✚ IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

PRIVATE REVIEW ANALYSIS IN SOCIAL NETWORK SITES

Miss. Pallavi D. Dhende *, **Prof. Manisha S. Otari**
* Computer Science & Engineering, N. K. Orchid College of Engg., Maharashtra, India.

ABSTRACT

Online social networking (**OSN**) has given an easy way to communicate over the world. In online review sites, communication takes place via exchanging of text, multimedia data from one party to the other party. Usually people post their reviews on such online review sites which may include disclosing of their private data. As a result, an unauthorized person may acquire such private data about a person from their posted information/review and can make the misuse of such private data. So there is need of the system that can prevent such disclosure of private information on social network. Our proposed system prevents user from posting such kind of information that contains some private/sensitive data before they post it on social network.

KEYWORDS: Social network (OSN), review, private/sensitive data, proposed system.

INTRODUCTION

Now days, people communicate more and more using digital technology, such as instant messengers, e-mails and social networking websites. But by using different online services, for example, e-shopping or the Internet forums, the users generate a ample of data about themselves. Such information may include sensitive data about location that users visit or the location of user homes, relations such as the fact that a user has a brother, or other temporal attributes like time-of-day, a specific date, or a special occasion. Such kind of disclosure enables third parties to predict a picture of the users behavior. A social network is a set of people or other social entities such as organizations connected by a set of socially meaningful relationships.

Online Social networking sites (OSN) are a type of online communities that have grown tremendously in popularity over the past years. Online social networking has given an easy way to maintain already existing relationships and present oneself to others. However, the increasing number of actions in online services also gives a rise to privacy concerns and risks.

When personal information is accessed by malicious third parties, additional risks associated with privacy become real. The nature of the risk depends on the type and the amount of information that has been provided: the information may, in certain cases, be extensive and very intimate. These online privacy risks range from identity theft to both online and physical Stalking . Unauthorized access to private information may cause economic losses to the individual. However, the **SNS** related privacy concerns are even more significant to both ones self-image and public identity. Loss of privacy and control over personal information may cause damages that are socially irreparable: losing face among friends, revealing secret information, making social blunders, or simply giving a wrong impression. Hence, it becomes important to protect private information disclosure on online social network.

MOTIVATION AND RELATED WORK

Much of the work has been done on characterizing the nature of privacy leaks on Twitter and focused on users divulging vacation plans, tweeting under the influence of alcohol, and revealing personal medical conditions.

Most of the research in online reviews is focused on opinion mining, which is the extraction of opinion features from sets of reviews. [1] Developed an opinion mining tool that would generate a list of product attributes (quality features, etc.) and aggregate opinions about each of them (poor, mixed, good). They developed a classifier for distinguishing between positive and negative reviews. **Mining Hu and Bing Liu** also mined product features from customer reviews [3].

More recent work on opinion mining and sentiment analysis by **B. Pang and L. Lee** [5] briefly acknowledges the broader issues of privacy, manipulation, and economic impact of online reviews. A difference between the opinion mining research and this work is the focus on privacy as well as the inclusion of restaurant, hotel, and movie reviews and so on rather than just product reviews.

One other related research area worth mentioning is data loss prevention. [2] Presented an automatic text classification algorithm for classifying enterprise documents as sensitive or not sensitive, which had a false negative rate of less than

3.0% and a false discovery rate of less than 1.0%.

EXISTING WORK

There exist some systems that perform only characterization of the nature of the privacy leak in social networking websites [4]. Such system analyzes the private information which is leak unknowingly by users in social networking sites. This system considers few categories of information disclosure for example information about relation, location and other some temporal attributes. This research has made the use of review scraping, keyword matching, & name annotation concept to achieve this aim. This research has developed tool in **Python**. The tool takes local HTML file containing review text as input. One of the review sites retrieve review text from the site server after the page has finished its loading. This process is accomplished by using asynchronous JavaScript requests. There exist various techniques through which reviews can be extracted; web scraping is one of them. In existing system reviews were scraped from these HTML files using Beautiful Soup, which is a Python HTML/XML parser.

But existing system is having some noticeable limitation. It doesn't provide any mechanism for preventing users from posting their reviews which may contain some kind of private information. It does only characterization of different kinds of private information leak.

EXISTING TECHNOLOGY

There exist many techniques for information extraction from online social networks and performing keyword matching. Information can be extracted in many ways like **Web scraping** or **web harvesting** or web data extraction is one of the techniques available for data extraction. Keywords were brainstormed for each sub-category and stored as static Python lists in the privacy-check tool Source code. A listing of the keywords can be found on our website1 [6].

Keyword matching can be done using regular expressions to split each review into a list of words and non-words (i.e., white space, punctuation, and numbers) [2]. Words are then matched against keywords in the given keyword list. For each match, the word can be annotated before re-joining the list of words and non-words.

OVERVIEW OF APPROACH

Definition of OSNs

Boyd and Ellison's widely used dentition [6] captures the key elements of any OSN:

Definition1. An OSN is a web-based service that allows individuals to:

1. Construct a public or semi-public profile within the service,
2. Articulate a list of other users with whom they share a connection,
3. View and traverse their list of connections and those made by others within the service.

A list of Online Social Networks (OSNs) where users post their reviews are **Facebook, MySpace** or relative **(Geni),** but also includes connections like follower **(Twitter),** professional **(LinkedIn)** or subscriber **(YouTube).**

In recent years, **Online Social Networks (OSNs)** have become an important part of daily life for many. Users build explicit networks to represent their social relationships, either existing or new. Users also often upload and share private information related to their personal lives. The potential privacy risks of such behavior are often underestimated or ignored. Aside from creating an actual network of social links, many OSNs allow their users to upload multimedia content, communicate in various ways and share many aspects of their lives. Because of the public nature of many social networks and the Internet itself, content can easily be disclosed to a wider audience than the user intended. When internet user post such private information it is possible that an unauthorized person may access such private information and can make misuse of it.

Hence, our aim is to prevent users of social network from the inadvertent disclosing of private information and to alert them about privacy leaks on social network and giving the option to remove such information or think twice before posting the sensitive information.

PROPOSED SYSTEM

Our proposed system aims at preventing inadvertently disclosure of private information by user on social network. The system architecture of our proposed system is shown below.
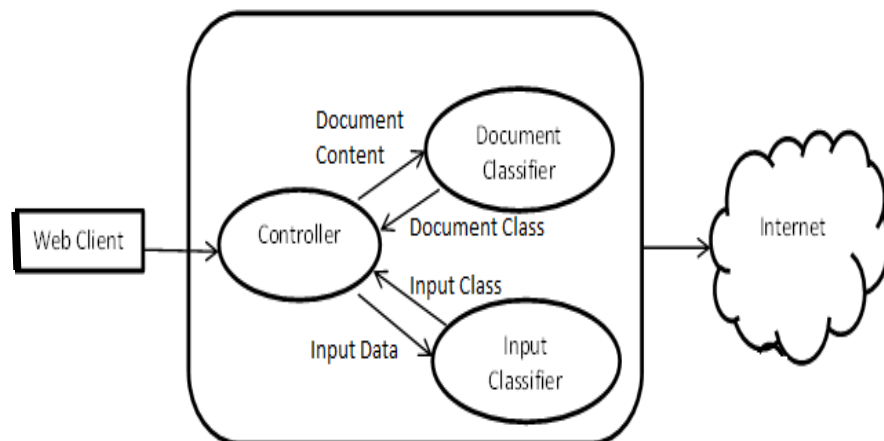
Proposed System Architeture

In our proposed system, first we are extracting data or information from websites. This extracted data and training dataset is then stored in database. Next, the database now contains extracted data and keyword list. This data is then transfer to Text classifier. Text classifier then performs keyword matching, named entity recognition and information/review annotation and after these tasks it prepares alert message for alerting user to prevent them from disclosure of private information being posted on social network. This message is transfer to controller and controller transfers this alert message on browser.

This proposed system goes through various set up phases, extracting data from various sites which contains information posted by user on social network is the first phase, second phase of running system is to take extracted data and training dataset and transfer it into database and store this datasets in database and Transfer this dataset to text classifier which performs keyword matching, named entity recognition and review annotation is the last phase.

System modules
A proposed system is having some modules they are listed below.
1. Text Classifier
2. Page classifier
3. Proxy Server
4.Classification services

Text Classifier classifies review posted by user on any Online Social review website. Output of this module helps to determine whether posted review contains any private information or not/whether it is sensitive or non-sensitive review.
Page Classifier this module focuses on context of page on which review is being posted.
Proxy Server In computer networks, a **proxy server** is a server (a computer system or an application) that acts as an intermediary for requests fromclients seeking resources from other servers. A client connects to the proxy server, requesting some service, such as a file, connection, web page, or other resource available from a different server and the proxy server evaluates the request as a way to simplify and control its complexity.

ANAYSIS OF PROPOSED SYSTEM
Existing work is based on characterizing the nature of private information leak in networking sites. Less work has been done in protecting private information discloser in social network before they are posted in social network by users.
More work has been done on opinion mining and sentiment analysis which acknowledges the broader issues of privacy, manipulation, and economic impact of online reviews. A difference between the opinion mining research and this work is the focus on privacy as well as the inclusion of websites where user can post their information and many more rather than just product reviews.
The increasing use of social network is creating new challenges for user privacy. People share their **feelings, emotions** on social network through the way of posting their information in the form of review on social network. Although reviews are public, many users inadvertently disclose private information to the world

about relationship, location, temporal, addiction, abuse, dating and so on. Hence there is possibility of an unauthorized person may acquire such private data about a person from their posted information and can make the misuse of such private data.

So there is need of the system that can prevent such disclosure of private information on social network. Instead of analyzing private information which is already posted on social network and making characterization of nature of private information leak in etworking sites, it is always better to prevent such private information disclosure before posting it on social network.

Existing system provide a way to reduce such disclosing of sensitive data but it is having some limitations like it is considering only of three categories like **relation, location or temporal data** and it fails to determine the context sensitivity of reviews means for example do not try to match the word ‒Flight‖ if the review is about a toy helicopter sold by Amazon.

Our proposed system is able to overcome such limitations. I our system keyword matching is enhanced using classifier based on context. It prevents user from posting such kind of information that contains some private/sensitive data before they post it on social network.

CONCLUSION
Our proposed system will provide best prevention against privacy leak in social network. It will provide an alert to the user from posting their information which may contain some private/sensitive data. And will give them option to remove such information containing private data or allow them to think twice before they post data on Social Networking sites(SNS's). This will give best prevention of posting of sensitive information on SNS's which none of the existing system does.

REFERENCES
1. Rosenblum, David. "What anyone can know: The privacy risks of social networking sites", IEEE Security and Privacy, Vol. 5, No.3 , pp 40-49, 2007.
2. Sheehan, K. B., and M. G. Hoy. "Flaming, Complaining, Abstaining: How Online Users Respond to Privacy Concerns", Journal of Advertising, Vol. 28, No. 3, pp 37-51, 1999.
3. Son, J.-Y., and S. S. Kim. "Interent users' information pricavyprotective response: a taxonomy and a nomological model", MIS Quarterly, Vol 32, No. 3, pp 503-529, 2008.
4. Swire, P. P. Markets, Self-Regulation, and Government Enforcement in the Protection of Personal Information In Privacy and Self- Regulation in the Information Age U.S. Department of Commerce, Washington, DC, pp. 3-20, 1997.
5. Swire, P. P., and R. E. Litan. None of your business: world data flows, electronic commerce, and the European privacy directive Brookings Institution Press, Washington, DC ,1998.
6. Wellman, B. "Computer Networks As Social Networks", Science, Vol. 293, pp. 2031-2034, 2001.
7. Yang, C. C. "Information sharing and privacy protection of terrorist or criminal social networks", IEEE International Conference on Intelligence and Security Informatics, pp. 40-45, 2008.
8. Yu, H., M. Kaminsky, P. B. Gibbons, and A. Flaxman. "SybilGuard: defending against Sybil attacks via social networks", the 2006 conference on Applications, technologies, architectures, and protocols for computer communications, pp. 267-278, 2006.
9. Zhou, B., and J. Pei. "Preserving privacy in social networks against neighborhood attacks", Proceedings of the 24th IEEE
10. Hung, S.-Y., C.-M. Chang, and T.-J. Yu. "Determinants of user acceptance of the e-Government services: The case of online tax filing and payment system",
11. C. Walters. Facebook's new terms of service: "we can do anything we want with your content. forever.", 2 2009. http://consumerist.com/2009/02/ facebooks-new-terms-of-service-we-can-do-anything-we-want-with-your-content-forever.
12. Hill hertz. Types of social networking websites. online, 2010. http: //www.hudsonhorizons.com/Custom-Website-Solutions/Social-Networking/ Types-of-Social-Networks.htm.

AUTHOR BIBLIOGRAPHY

Miss. Pallavi. D. Dhende

She received B.E degree and pursuing the M.E. degree in Computer Science and Engineering in Nagesh Karajagi Orchid College of Engg. & Technology, Solapur, India. She is doing his dissertation work under the guidance of Prof. M. S. Otari ,Assistant Professor at Nagesh Karajagi Orchid College of Engg. & Technology, Solapur, Maharashtra, India..

Prof. M. S. Otari

She is an Assistant Professor in Nagesh Karajagi Orchid College of Engineering and Technology, Solapur, Maharashtra, India. Her research area includes Machine learning, Natural Laguage Processing, Web Technology, networking.