A Nlp-based Solution to Prevent from Privacy Leaks in Social Network Posts

Gerardo Canfora, Andrea Di Sorbo, Enrico Emanuele,
Sara Forootani, Corrado A. Visaggio
University of Sannio, Benevento, Italy
Email: {canfora,disorbo}@unisannio.it, enricoemanuele@gmail.com, {forootani, visaggio}@unisannio.it

ABSTRACT

Private and sensitive information is often revealed in posts appearing in Social Networks (SN). This is due to the users' willingness to increase their interactions within specific social groups, but also to a poor knowledge about the risks for privacy. We argue that technologies able to evaluate the sensitiveness of information while it is being published could enhance privacy protection by warning the user about the risks deriving from the disclosure of a certain information. To this aim, we propose a method, and an accompanying tool, to automatically intercept the sensitive information which is delivered in a social network post, through the exploitation of recurrent natural language patterns that are often used by users to disclose private data. A comparison with several machine learning techniques reveals that our method outperforms them, since it is more precise, accurate and not dependent on (i) a specific training set, or (ii) the selection of particular features.

CCS CONCEPTS

• **Security and privacy → Privacy protections**; • **Computing methodologies → Natural language processing**;

KEYWORDS

Security and Privacy, Natural Language Processing, Empirical Study

ACM Reference Format:
Gerardo Canfora, Andrea Di Sorbo, Enrico Emanuele, Sara Forootani, Corrado A. Visaggio. 2018. A Nlp-based Solution to Prevent from Privacy Leaks in Social Network Posts. In *ARES 2018: International Conference on Availability, Reliability and Security, August 27–30, 2018, Hamburg, Germany.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3230833.3230845

ARES 2018, August 27–30, 2018, Hamburg, Germany

ACM ISBN 978-1-4503-6448-5/18/08…$15.00
https://doi.org/10.1145/3230833.3230845

1 INTRODUCTION

Anecdotal and empirical evidence indicate that individuals do not protect adequately their privacy, especially on SN. According to the Consumer Reports' 2010 State of the Net analysis[1] more than half of SN users usually share private information. It seems that this is not necessarily due to a poor knowledge about the risks for privacy, but rather to a cognitive dissonance between knowledge and behavior: the study accomplished in [21] found that there was little correlation between participants' broader concern about privacy on Facebook and their posting behavior.

Social networking allows to gain social capital. Social capital refers to the increasing benefits deriving from the relationships among people within a specific social context or network [10]. Individuals disclose personal information in order to increase their social capital. For example, an individual that shares a medical diagnosis will be more likely to receive comfort from the community [23].

Previous research demonstrated that privacy preservation is also conditioned by (i) *optimism bias* (*i.e.,* the tendency of believing to be less exposed to risks than others) [2], and (ii) *overconfidence bias* (*i.e.,* individuals tend to consider themselves with higher skills than those they actually exhibit) [14]. People are not able to evaluate all the relevant parameters for estimating privacy risks [5]. Thus, their privacy decisions are compromised by incomplete information and bounded rationality [1]. Moreover, Jabee and Alam [13] revealed that users of SNs are not aware of how to protect properly privacy on SNs.

For all these reasons it is needed to provide users with a technology that evaluates the sensitiveness of information while the user is publishing it, and warns them about the risk determined by sharing a certain information. Current solutions that SNs adopt for protecting users' privacy are mainly based on setting customizable privacy preferences, a mechanism that turns out to be insufficient for contrasting

[1] Consumer Reports National Research Center (2010) State of the net 2010, Consumer Reports Magazine, June 2010

leakages of privacy[20], as it does not allow users to have control over data.

The solution that we propose with this paper aims at identifying when a sentence (i.e. a post to a SN) entails a risk of privacy leakage. This is an urgent issue to face, as stated by Kayes and Iamnitchi[15]: "Research could explore a comprehensive risk assessment and solutions considering SNs as potential trackers [of sensitive information]". Our solution recognizes specific patterns used in natural language for expressing specific classes of privacy leakage (*i.e.*, domains concerned by privacy, such as sexual or political orientation). A particular pattern is recognized in a post when given conditions (that are manually defined) on its syntactic structure are verified.

Machine learning has been largely investigated in literature for intercepting privacy leakage in natural language [6, 17, 22, 22]. Although machine learning showed to be effective in this context, its performances could degrade if: (i) the features are not selected properly; (ii) the training set is not adequate to the testing set; and (iii) the correct machine learning algorithm is not used.

The effectiveness of the method we propose depends only on the completeness and correctness of the formalized patterns for any domain of privacy leakage. Thus, the performance of sentences classification does not change with the features' selection nor a training set. In addition, our proposed method has a great accuracy, since it is potentially able to distinguish the different impact on privacy of sentences that exhibit very similar structures, i.e. when the intersection of lemmas used by the two sentences is wide.

2 THE METHOD

The main purpose of our work is to support users while sharing textual data in social network posts, in order to avoid unintentional revelations of sensitive information. To this aim, we propose a method based on Natural Language Parsing (NLP) which is able to automatically identify privacy leakage in social network posts. Specifically, we argue that users release sensitive information according to recurrent natural language patterns, and the identification of these patterns could be useful for automatically recognizing such privacy leaks. For instance, considering the following posts:

1. *Finally, I'm in Rome!*
2. *Unfortunately, at that time, I was in London*

We can notice that they both reveal sensitive information about the location of the post's author that malicious users could leverage for tracing her [26]. Moreover, both posts match a shared recurrent language pattern (*i.e.*, *"[someone] is in [somewhere]"*). This example demonstrates how recurrent patterns could be exploited to identify privacy leakage in natural language texts.

Table 1: Examples of Stanford Typed Dependencies representation

Post 1	Post 2
advmod('m-4, Finally-1)	advmod(was-8, Unfortunately-1)
nsubj('m-4, I-3)	det(time-5, that-4)
root(ROOT-0, 'm-4)	prep_at(was-8, time-5)
prep_in('m-4, Rome-6)	**nsubj(was-8, I-7)**
	root(ROOT-0, was-8)
	prep_in(was-8, London-10)

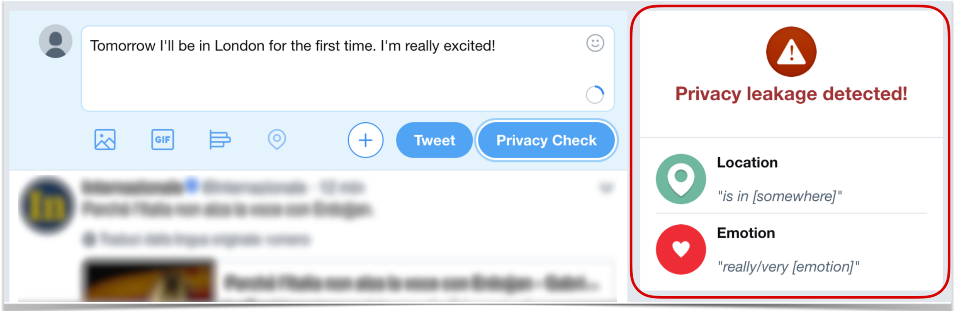
Thus, we leverage a technique, previously proposed for the recognition of useful text fragments within discussions among developers [9]. The technique relies on the identification of natural language patterns within sentences contained in the target texts, using the Stanford Typed Dependencies (SD) representation [8] of these sentences. The SD representation of a sentence models the sentence's grammatical structure as a list of triples, in which each triple describes the grammatical relation existing between two words: the *governor* and the *dependent*. In Table 1, the SD representations of posts 1 and 2 are showed.

Listing 1: Example of a NLP heuristic's definition

```
<NLP_heuristic>
  <sentence type=``declarative''/>
  <type>nsubj/prep_in</type>
  <text>[someone] is in [somewhere]</text>
  <conditions>
    <condition>nsubj.governor=``be''</condition>
    <condition>nsubj.governor=prep_in.governor</condition>
  </conditions>
  <sentence_class>LOCATION</sentence_class>
</NLP_heuristic>
```

We can notice that the posts share similar grammatical frames (*i.e.*, the verb "to be" as the principal predicate of the sentence, a nominal subject and a prepositional phrase connected with this predicate, as reported in Table 1). Thus, the presence of this particular syntactic structure in a generic post may indicate a possible disclosure of sensitive information. We can therefore formalize a *NLP heuristic* aimed at automatically detecting the presence of this natural language pattern through the identification of specific keywords in precise grammatical roles and/or specific syntactic frames [9]. More precisely, a NLP heuristic is a rule able to recognize a particular path in the Stanford Typed Dependencies tree of a generic sentence. Identified heuristics are collected in a XML file, in order to (i) make existing heuristics more flexible to further refinements, and (ii) facilitate the integration of new heuristics.

Listing 1 illustrates an example of NLP heuristic's definition. We implemented an engine able to automatically parse the XML file in which heuristics are defined. Specifically, such engine takes in input (i) a natural language text to be



Tomorrow I'll be in London for the first time. I'm really excited!

Privacy leakage detected!

Location
"is in [somewhere]"

Emotion
"really/very [emotion]"

Figure 1: An example of privacy leakage detection

scanned, and (ii) the XML file containing the NLP heuristics. The engine analyses the SD representation of sentences composing the input text and, for each NLP heuristic, it verifies the satisfaction of the defined `<conditions>`. When all the `<conditions>` of a specific heuristic are simultaneously satisfied the engine assigns the value indicated in the `<sentence_class>` tag to the sentence.

Researchers in [4] identified a list of categories of private information often appearing in tweets. We leverage this taxonomy, in order to classify the kinds of sensitive information that are disclosed in social network posts. We focus on two categories of the taxonomy: *Location* (*i.e.*, posts giving out location information) and *Emotion* (*i.e.*, posts providing highly emotional content, such as frustration, hot states, etc.). For both of these categories we identified a set of natural language patterns that social network users often use to reveal sensitive data. For each identified pattern, we defined NLP heuristics (aimed at automatically recognizing it), as the example in Listing 1.

We, finally, developed a browser extension (*i.e.*, a plug-in that extends the functionality of the web browser) for privacy preserving when users write posts on Twitter. In particular, such extension integrates the `Privacy Check` button within the Twitter home page (as showed in Figure 1). When pressed, this button allows users to check (before posting) the tweet they are actually writing against the disclosure of sensitive information belonging to the *Location* and *Emotion* categories. For example, in Figure 1 the system has been able to detect sensitive data belonging to both categories of interest (*i.e.*, *Location* and *Emotion*).

3 STUDY DESIGN

The *goal* of this study is to analyze social network posts with the *purpose* of investigating the effectiveness of the proposed method (presented in Section 2) in detecting the presence of sensitive information (belonging to the *Emotion* and *Location* categories) in such posts. Therefore, the study aims at investigating the following research questions:

Table 2: The Dataset

Category	$T_{training}$	T_{test}	Total
Location	113	169	282
Emotion	82	126	208
Other	183	183	366
Total	**378**	**478**	**856**

- **RQ1**: *To what extent is the proposed method effective in recognizing sensitive data in social network posts?*
- **RQ2**: *Is the proposed method more effective than machine learning approaches in detecting sensitive data in social network posts?*

3.1 Context

To answer our research questions we collected data from an online available dataset [7]. Such dataset, contains about 9900 Facebook status updates of 250 different users from the myPersonality sample [16].

In particular, from all the available data, we randomly sampled a set of 856 Facebook statuses. An author of this work and an external evaluator (a Master's student in Computer Science) manually labeled the sampled posts according to the categories of the taxonomy presented in [4]. More specifically, annotators were asked to separately mark all the posts in which, according to their opinion, sensitive information belonging to the *Location* or *Emotion* category were disclosed, while all other posts could be labeled as *Other*. To assure that both annotators applied the same criteria when labeling the data, the categories' definitions were discussed among them before any labeling was done. To answer our research questions, we split the data collection, in two subsets: the *training set*, $T_{training}$ and the *test set*, T_{test}. Table 2 reports, for each category, the amount of posts contained in each subset.

3.2 Research Method

To answer RQ1, we manually analyzed the posts contained in the *training set* (*i.e.*, $\mathrm{T}_{training}$) and belonging to the categories *Location* and *Emotion*). From the manual analysis of these posts, 25 and 10 recurrent natural language patterns have been identified for the categories *Location* and *Emotion*, respectively. For each identified pattern, a set of heuristics has been defined, in order to automatically recognize it, since different grammatical structures can be associated to a specific pattern.

For the *Emotion* category we leveraged all the terms contained in the emotional word wheel[2] designed by the English teacher Kaitlin Robbs, a support for finding the word that best expresses a specific emotional state. Such terms have been collected in an emotions dictionary, $\mathrm{D}_{emotions}$.

Relying on the 97 defined heuristics (*i.e.,* 72 for the category *Location* and 25 for the category *Emotion*), we used our engine (see Section 2) to automatically classify the posts contained in our *test set.* We compared the outputs against the human generated oracle and evaluated the method effectiveness with widely adopted metrics in the field of Information Retrieval[3]: Precision, Recall, F-Measure, Accuracy.

To answer RQ2, we used the posts in $\mathrm{T}_{training}$ to train a set of machine learning techniques to automatically classify the posts contained in T_{test}. In order to automatically classify the posts contained in our T_{test} we performed the following steps:

1. *Textual features*: In this step, we used all the posts contained in $\mathrm{T}_{training}$ and T_{test} as base information to build a textual corpus. We preprocessed all the posts by applying (i) tokenization, (ii) lowercase conversion, (ii) stop-words removal, and (iii) stemming [11] to reduce the number of features for the ML techniques. The output of this phase is a Term-by-Documents matrix M where each row represents a post and each column represents a term. Thus, each entry $M_{[i,j]}$ of the matrix represents the weight (or importance) of the *j-th* term contained in the *i-th* post. we weighted terms using the tf-idf (term frequency - inverse document frequency).
2. *Split training and test features*: This second step splits the matrix M in two submatrices $\mathrm{M}_{training}$ (containing the rows of M modeling the posts of $\mathrm{T}_{training}$) and M_{test} (containing the rows of M modeling the posts contained in T_{test}).
3. *Integrate Labeling Information*: In this step, a new column containing the semantic category of the post (*i.e.*, the category to which the post represented by each row was assigned) is added to $\mathrm{M}_{training}$ and M_{test}.
4. *Classification*: Data contained in $\mathrm{M}_{training}$ was used to train a set of machine learning classifiers which predicted the labels to assign to each post P_i in T_{test}, leveraging the values contained in the columns of M_{test} related to P_i . In particular, we experimented (using the Weka tool [25]) different machine learning algorithms: Logistic Regression, Simple Logistic, J48, FT, Random Forest, the standard probabilistic naive Bayes classifier, which have been proved effective in several text categorization tasks [12, 19, 24].

Table 3: The values of *True Positives (TP), False Negatives (FN), False Positives (FP), True Negatives (TN), Precision (P), Recall (R)* and *Accuracy (A)* for the categories Location and Emotion

Category	TP[%]	FN[%]	FP[%]	TN[%]	P	R	A
Location	71	29	6	94	0.863	0.710	0.858
Emotion	70	30	2	98	0.936	0.698	0.908

We, therefore, compared the Precision, Recall and F-measure obtained by each machine learning approach with the results achieved by our NLP-based method.

4 RESULTS

In this section, the results of our experiment are discussed.

4.1 RQ1 Results

Table 3 shows the values of true positives(TP), false positives(FP), true negatives(TN), false negatives(FN), precision (P), recall (R) and accuracy (A) obtained with the experimentation for the two categories.

The analysis shows that the TN rate is very high for both the categories, while the FP rate is very low. This is due to the peculiar strength of the method, which is able to identify precisely which are the expressions *not used* in the natural language when a privacy leakage (in one of the two categories) occurs. As a matter of fact, the method rejects all the expressions that are not described by the defined heuristics, and this lets discarding negatives with a high precision. On the contrary, the capability of recognizing what belongs to each category (TP and FN) is affected by the incompleteness of the heuristics' set: if an expression has been not coded in a heuristic, the system will be not able to recognize it. Such a limitation can be easily overcome, as by a manual inspection the analyst could select the false negatives and produce the corresponding heuristic.

We can observe that the performances of the method are similar for the two categories, so the category scarcely affects the results of the method. We obtained better results for the *Emotion* category than for the *Location* category, because the domain of expressions used in the former category is more restricted than the broader collection of expressions used for revealing sensitive data about *Location.* Of course, it is evident how the *recall* is affected by the incompleteness of the heuristic set, while the *precision* benefits from the

[2] https://goo.gl/B64Bex

Table 4: The comparison between machine learning algorithms and heuristics for the Location and Emotion categories

Algorithm	Location			Emotion		
	P	R	F1	P	R	F1
Logistic Regression	0.615	0.651	0.632	0.424	0.619	0.503
Simple Logistic	0.748	0.473	0.580	0.835	0.603	0.700
J48	0.750	0.462	0.571	0.866	0.563	0.683
FT	0.755	0.438	0.554	0.752	0.675	0.711
Random Forest	0.564	0.751	0.645	0.881	0.413	0.562
NNge	0.433	**0.905**	0.586	0.805	0.524	0.635
Naive Bayes	0.574	0.621	0.597	0.528	0.603	0.563
Heuristics	**0.863**	0.710	**0.779**	**0.936**	**0.698**	**0.800**

capability of the method to correctly discard expressions that do not fall in the examined category. We can conclude that the *recall* is more sensitive to the incompleteness of the heuristic set than the *precision*, but, as previously clarified, this is not a limitation of the method per se, but of the process to define the heuristic. The heuristics set can be completed with additional runs of analysis in the method's tuning.

RQ1 Summary: *The accuracy of our method in detecting sensitive data in social network posts is above the 85% for both the considered categories.*

4.2 RQ2 Results

Table 4 reports the comparison between the considered algorithms of machine learning and the heuristics for the *Location* and *Emotion* categories. The comparison shows that the value of *Precision* (P) obtained through the heuristic approach is greater than the ones obtained with the algorithms of machine learning. In addition, the classification with heuristics exhibits a value of *Recall* (R) that is similar for the two categories, but (i) for the *Emotion* category it is the greatest among all the techniques compared, while (ii) for the *Location* category the *Recall* value achieved through our method is smaller than the *Recall* values obtained by the *NNge* and *Random Forest* machine learning techniques. This difference of performances between the two categories is reflected in the *F-Measure* value. However, our method achieves the best *F-Measure* values for both the categories of interest. Finally, we can conclude that the technique of classification based on the heuristics outperforms the best machine learning algorithms; nevertheless, the performances depend on the largeness of the category's domain, i.e. how many expressions in the natural language can be associated to a privacy leakage for that domain.

RQ2 Summary: *Our method outperforms all the considered machine learning techniques, in terms of F-Measure when detecting sensitive information related to both Location and Emotion categories.*

5 RELATED WORK

The literature of privacy leakage detection in social networks counts several approaches, but only a little part of them makes use of Nlp techniques, while most existing works employ machine learning.

Approaches to detect messages carrying sensitive information have been proposed in the literature for enhancing the users' awareness by signaling privacy leaks in unstructured posts [6, 18, 22]. Differently from these approaches, that are more or less effective in recognizing messages containing private data, our method is also able to clearly identify the type of sensitive information that is being disclosed.

An approach for automatically recognizing the type of private data contained in a social network post was designed by Mao *et al.* [17], who analyzed status updates and conversation posts on Twitter related to vacation plans, tweets posted under the influence of alcohol and tweets revealing medical conditions. They extracted different features (*e.g.*, regular expressions, named entities, part-of-speech tags) from these kinds of tweets in order to automatically train a Naive Bayes machine learning algorithm, obtaining a classification accuracy of about 0.8 for all the considered classes.

Existing methods for detecting privacy leakage within text make mainly use of machine learning for the classification of privacy leakage.We propose a novel approach that identifies natural language patterns within sentences that model classes of privacy leakages; furthermore the performances observed with the experimentation are much better than those presented in literature by the existing solutions.

6 CONCLUSIONS

The spread of user-generated content web applications led users to disseminate the web with sensitive information: social networks are an evident example of such a trend. Previous study demonstrated that Social network users exhibit a behaviour that is generally poorly aware of privacy risks. In such a scenario it is urgent to develop defensive technologies that are able to assist the user in limiting the risk related to privacy leakage.

We proposed a method that intercepts the sentences delivering sensitive information, by identifying the domain of privacy that is affected by the leakage. The proposed method models the structures of the sentences with a mechanism that considers the roles of the lemmas in the sentence. We compared our approach to machine learning, which is the main antagonistic approach for intercepting privacy leakage in texts: our method outperforms seven algorithms of machine learning. Finally, our method exhibits additional advantages with respect to machine learning: it is more accurate, and is not dependent on a specific training set or the selection of particular features.

Since the only limitation of the method is the incompleteness of the heuristics' set, we aim at developing mechanisms to automatically propose new sentences' structures by measuring the similarity among sentences mistakenly classified as negatives. As a future work, we also plan to (i) identify language patterns for other categories of sensitive information, and (ii) define heuristics aimed at recognizing the identified patterns.

REFERENCES

[1] A. Acquisti and J. Grossklags. 2005. Privacy and rationality in individual decision making. *IEEE Security Privacy* 3, 1 (Jan 2005), 26–33. https://doi.org/10.1109/MSP.2005.22

[2] Young Min Baek, Eun mee Kim, and Young Bae. 2014. My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns. *Computers in Human Behavior* 31 (2014), 48 – 56. https://doi.org/10.1016/j.chb.2013.10.010

[3] Ricardo A. Baeza-Yates and Berthier A. Ribeiro-Neto. 1999. *Modern Information Retrieval.* ACM Press / Addison-Wesley. http://www.dcc.ufmg.br/irbook/

[4] Aylin Caliskan Islam, Jonathan Walsh, and Rachel Greenstadt. 2014. Privacy Detective: Detecting Private Information and Collective Privacy Behavior in a Large Social Network. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES '14).* ACM, New York, NY, USA, 35–46. https://doi.org/10.1145/2665943.2665958

[5] Colin Camerer. 1998. Bounded Rationality in Individual Decision Making. *Experimental Economics* 1, 2 (01 Sep 1998), 163–183. https://doi.org/10.1023/A:1009944326196

[6] Paolo Cappellari, Soon Ae Chun, and Mark Perelman. 2017. A Tool for Automatic Assessment and Awareness of Privacy Disclosure. In *Proceedings of the 18th Annual International Conference on Digital Government Research (dg.o '17).* ACM, New York, NY, USA, 586–587. https://doi.org/10.1145/3085228.3085259

[7] Fabio Celli, Fabio Pianesi, David Stillwell, Michal Kosinski, et al. 2013. Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition.*

[8] Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, Vol. 6. Genoa, 449–454.

[9] Andrea Di Sorbo, Sebastiano Panichella, Corrado Aaron Visaggio, Massimiliano Di Penta, Gerardo Canfora, and Harald C. Gall. 2015. Development Emails Content Analyzer: Intention Mining in Developer Discussions (T). In *30th IEEE/ACM International Conference on Automated Software Engineering, ASE 2015, Lincoln, NE, USA, November 9-13, 2015.* 12–23. https://doi.org/10.1109/ASE.2015.12

[10] Nicole B. Ellison, Jessica Vitak, Charles Steinfield, Rebecca Gray, and Cliff Lampe. 2011. *Negotiating Privacy Concerns and Social Capital Needs in a Social Media Environment.* Springer Berlin Heidelberg, Berlin, Heidelberg, 19–32. https://doi.org/10.1007/978-3-642-21521-6_3

[11] William B. Frakes and Ricardo Baeza-Yates. 1992. *Information Retrieval: Data Structures and Algorithms.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[12] E. Guzman, M. El-Haliby, and B. Bruegge. 2015. Ensemble Methods for App Review Classification: An Approach for Software Evolution (N). In *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE).* 771–776. https://doi.org/10.1109/ASE.2015.88

[13] Roshan Jabee and M Afshar Alam. 2016. Issues and Challenges of Cyber Security for Social Networking Sites (Facebook). (2016).

[14] Carlos Jensen, Colin Potts, and Christian Jensen. 2005. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies* 63, 1 (2005), 203 – 227. https://doi.org/10.1016/j.ijhcs.2005.04.019 HCI research in privacy and security.

[15] Imrul Kayes and Adriana Iamnitchi. 2017. Privacy and security in online social networks: A survey. *Online Social Networks and Media* 3 (2017), 1–21.

[16] Michal Kosinski, Sandra C Matz, Samuel D Gosling, Vesselin Popov, and David Stillwell. 2015. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist* 70, 6 (2015), 543.

[17] Huina Mao, Xin Shuai, and Apu Kapadia. 2011. Loose Tweets: An Analysis of Privacy Leaks on Twitter. In *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society (WPES '11).* ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/2046556.2046558

[18] J. Neerbeky, I. Assentz, and P. Dolog. 2017. TABOO: Detecting Unstructured Sensitive Information Using Recursive Neural Networks. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE).* 1399–1400. https://doi.org/10.1109/ICDE.2017.195

[19] Sebastiano Panichella, Andrea Di Sorbo, Emitza Guzman, Corrado A. Visaggio, Gerardo Canfora, and Harald C. Gall. 2015. How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution. In *Proceedings of the 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME) (ICSME '15).* IEEE Computer Society, Washington, DC, USA, 281–290. https://doi.org/10.1109/ICSM.2015.7332474

[20] Shailendra Rathore, Pradip Kumar Sharma, Vincenzo Loia, Young-Sik Jeong, and Jong Hyuk Park. 2017. Social network security: Issues, challenges, threats, and solutions. *Information Sciences* 421 (2017), 43–69.

[21] Bernardo Reynolds, Jayant Venkatanathan, Jorge Gonçalves, and Vassilis Kostakos. 2011. Sharing Ephemeral Information in Online Social Networks: Privacy Perceptions and Behaviours. In *Human-Computer Interaction – INTERACT 2011*, Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 204–215.

[22] A. Srivastava and G. Geethakumari. 2013. Measuring privacy leaks in Online Social Networks. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI).* 2095–2100. https://doi.org/10.1109/ICACCI.2013.6637504

[23] Frederic Stutzman, Jessica Vitak, Nicole Ellison, Rebecca Gray, and Cliff Lampe. 2012. Privacy in Interaction: Exploring Disclosure and Social Capital in Facebook. (2012). https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4666

[24] Lorenzo Villarroel, Gabriele Bavota, Barbara Russo, Rocco Oliveto, and Massimiliano Di Penta. 2016. Release Planning of Mobile Apps Based on User Reviews. In *Proceedings of the 38th International Conference on Software Engineering (ICSE '16).* ACM, New York, NY, USA, 14–24. https://doi.org/10.1145/2884781.2884818

[25] Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems).* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[26] Justin Zhan and Xing Fang. 2011. Location Privacy Protection on Social Networks. In *Proceedings of the 4th International Conference on Social Computing, Behavioral-cultural Modeling and Prediction (SBP'11).* Springer-Verlag, Berlin, Heidelberg, 78–85. http://dl.acm.org/citation.cfm?id=1964698.1964710