International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 2, No 01, 2017 ISSN: 2348-6600
Reference ID: IJCS-302 PAGE NO: 2020-2026

Design and Implementation of an Online Social Network Application with Privacy Violation Detector Based On Text Mining Technique

V. Ramya M.Phil Scholar
Department of Computer Science
Tamil University, Tanjavur.

S.Baskaran
Head of the Department, Department of Computer Science
Tamil University, Tanjavur.

***Abstract*— Now a days Online Social Networks [OSN] are integral part of human society. In the present scenario of Networked Society, a person maintains an account for social networks for building interpersonal relations, business exchanges/transactions, sharing of ideas and information among like-mined peoples for various activities. Because of its enormous benefits, the use of OSN's is increasing day by day. The users of such systems are badly in need of the facilities for protection of their private/personal information and their likes and dislikes. Here, the term privacy is defined as: it is the maintenance of confidentiality. To keep secrecy of personal information of users and prevent privacy violation, various techniques are available and supported by different OSNs. However, the existing system could not fulfill privacy maintenance as the expected level of users. This paper discusses a model that allows user to take over the control of their profiles in OSN and thereby control the privacy. The key functions of the proposed model are that the message posted on user wall's will be filtered for unwanted content in any form and type and will be posted with the users consent only. The feasibility of this model is considered with the present OSN's scenarios.**

Keywords: Privacy Violation, Online Social Networks, Information Filtering, Text Mining.

1. INTRODUCTION

Privacy is the right of an individual to expose them selectively. An individual may prefer to expose certain information about them self a certain group of others, but may choose to hide another set of information. This right of user is difficult to maintain on the Web since information can propagate easily. It is even worse on OSN since different users can share content about an individual, without expecting an explicit confirmation from that individual. This results in tremendous privacy violations to take place. Consider the following examples: A user herself misconfigures the system and reveal unintended content (e.g., the user shares holiday pictures with colleagues when not intending to); or a friend of a user shares a content not knowing that the user would not want the content online (e.g., a friend shares a picture where the user is drunk). These examples show that both a user and friends can take simple actions that lead to privacy violations. More importantly, sometimes the privacy violations are more subtle. In order to discover the violation, it requires various pieces of information to be put together. For example, a user does not reveal her location but shares a picture. Any software that can process the geodes can help others to discover the user's location. Sometimes, the information needed to decipher the violations is not straightforward like this example. For instance, looking at two friend's check-ins to a remote island could signal

All Rights Reserved ©2017 International Journal of Computer Science (IJCS Journal)
Published by SK Research Group of Companies (SKRGC) - Scholarly Peer Reviewed Research Journals

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 2, No 01, 2017 ISSN: 2348-6600
Reference ID: IJCS-302 PAGE NO: 2020-2026

that they are together. Inferring this information, when neither has explicitly specified it, could easily violate their privacy. In all cases, the users seek tools that will help them to preserve their privacy and catch privacy breaches if any, so that they can take an action.

Violation of Privacy Mishandling private information, such as customer passwords or social security numbers, can compromise user privacy and is often illegal.

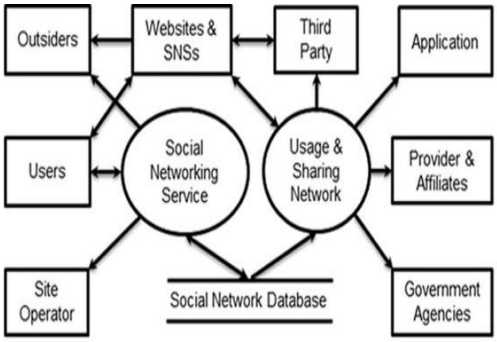

Figure 1 Social Network Service

Privacy violations occur when: [i] User personal information enters the program. [ii] Data is written to an external location, such as the console, file system, or network.

Personal data can enter a program in a variety of ways:

- Directly from the user in the form of a password or personal information
- Accessed from a database or other data store by the application
- Indirectly from a partner or other third party

A privacy violation reveals two important axes for understanding:

- The first axis is the main contributor to the situation.
- The second axis is how the information is revealed.

Most of the existing commercial systems on the Web allow a user to specify constraints on own posts and enforce them. However, this does not necessarily prevent privacy violations. That is, if a user does not want her colleagues to see her holiday pictures but her group holiday picture is shared publicly by a friend, her privacy is still violated. Various approaches to deal with privacy violations exist in the literature. Some such approaches are outlined here. One set of approaches aim to prevent privacy violations in the first place. The approaches that employ argumentation or negotiation techniques among users to reach agreements before sharing content fall into some category. Another set of approaches represents user's privacy constraints formally and try to find out if the network evolves into a state where these constraints are violated. An important work in this approach is that privacy concerns are represented as multiparty access control rules. This type of work is based on a social network model, a multiparty policy specification scheme and a mechanism to en-force policies to resolve multiparty privacy conflicts. They benefit from Answer Set Programming (ASP) to represent their proposed model. Another important work is that of Carminative that studies a semantic web based framework to manage access control in OSNs by generating semantic policies.

2. LITERATURE SURVEY

2.1. R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Preventing private information inference attacks on social networks":

This paper deals how to launch inference attacks using released social networking data to predict personal information. The Online Social Networks [OSN], such as Facebook, are being utilized by several people. OSN allows users to issue details about themselves and to connect to their friends. Some of the information revealed inside OSN is meant to be private. But, it is possible to use learning algorithms on released data to predict private information. This paper devises three possible sanitization techniques that could be used in various situations. They are discovering the effectiveness of these technique and effort to use methods of combined inference to discover sensitive attributes of the data set. It shows that can decrease the effectiveness of both Local and relational classification algorithms by using the sanitization methods described.

All Rights Reserved ©2017 International Journal of Computer Science (IJCS Journal)
Published by SK Research Group of Companies (SKRGC) - Scholarly Peer Reviewed Research Journals

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 2, No 01, 2017 ISSN: 2348-6600
Reference ID: IJCS-302 PAGE NO: 2020-2026

2.2 C. G. Akcora, B. Carminati, and E. Ferrari, "Risks of friendships on social networks":

This paper explores the risks of friends in social networks caused by their friendship patterns, by using real life social network data and starting from a previously defined risk model. Particularly, it is observed that risks of friendships can be mined by analyzing users' attitude towards friends of friends. This paper gives insights into friendship and risk dynamics on social networks.

2.3 K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks":

A large body of work has been dedicated to address corporate-scale privacy concern related to social networks. The main focus was on how to share social networks owned by organizations without revealing the identities or sensitive relationships of the users involved. Not a lot attention has been given to the privacy threat of users posed by their information-sharing activities. The approach in this paper privacy concerns arising in online social networks from the individual users' viewpoint: it propose a framework to compute a privacy score of a user, which indicates the potential privacy risk caused by his participation in the network. Our descriptions of privacy score satisfy the following instinctive properties: the more sensitive the information revealed by a user, the higher his privacy risk. Also, the more visible the disclosed information becomes in the network, the higher the privacy risk. It develops mathematical models to estimate both sensitivity and visibility of the information and also apply our methods to synthetic, real-world data and demonstrate their efficacy and practical utility.

2.4 L. Fang and K. Le Fevre, "Privacy wizards for social networking sites,"

Privacy is a huge problem in online social networking site. While sites such as Facebook allow users fine-grained control over who can see their profiles, it is difficult for average users to specify this kind of detailed policy. This paper proposes a template for the design of a social networking privacy wizard. The intuition for the design comes from the observation that real users conceive their privacy preferences (which friends should be able to see which information) based on an implicit set of rules. Thus, with a limited amount of user input, it is usually possible to build a machine learning model that concisely describes a particular user's preferences, and then use this model to configure the user's privacy settings automatically. The efficiency of the proposed technique and approach is evaluated by collecting detailed privacy Preference data from 45 real Facebook users. This paper revealed two important things. First, real users tend to conceive their privacy preferences in terms of communities, which can easily be extracted from a social network graph using existing techniques. Second, the active learning wizard, using communities as features, is able to recommend high-accuracy privacy settings using less user input than existing policy-specification tools.

3. REVIEW OF EXISTING SYSTEM

In general, the system architecture of existing system of social network comprises the following modules: user registration, user authentication, share personal information and detect privacy violation. From this design of architecture one can infer that a privacy violation on social networks looks like a violation of access control. In typical access control scenarios, there is a single authority (i.e., administrator) that can grant accesses as required by user. However, in social networks, there are multiple sources of control. Here each user can contribute to the sharing of content by putting up posts and pictures about her as well as others. Further, audience of a post can are sharing the content, making it accessible for others. These interactions lead to privacy violations, some of which are difficult to detect by users. Even though the existing systems have many advantages of their own, they have some disadvantage also. The representative list of disadvantages can be listed as: [i] Increases the danger of people declining quarry to online scams that seem genuine, resulting in data or individuality theft. [ii] Potentially marks in negative remarks from employees about the organization or potential legal consequences if employees use these sites to view objectionable. [iii] Open up the possibility for hackers to execute fraud, begin spam and virus attacks. Hence, it needs to device

All Rights Reserved ©2017 International Journal of Computer Science (IJCS Journal)
Published by SK Research Group of Companies (SKRGC) - Scholarly Peer Reviewed Research Journals

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 2, No 01, 2017 ISSN: 2348-6600
Reference ID: IJCS-302 PAGE NO: 2020-2026

new techniques and approaches which would overcome the disadvantages of the existing systems.

4. REVIEW OF PROPOSED SYSTEM

There is a very high chance of posting unwanted content on particular public/private areas, in OSN, called in general walls. This causes privacy violations. In order to maintain privacy, these types of unwanted publication/posting can be controlled and thereby prevented the unwanted messages which are written on user's wall. This can be executed by implementing Filtering Rules (FR) in our system. For this purpose, Block List (BL) will also be maintained in this system. The proposed system creates a web site on the internet. This empowers the users to automatically control the messages write on their own walls after filtering unwanted messages. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data. OSNs give support to avoid unwanted messages on user walls and identify the user behavior to avoid risk friendship.

4.1 Advantages of Proposed System:

In addition to overcoming the disadvantages of the existing systems, our proposed system will have the following advantages: [i] It is based classifying comments. [ii] Short texts are classified based STC approach and [iii] it is also blocked malicious users.

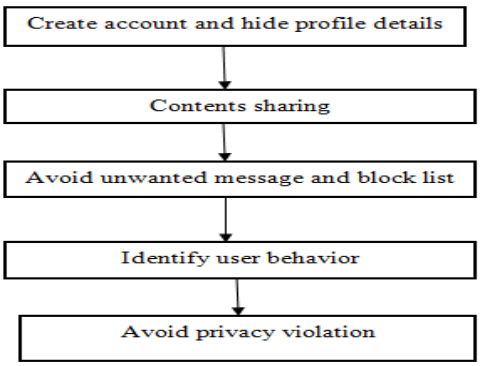

Figure 2 System Architecture of Proposed System

5. IMPLEMENTATION

To prevent privacy violation, a web based application is designed, developed and implemented. In this application, PHP is used for the front-end development and SQL server is used for the back-end development. In general, the social networks are vulnerable to privacy violation. Our technique prevents privacy violation to some extent. For this, a gateway should be developed for user authentication and data collection. The integrated Friendsmate application login procedure in to our web application is to enable user authentication. There also implemented a Friendsmate gateway to collect data from Friendsmate users.

The figure (3) shows the information flow. The tasks are represented as rectangles. A user task is depicted as a task with a figure on top while the other tasks are automated tasks. The solid arrows represent the flow between tasks. The data operations are shown as arrows. First, the user logs into the system by providing her Facebook credentials. The tool collects the user data and stores in a Structured Query Language DataBase (SQLDB).

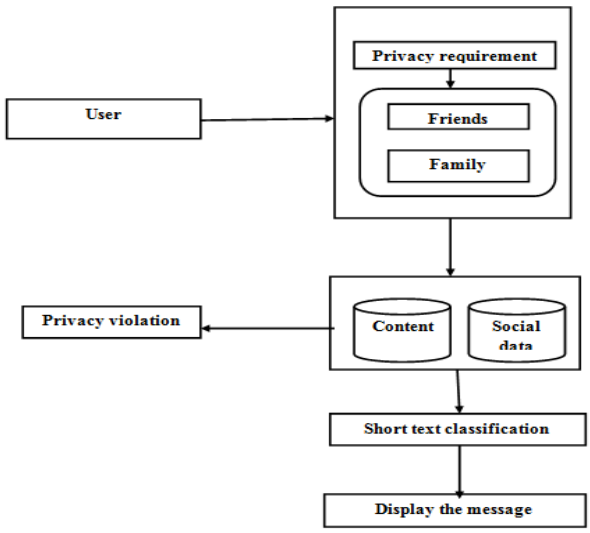

Figure 3 Implementation Steps

The user inputs her privacy concerns, which are stored as a JSON (JavaScript Object Notation) document. These privacy concerns are transformed into

All Rights Reserved ©2017 International Journal of Computer Science (IJCS Journal)
Published by SK Research Group of Companies (SKRGC) - Scholarly Peer Reviewed Research Journals

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 2, No 01, 2017 ISSN: 2348-6600
Reference ID: IJCS-302 PAGE NO: 2020-2026

commitments between the user and the social network (Facebook) operator, and the corresponding violation statements SPARQL queries (SPARQL Protocol and RDF (Resource Description Framework) Query Language) are generated as well. On the other branch, Generate Ontology's task takes care of reading user data from SQLDB, creating and storing ontology in SQLDB. Detect Privacy Violations task uses SPARQL queries and the user's ontology to monitor the social network for privacy violations. Finally, the user is shown a list of posts that violate her privacy if any. Then, the user can take an action such as modifying a post (e.g., removing a person from the audience of that post). Once the user logs out from the system, the tool removes the user data and the generated ontology's. This ensures that no information remains in the database after the detection is completed.

Data Collection: it extracts information about the user from Facebook by the use of Facebook Graph API. We request the following login permissions: email, public profile, user friends, user photos, user posts. These permissions allow us to collect information about Facebook posts together with the comments and likes of other users. We use SQLDB, which is an open-source document-oriented database, to keep the extracted information. Graph API supports the exchange of JSON documents, and it becomes reasonable to store the user data as a JSON document in SQLDB. Note that we only extract information of the user, which may be shared by the user itself or by a friend of the user (i.e., the user is tagged in a post shared by a friend).

Friendsmate enables extraction of some information of a user, such as the user's posts, the comments on the posts or the likes of the posts. However, it does not allow us to extract some important information about the users, such as the list of friends of a user. Further, it is not possible to extract any information about the posts of other users. As another limitation, one cannot extract information about user-defined lists (e.g., if the user has a family list, it is not possible to get users that belong to that list). We analyze the collected Information of the user so that we can come up with an approximate list of friends. For this, we analyze the interactions of other users with the user. For example, if a person makes a comment about a post shared by the user, then we consider this person as being a friend of the user. So, this list includes more users than the actual list of friends of the user. Consider the user actual number of friends for this user. However, by analyzing the interaction data of the user, we come up with a list of 5 users. Since the constructed list is only a partial view of the social network, our tool may not detect all of the violations. Moreover, the approximate list of friends may contain users who are not actual friends of the user (e.g., a friend of friend of the user will be included in the approximate list as a result of liking a post of the user).

6. MODULE OF PROPOSED SYSTEM

This system is developed with modules approach. This consist five modules for the following purposes Social Network Creation, Privacy Requirements, Identifier of Privacy Violation, Social Group Updating and Evaluation.

6.1MODULE DESCRIPTION

6.1.1 Social Network creation:

Social Network refers to the interface which helps user to get connection with people to whom they share and/or exchange information and ideas. A Social Network Manager of an organization is trusted with monitoring, contributing to, and filtering, measuring and otherwise guiding the social media presence of a brand, product, individual or corporation. The role is related to that of a community manager on a website forum or public relations representative. Social media managers are often found in the marketing and public relations departments of large organizations. In our application the user interface is designed with GUI concepts like in Face Book. This facilitates users to use the optional available in the package over graphical icons and visual indicators such as secondary notation, as opposed to text-based interfaces, typed command labels or text navigation.

6.1.2 Privacy Requirements

This module is responsible for testing whether privacy requirements are met by the OSN. This module needs to formally represent the expectations from the system for maintain the privacy. Since privacy requirements differ per person to person, this module is responsible for creating on-demand privacy agreements

All Rights Reserved ©2017 International Journal of Computer Science (IJCS Journal)
Published by SK Research Group of Companies (SKRGC) - Scholarly Peer Reviewed Research Journals

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 2, No 01, 2017 ISSN: 2348-6600
Reference ID: IJCS-302 PAGE NO: 2020-2026

with the system. Formalization of users' privacy requirements is important since privacy violations are due to the variance in expectation of the users' in sharing. What one person reflects a privacy violation may not essentially be a privacy violation for a second user.

6.1.3 Identifier of Privacy Violation:

Identifier of Privacy violation can be useful in two ways. First is to find out whether the current system violates privacy constraint of a user. That is, to decide if the actions of others or the user have already created a violation. Second is to find out whether taking a particular action will lead to a violation (e.g., becoming friends with a new person). That is, to decide if a future state will cause a violation. If so, the system can act to prevent the violation, for example by disallowing a certain friendship or removing some contextual information from a post. Ideally, it is best to opt for the second usage so that violations are caught before they occur.

6.1.4 Social group updating:

This module named as Filtered Wall (FW) is able to filter unwanted messages from OSN user walls. The architecture in support of OSN services is a three-tier structure. The first layer commonly aims to provide the basic OSN functionalities (i.e., profile and relationship management). In addition, some OSNs provide an additional layer allowing the support of external Social Network Applications (SNA) finally, the supported SNA may involve an additional layer for their needed graphical user interfaces (GUIs).

6.1.5 Evaluation:

This module find out whether the user is create Privacy Agreement Violator or Privacy Agreement Binder. Based on the text of communications among the users of this application, it is also found out that whether an individual / group exchanges info within the scope and permitted limits or not.

7. CONCLUSION

This paper discussed a new technique for maintaining privacy and preventing privacy violation in OSN. This innovation concept is implemented in this package with 5 modules namely Social Network creation, Privacy Requirements, Identifier of Privacy Violation, Social Group updating and evaluation. In social network creation modules, user's personal data like name, Dob, sex, place, phone number etc., are collected. In Privacy Violation Detector module, the action of the user which violates the maintenance of privacy policy is detected and violated action is denied to be executed. According with general principles of personal data maintenance, the users are permitted to edit their personal data. In updating module, based on textual communication the non acceptable words, phrases are identified and find out the violation of acceptable communication limit. This helps to grade and label the friendship as good, bad, average etc., the functioning and performance of each module is tested and found they work well. The core functions of the proposed method are that the message posted on user wall's will be filtered for unwanted content in any form and type and will be posted with the users consent only. In addition, it identifies the level of maintainability of privacy of an individual user as well as group. Since the implementation is based on Text Mining technique, it has enormous scope for further developments.

8. REFERENCES

[1] C. G. Akcora, B. Carminati, and E. Ferrari, "Risks of friendships on social networks," in IEEE International Conference on Data Mining (ICDM), 2012, pp. 810–815.

[2] K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 5, no. 1, pp. 6:1–6:30, 2010.

[3] L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in Proceedings of the 19th international conference on World Wide Web.ACM, 2010, pp. 351–360.

All Rights Reserved ©2017 International Journal of Computer Science (IJCS Journal)
Published by SK Research Group of Companies (SKRGC) - Scholarly Peer Reviewed Research Journals

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 2, No 01, 2017 ISSN: 2348-6600
Reference ID: IJCS-302 PAGE NO: 2020-2026

[4] A. C. Squicciarini, D. Lin, S. Sundareswaran, and J. Wede, "Privacy policy inference of user-uploaded images on content sharing sites," IEEE Trans. Knowl. Data Eng., vol. 27, no. 1, pp. 193–206, 2015.

All Rights Reserved ©2017 International Journal of Computer Science (IJCS Journal)
Published by SK Research Group of Companies (SKRGC) - Scholarly Peer Reviewed Research Journals