer Science and Technology, Weihai, Shandong, China
hbpark@hit.edu.cn)

Pankoo Kim[1]
(Chosun University, Gwangju, Republic of Korea
pkkim@chosun.ac.kr)

Abstract: Social networking services (SNSs) such as Twitter and Facebook can be considered as new forms of media. Information spreads much faster through social media than any other forms of traditional news media because people can upload information with no time and location constraints. For this reason, people have embraced SNSs and allowed them to become an integral part of their everyday lives. People express their emotional status to let others know how they feel about certain information or events. However, they are likely not only to share information with others but also to unintentionally expose personal information such as their place of residence, phone number, and date of birth. If such information is provided to users with inappropriate intentions, there may be serious consequences such as online and offline stalking. To prevent information leakages and detect spam, many researchers have monitored e-mail systems and web blogs. This paper considers text messages on Twitter, which is one of the most popular SNSs in the world, to reveal various hidden patterns by using several coefficient approaches. This paper focuses on users who exchange Tweets and examines the types of information that they reciprocate other's Tweets by monitoring samples of 50 million Tweets which were collected by Stanford University in November 2009. We chose an active Twitter user based on "happy birthday" rule and detecting their information related to place to live and personal names by using proposed coefficient method and compared with other coefficient approaches. As a result of this research, we can conclude that the proposed coefficient method is able to detect and recommend the standard English words for non-standard words in few conditions. Eventually, we detected 88,882 (24.287%) more name included Tweets and 14,054 (3.84%) location related Tweets compared by using only standard word matching method.

Keywords: Text Analysis, Personal Information leakage, Social Network Services, Twitter, Personal Identifiable Information
Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

[1] Corresponding author

1 Introduction

With rapid advances in mobile internet systems and smartphones, people can find and share information without time or location constraints. Before the advent of these technologies, which was only a few years ago, people had to go home or internet cafes to search for information or upload contents. However, there is no longer any need to use desktops because people can easily obtain and share diverse information by using smartphones through mobile web browsers or social networking services (SNSs). The SNSs have provided a convenient and useful platform for obtaining information and knowledge, among others. People are willing to upload their experience gained during their travel or knowledge from research. Despite of this great convenience, leakages of private information are highly likely. The problem is that this leakage has been exacerbated by SNSs, which provide users with an online web platform for engaging in various social activities. SNSs users share interests, activities, knowledge, and events, among others, to strengthen their social relationships with others anytime, anywhere. When the U.S. Airways jet crashed into the Hudson River on January 15, 2009, Twitter was the first to show a photo of the crash, beating even local news media outlets. This event highlights that Twitter is not only an SNS but also an important part of the media. Although the speed at which information is exchanged through social media has many benefits, there are also some negative consequences. Tweets can spread across the world within few seconds, and this can pose serious challenges. The main reason why Twitter data are particularly dangerous to organizations is that anyone with a Twitter account can access any other users by using the "following" function, which requires no permission from the target user for accessing his/her public timeline (or Twitter messages). Based on the Twitter policy, if user *A* sends a following request to user *B*, then *A* is automatically authorized to access the public timeline of *B*.

Assuming that user *C* sends a Tweet to friends to share information that it is his/her birthday tomorrow or that user *D* sends a Tweet to celebrate his/her friend's birthday. In this case, although the date of birth is personal information, users typically reveal the date of birth unintentionally. Here the problem is that as long as user *E* is following user *C* or *D*, user *E* can access personal messages between user *C* and *D*. In addition, users send Tweets to others by using their real name or place of residence. This is the main reason why there is a critical need to monitor leakages of personal information on Twitter. Many researchers have suggested that SNSs have the great potential to reveal unknown personal attitudes or sentiments but that they remain problematic because of information leakages. If personal information is revealed by users with inappropriate intentions, there may be serious problems such as online stalking. To address this problem on Twitter, this study examines Twitter data to reveal various hidden patterns related to personal information. For this, we focus on Twitter users who reciprocate others' Tweets by monitoring samples of 50 million Tweets collected by Stanford University in November 2009 and define simple syntactic patterns to uncover the date of birth in Tweets. We also proposed a coefficient method for detecting and recommending personal name and location information if it is not in a standard English form. And compared with other coefficient approaches and give you comparisons.

The rest of this paper is organized as follows: Section 2 provides a review of previous research. Section 3 explains the method for monitoring leakages of personal information on Twitter based on syntactic patterns. Section 4 gives experiment for detecting personal private information, and Section 5 concludes with some suggestions for future research.

2 Previous Research

Digital personal information or personally identifiable information (PII) is always secured from other users. PII is information for uniquely identifying or distinguishing a single individual's identity. Here the individual's full name, national identification number, driver license number, credit card number, and date of birth, among others, are commonly used to distinguish his or her identity [Krishnamurthy, 10]. For example, if there is an individual who wants to withdraw a huge amount of money from his or her bank account, then he or she will be asked for a valid piece of PII to authenticate his or her identity. In addition, when individuals forgot their passwords for certain websites, they are asked for PII data such as their date of birth and e-mail address. Such digitized data can be easily duplicated for others and are more likely to be exposed to others than traditional physical resources [Yim, 12]. Therefore, they can be a source of serious problems if digital PII is revealed to those with inappropriate intentions either intentionally or accidently [Sasaki, 12].

In recent years, many researchers have explored various ways to prevent leakages of personal information not only for traditional websites but also SNSs. Packets containing encrypted messages are generally considered as an important factor in improving security for personal information by monitoring transferred packets in the network [Choi, 06]. In addition, previous research has proposed a conceptual model to support the supply chain and better understand how organizations' confidential information may be leaked [Zhang, 11]. To infer private information on SNSs, Facebook data based on the Naïve Bayes classification method have been used to predict privacy-sensitive trait information exchanged between users [Lindamood, 09] by examining the relationships between users. Here the finding shows that personal information can be leaked to unknown users. A new architecture for protecting private information on Facebook has been proposed to mitigate privacy risks. [Lucas, 08] presented a new architecture for protecting personal private information published on Facebook for mitigating the privacy risks. In addition, previous research has traced the social footprint of a user's profile and found leakages of a diverse range of personal information on various SNSs such as Flickr, LiveJournal, and MySpace [Irani, 11]. Here an issue of great concern is the involuntary leakage of information on SNSs [Lam, 08].

Users are paying closer attention to prevent leakages of their personal information when making web documents. However, the problem with SNSs is that they provide wide freedom and vast metadata and thus that it is relatively easy to infer users' personal information. SNS users are likely to reveal their private information unintentionally, and this is why this paper focuses on Tweets in the Twitter public timeline to examine how users reciprocate their personal information with others without paying close attention. Even they do beware of it, private information is leaking unintentionally.

Paper	Method	Data sources
[Choi, 06]	Packet monitoring	Personal info, Transfer log, queries
[Lucas, 08]	Detecting transmission messages by using AES Approach	Transmission messages
[Lindamood, 09]	Monitoring social links	Friendship links
[Lam, 08], [Zhang, 11]	Monitoring messages in Social Network Services	Profiles of users Time lines of SNSs
[Kiyomoto, 11]	Generalization method	Queries and database responses

Table 1: Comparison of other researches methods and data sources

Although other researches analyzed time lines or messages in SNSs, they did not consider the fact that there is a high possibility users are likely to make misspelled words when they text messages. Therefore, we need to monitor the text messages even though they were in non-standard English form. This is one of the major contributions of our research.

3 Monitoring Personal Information Leakage on Social Network Services

In order to detect personal information such as date of birth, place to live and real name which might be exposed via Twitter, we propose a system to detect personal information leakage by using syntactic patterns which are described in table 4 for detecting date of birth information. We also propose a method to detect location and name information by using 3,604 town information and 10,528 name lists described in Wikipedia followed by figure 1. Although we have town and name lists for detecting personal information leakage, it is guarantee that we can detect those information well due to the fact that people are likely to make a misspell in SNSs. Therefore we need to normalize non-standard word into standard English words by using diverse coefficient approaches before detecting private information as shown in figure 1.

3.1 A Method for Detecting Date of Birth

This section describes a method for monitoring leakages of personal information on Twitter. Twitter, one of the most popular SNSs, is a microblogging service for sharing information by sending text-based messages restricted to 140 characters, namely Tweets [Kwak, 10, Java, 07]. Unlike in the case of most SNSs, including Facebook and MySpace, Twitter users can freely follow others or be followed. Most SNSs require a user to obtain permission from another user whose webpage he or she wishes to access, but this is not the case on Twitter. According to the Twitter policy, being a follower of a particular user on Twitter means that the follower can access all of the Tweets from the user [Kwak, 10]. This guarantees freedom for anyone sharing information. However, personal information should not be revealed in this process, and therefore this paper proposes a method for detecting personal information from unknown users. For this, we first define syntactic patterns to detect the date of birth in

Tweets. When an individual celebrates someone's birthday through Twitter or other SNSs, the text message normally includes the keywords "birthday" and "b-day." Based on this assumption, we can identify Tweets containing these keywords from a huge Twitter data set by taking a simple text-matching approach. As shown in Table 2, the Twitter data set (8.27 GB) collected by Stanford University [Yang, 11] provides information on the time of the Tweet, the user, and the Tweet.

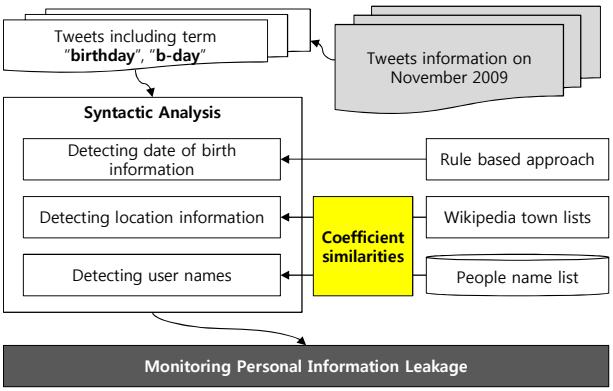

Figure 1: Processes for detecting personal information on Twitter

Type	Information
T	2009-11-06 21:12:16
U	http://twitter.com/femisex
W	RT TheNewAgenda Fort Hood soldier is a true heroine http://bit.ly/1GHi1L
T	2009-11-06 21:12:25
U	http://twitter.com/rudyalba1
W	@ninfest I feel unfaithful I just bought a PSPGO and I like it.
T	2009-11-06 21:28:25
U	http://twitter.com/chacedollars
W	Enjoying my time in Dc @ Schott conference...

Table 2: Examples of the Twitter data set collected by Stanford University

Here T denotes the time at which a given Tweet was uploaded on Twitter; U indicates the ID of the Twitter user who wrote that Tweet; and W represents the Tweet by U at T.

To uncover the pattern hidden in the natural language on Twitter, we simply extract Tweets that include only "birthday" and "b-day" from the data set. There are about 189,000 Tweets, and the total size is only about 30 MB. Table 3 shows some of the extracted Tweets.

Type	Information
T	2009-11-02 05:47:11
U	http://twitter.com/daffaneyb00
W	A special day! Happy Birthday to me baby!!!
T	2009-11-02 05:47:21
U	http://twitter.com/monicabhasin
W	Happy Birthdayyyyyyyyyyyyyyy SRK.. God bless you :D
T	2009-11-02 05:48:09
U	http://twitter.com/teamkatelyn
W	HAPPY BIRTHDAY KATELYN!!! (on the east coast anyways) @katelyntarver

Table 3: Examples of the extracted tweets from Twitter data set

Twitter has several functions, including "RT," "@," and "#" among others. Here "RT" refers to retweets; "@" to a specific user (user ID) who is the target of the Tweet; and "#" to hashtags, which represent keywords or Tweet topics. Here we define syntactic patterns to infer the date of birth from the extracted Tweets (see Table 4). To define these patterns, we select 1,000 Tweets randomly and check those Tweets manually.

Index	Information
1	***someone* happy birthday (b-day) to *someone***
Ex)	@Anissarachma happy birthday to you Frank happy happy birthday from my heart happy birthday to my cousin Nathan and @Franklero
2	***someone* happy birthday (b-day) *someone***
Ex)	@LauraThomas34 happy birthday friend @kerronclement happy birthday Lolo happy birthday
3	***date* is *possessive* birthday (b-day)** or **it is *possessive* birthday**
Ex)	@justinbieber today is my b-day @JackAllTimeLow today is my birthday looooh @Teairra_Monroe Its my birthday and
4	**wish *someone* (a) birthday**
Ex)	@drew's birthday! Everyone wish him happy bday will you please wish me a happy birthday @MixMastaMario I wish her a happy birthday

Table 4: Syntactic patterns for detecting date of birth

According to the patterns in Table 4, patterns 1 and 2 celebrate someone else's birthday; pattern 3, the user's own birthday; and pattern 4, both. In the case of patterns 1 and 2, the first "someone" indicates the name of the user, that is, his or her ID or @user ID. The second "someone" represents the name of the user, that is, his or her ID, @user ID, or the pronoun. To detect the user's name in Tweets, we collect a list of all names for males and females from a website containing 10,532 names. In addition, it is easy to detect the user ID because Twitter users are likely to add the "@" function when sending Tweets. Therefore, if a Tweet satisfies patterns 1 and 2 when

the user ID comes first, then the upload date of the Tweet is the date of the user's birthday. In the case of pattern 3, the important factors in determining the date of someone's birthday are date and possessive terms. The date terms include, among others, "today," "tomorrow," and those referring to specific dates. The possessive terms include, among others, "my," "his," "her," and those for a particular name or user ID. In pattern 4, the term "someone" can be a name, user ID, or objective term such as "me," "her," or "him." To identify the date of birth on Twitter, we develop a simple extraction program based on these patterns by using Python.

Algorithm for detecting date of birth information

```
T_j = Set of Tweet text messages
U_i = Set of userIDs or user names
Bday = date
posv = "my, his, her, someone's"

def pattern_1 ( ):
    U_i happy birthday (b-day) to U_i
def pattern_2 ( ):
    U_i happy birthday (b-day) U_i
def pattern_3 ( ):
    Bday is posv birthday (b-day) or it is posv birthday
def pattern_4 ( ):
    wish U_i (a) birthday

for all subsets T_j :
    if T_j in pattern_1 or pattern_2 :
        if U_i comes in front :
            retrun Bday to U_i
    if T_j in pattern_3 :
        retrun Bday to posv
    if T_j in pattern_4 :
        U_i = userID or name or possessive
        return Bday to U_i
end for
```

3.2 A Method for Detecting Location Information

This section describes a method for detecting information related with places where people might be staying currently. According to previous section, we obtained 24,922 user's tweets (29,399) which exposed date of birth information. The reason why we choose the tweets of users who exposed date of birth information is that there is a high possibility to expose other personal information again and again. The problem is that users are likely to mention where they are intentionally or unintentionally via their Twitter account in order to share their currently location information with their

friends. If someone such as attackers obtains current place information of normal users, it will be a serious problem.

In order to detect place information in tweets, we collected 3,604 town information around world described in Wikipedia. As we can see in table 2, 3, and 4, people do not always wrote a text message in standard English form. Many words are likely to be described in non-standard words such as contraction, abbreviation, acronym, mixed words, funny spelling and misspelling words. Therefore, we need to normalize non-standard words into formal English words for detecting location information. The following table 5 shows the examples of non-standard words.

Taxonomy	**Information**
Contraction	I'm, can't, won't, haven't, …
Abbreviation	uni., dept., ref., max., …
Acronym	SCUBA, ROM, NLP, NER, FBI, …
Mixed	WS99, x220, MS-Dos, ...
Funny spelling	cooool, hooooot, lol, :), b4, …
Misspelling	geogaphy, univercity, knowlege, …

Table 5: Examples of non-standard words

In order to find place information from the given tweets messages, we applied Dice, Jaccard, Ochiai, and proposed coefficient methods. In this research, we only focus on misspelling and funny spelling words.

Dice coefficient is a statistic approach for comparing the similarity of two samples developed by Lee Raymond Dice [Dice, 1945]. Let us assume that we have two samples "havent" and "haven't". Bigrams of these two words can be described by as follows: $bigram_{\text{havent}} = \{ha, av, ve, en, nt\}$ and $bigram_{\text{haven't}} = \{ha, av, ve, en, n', 't\}$. Dice coefficient of these two words can be calculated by following the equation 1.

$$Dice\_coeff(w_i, w_j) = \frac{2 \times |bigram_{w_i} \cap bigram_{w_j}|}{|bigram_{w_i}| + |bigram_{w_j}|} \tag{1}$$

Where, $|bigram_{w_i}|$ and $|bigram_{w_j}|$ are the number of total bigram of given words for $|w_i|$ and $|w_j|$. $|bigram_{w_i} \cap bigram_{w_j}|$ denotes the number of bigrams which appeared in $|w_i|$ and $|w_j|$ at the same time.

Jaccard coefficient which was developed by Paul Jaccard is statistic used for measuring similarity and diversity of samples followed by the equation 2.

$$Jaccard\_coeff(w_i, w_j) = \frac{|bigram_{w_i} \cap bigram_{w_j}|}{|bigram_{w_i} \cup bigram_{w_j}|} \tag{2}$$

Ochiai coefficient or also known as Ochiai-Barkman coefficient, or Otsuka-Ochiai coefficient was considered as a superior coefficient measurement in research [Jackson, 89] which can be calculated by the following equation 3.

$$Ochiai\_coeff(w_i, w_j) = \frac{|bigram_{w_i} \cap bigram_{w_j}|}{\sqrt{|bigram_{w_i}| \times |bigram_{w_j}|}} \quad (3)$$

Our proposed coefficient measurement is performed well when two given words has the same number of characters such as words {“Seoul” and “Seuol”} followed by the equation 4.

$$Proposed\_coeff(w_i, w_j) = \frac{|bigram_{w_i} \cap bigram_{w_j}|}{\text{Ave}(|bigram_{w_i}| + |bigram_{w_j}|)} \quad (4)$$

We considered a tweet which has terms {“live in,” “stay in, ” and “fly to”} as a candidate message might expose location information. The following table 6 shows the examples of extracted cadidate messages

Index	Tweet message
1	@jes2go that's what happens when you ***live in*** Hooterville.
2	@McFly_xX hey :D in my history class today,we were talking about Israel,and I said about you haha ! but,do you ***live in*** Jerusalem ? :D xx
3	@JessicaG_JB you're soo lucky u ***live in*** toronto .. everyone goes there :P nobody comes here !! :P
4	RT @lancearmstrong: Driving to the airport now to ***fly to*** LA to receive American Cancer Society's Medal of Honor... http://bit.ly/1hpcqY
5	I envy my grandmom so much. The day after tomorrow she'll go to korea, and when she returned she would immediately ***fly to*** new zealand. -_-"
6	@DonnieWahlberg Donnie i hate FLYING but what an i going to do Im going to ***fly to*** MONTREAL to see YOU!!!! TWUGS!

Table 6: Examples of candidate tweet messages

As we can see in table 6, people are likely to expose private information via Twitter although strangers can obtain their information. The first row indicates that user *jes2go* used to live in Hooterville but not anymore. The second row has information that user *McFly_xX* used to live in Israel but now he might be staying in Jerusalem and the third row gives information that user *JessicaG_JB* is living in Toronto currently. The fourth low has interesting information that user *lancearmstrong* is driving to the airport and he/she will be in Los Angeles for few days. In other words, he/she will not be in his/her home for few days. The fifth and sixth rows are the same as the fourth low that we can infer whether given users are in there home for next few days or not easily. Therefore, we want to detect this private information on Twitter for giving a warning to users. The following table 7 shows how we obtain place information from given Twitter messages which contains non-standard words.

Algorithm for detecting location information

```
1:  T_j = Set of Tweet text messages
2:  U_i = Set of userIDs or user names
3:  L_t = Set of place names
4:  for all subsets T_j:
5:      if "live in" or "stay in" or "fly to" in T_j :
6:          save T_j , L_t , U_i
8:  end for

9:  for all subsets T_j:
10:     if "fly to" in T_j :
11:         find U_i. L_t ( )
12:         if U_i. L_t [n] != U_i. L_t [n + 1]
13:             return "warning"
14: end for
```

4 Experiment

The test data set was collected by Stanford University on November, 2009 and it contains text, user ID, and time information approximately 50 million Tweets. Firstly, we obtained Tweets and their user IDs who exposed birthday related information by using syntactic patterns. The total number of tweets which satisfied this condition is 29,399. The reason why we choose the tweets of users who exposed date of birth information is that there is a high possibility to expose other personal information again and again.

NSW indicates a word which might represent certain place and suggested words are recommended words by using *PyEnchant*[2] module which is a spellchecking library for Python.

According to the results shown in table 7, names of towns or cities are not always described in a standard English form. For example, one of the well-known cities in Australia "Brisbane" can be misspelled like "Brisbane," "brizban," "brizbnae," and more when people typed in by using Smartphone devices which have limited input system. The coefficient values between "#brisbane" and suggested words {"Brisbane," "disbandment," "briskness," "disbarring"} are shown at the first row in table 7. We can see that the most adequate standard word for "#brisbane" is "Brisbane." However, those four kinds of coefficient methods have failed to distinguish the standard word for "Brizbnae" shown in the third row. The important thing is that all of three coefficient methods have failed to find the standard word for "montrael" except our proposed method has succeeded shown in sixth row. This is the main contribution of this research that our proposed coefficient method can find the standard English words when other coefficient methods cannot.

[2] http://pythonhosted.org/pyenchant/

NSW	Suggested Words	Dice	Jaccard	Ochiai	Proposed
#brisbane	**Brisbane**	**0.933**	**0.875**	**0.935**	**0.8**
	disbandment	0.444	0.285	0.447	0.555
	briskness	0.5	0.333	0.5	0.625
	disbarring	0.470	0.307	0.471	0.705
#brizban	whizzbang	0.533	0.363	0.534	0.533
	Brisbane	**0.571**	**0.4**	**0.571**	**0.571**
Brizbnae	Brigandage	0.25	0.142	0.251	0.625
	Brisbane	0.285	0.166	0.285	0.714
	Brittaney	0.266	0.153	0.267	0.666
	Brianna	**0.461**	**0.3**	**0.462**	**0.769**
montreol	**Montreal**	**0.714**	**0.555**	**0.714**	**0.857**
	Montrachet	0.5	0.333	0.503	0.75
	Montmartre	0.625	0.454	0.629	0.625
	Monterrey	0.533	0.363	0.534	0.8
montrel	**Montreal**	**0.769**	**0.625**	**0.771**	**0.923**
	Montrachet	0.533	0.363	0.544	0.8
	Montmartre	0.666	0.5	0.680	0.666
	Montpelier	0.533	0.363	0.544	0.666
montrael	Montrachet	**0.625**	**0.4545**	**0.629**	0.875
	Montreal	0.571	0.4	0.571	**1.0**
	Montserrat	0.5	0.333	0.503	0.875
	Montmartre	0.5	0.333	0.503	0.75
melburne	**Melbourne**	**0.714**	**0.555**	**0.721**	**0.857**
	melamine	0.307	0.1818	0.308	0.461
	Swinburne	0.428	0.2727	0.433	0.485
	Burnett	0.5	0.333	0.5	0.666

Table 7: Coefficient results between NSW and suggested words for detecting location information

In order to detect location related information, we used town lists defined in Wikipedia. If location related words are detected, we conducted coefficient similarities for normalizing non-standard words into standard English forms because there are many misspelled words in SNSs. The total number of tweets which contain location terms is 11,408 (over 38%). We only consider place information that closely related to current user location. Therefore, we only considered tweets included “live in,” “stay in,” and “fly to.” If we detect name related information, we also take the same step conducted for detecting location related information. Therefore, we acquired 11,425 (over 38%) tweets contained people name information as shown in figure 2.

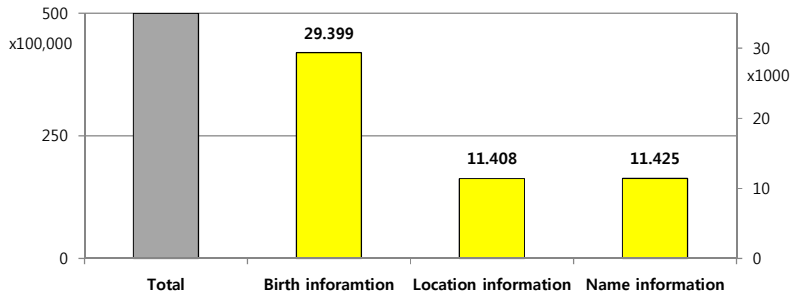

Figure 2: Graphs of detecting personal information

Consider the Tweet "RT @rhonda_Happy birthday @JanetRN" by the user "apostlethatroks." This Tweet indicates that "apostlethatroks" sends a celebration message originally written by the user "rhonda" to the user "JaneRN." That is, it is the birthday of "JanetRN." However, the problem is that there is no guarantee that the extracted date of birth is the actual date of the user's birthday because the proposed syntactic patterns cannot represent various types of written messages. To test whether the proposed method can accurately detect the date of birth, we randomly select 100 observations to manually compare its accuracy. As a result, we can infer 61% of all dates from the test data set with 75% accuracy, although we define only four types of syntactic patterns.

In order to test the performance of the proposed method, we conducted another experiment to compare how much private information can be detected by using only standard words and non-standard words both. As we can see in the figure 3, we can detect 233,019 (63.672%) personal name related tweets which is more than 144,137 (39.385%) by considering only standard words.

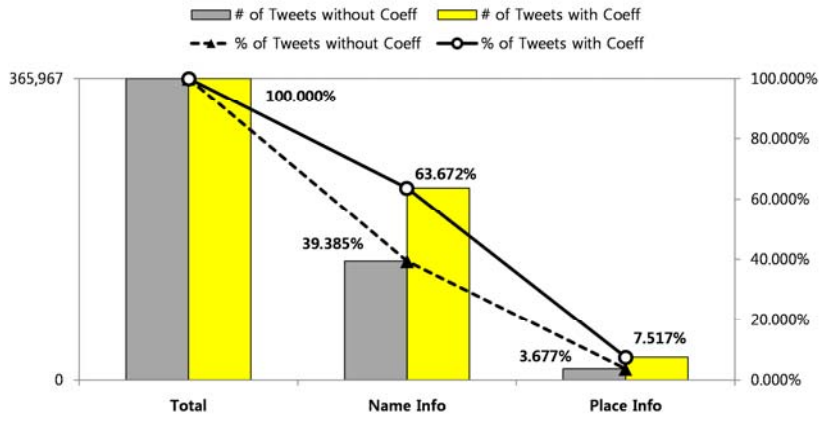

Figure 3: The results of detecting personal information by using standard words and non-standard words

The total number of test data was 365,967 which were the Tweets of 5,000 random users who exposed their birthday related information. When we conducted the experiment to detect name related information by using only standard words, we only can detect 144,137 Tweets which is 63.672% of the test data set. However, when we applied our proposed method for detecting name related information by also using non-standard words, we can detect 233,019 Tweets which is the 63.672% of test data set. 13,456 (3.677%) number of location related Tweets were extracted by using only standard words but we can detect 27,510 (7.517%) number of Tweets by using the proposed method. This is the major contribution of this paper that if we applied a method to normalize non-standard words for detecting personal information, we can detect much more information via Twitter due to the fact that people are likely to text a message in non-standard words.

5 Conclusions

This paper proposes a method for monitoring leakages of personal information on Twitter by inferring the date of birth based on a set of four syntactic patterns, location and name related information by using town and people name lists in Wikipedia and proposed coefficient measurement. Users can upload a diverse range of information on SNSs with no time or location constraints, which entails not only benefits but also some serious challenges. The serious problem is that users are likely to expose their personal information unintentionally via SNSs. In this regard, this paper proposes simple syntactic patterns and our proposed coefficient measurement to detect private information. Given that only four types of patterns are applied, the inference rate for the test data set is considered to be acceptable. Although, extracted private information is not in standard English form, we can detect it by using the proposed coefficient method for normalizing into standard words. However, it is difficult to identify users when test Tweets include pronouns or possessive/objective terms so syntactic patterns we proposed in this paper have to be improved in the future. Future research should also take a named-entity disambiguation approach for improved performance.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2013R1A1A2A10011667) and also financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation.

References

[Choi, 06] Choi, D., Jin, S., Yoon, H.: A personal Information Leakage Prevention Method on the Internet, IEEE 10th International Symposium on Consumer Electronics, pp. 1-5 2006.

[Dice, 1945] Dice, L.R.: Measures of the Amount of Ecologic Association Between Species, Ecology 26 (3): 297–302 1945.

[Irani, 11] Irani, D., Webb, S., Pu, C., Li, K.: Modeling Unintended Personal-Information Leakage from Multiple Online Social Networks, IEEE Internet Computing, vol. 15, no. 3, pp. 13-19 May 2011.

[Jackson, 89] Jackson, D.J., Somers, K.M., Harvey, H.H.: Similarity Coefficient: Measures of Co-occurrence and Association or Simply Measures of Occurrence?, The American Naturalist, vol. 133, no. 3, pp. 436-453 1989.

[Java, 07] Java, A., Song, X., Finin, T., Tseng, B.: Why We Twitter: Understanding Microblogging Usage and Communities, In Proc. of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56-65 2007.

[Kiyomoto, 11] Kiyomoto, S., Martin, K.M.: Model for Common Notion of Privacy Leakage on Public Database, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), vol. 2, no. 1, pp. 50-62 March 2011.

[Kwak 10] Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a Social Network or a News Media, 19th International Conference on World Wide Web, pp. 591-600 2010.

[Lam, 08] Lam, I.F., Chen, K.T., Chen, L.J.: Involuntary Information Leakage in Social Network Services, Proceedings of the 3rd International Workshop on Security: Advanced in Information and Computer Security, pp. 167-183 2008.

[Lindamood, 09] Lindamood, J., Heatherly, R., Kantarcioglu, M., Thuraisingham, B.: Inferring private information using social network data, Proceedings of the 18th international conference on World wide web, pp. 1145-1146 2009

[Lucas, 08] Lucas, M.M., Borisov, N.: flyByNight: Mitigating the Privacy Risks of Social Networking, Proceedings of the 7th ACM workshop on Privacy in the electronic society, pp. 1-8 2008

[Sasaki, 12] Sasaki, T.: A Framework for Detecting Insider Threats using Psychological Triggers, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), vol. 3, no. 1/2, pp. 99-119 March 2013

[Yang, 11] Yang, J., Leskovec, J.: Patterns of Temporal Variation in Online Media, ACM International Conference on Web Search and Data Mining, pp. 177-186 2011

[Yim, 12] Yim, G., Hori,Y.: Guest Editorial: Information Leakage Prevention in Emerging Technologies, Journal of Internet Services and Information Security, vol. 2, no. 3-4, pp. 1-2 November 2012

[Zhang, 11] Zhang, D.Y., Zeng, Y., Wang, L., Li, H., Geng, Y.: Modeling and evaluating information leakage caused by inference in supply chains, Computers in Industry, vol. 62, no. 3, pp. 351-363 April 2011