

“PrivacyDoc”: A Study on Privacy Protection Tools for Children in SNS

Ma Meng¹, Nasriah Zakaria², Salah Bindahman², Nik Mohd Asrol Alias³
and Wahidah Husain²

¹*Information School, University of Sheffield, United Kingdom*

²*School of Computer Sciences, Universiti Sains Malaysia
11800 Penang, Malaysia*

³*Faculty of Electrical Engineering, Universiti Teknologi MARA
11300 Permatang Pauh Penang, Malaysia*

*esionma@gmail.com; nasriah@cs.usm.my, bindohman2004@yahoo.com,
asrol@ppinang.uitm.edu.my; wahidah@cs.usm.my*

Abstract

With the development of social networking sites, the need to manage privacy levels has been demanded by many users and discussed by many researchers. Due to the fact that a large number of users of these sites are children, this need is becoming even more pressing. Protecting online privacy is crucial, especially for children who do not have the knowledge, ability and awareness to avoid risks caused by privacy breaches. This paper presents and evaluates a tool called PrivacyDoc that will help parents control their children's privacy level and recommend a suitable level of privacy. This tool has been implemented and its usability has been tested based on the user's point of view.

Keywords: *Social networking sites (SNS), PrivacyDoc, Case-based reasoning*

1. Introduction

Social networking sites (SNS's) have created new ways to communicate and share information via the Internet. They have become popular web based communities which millions of people have joined due to the ease of getting and sharing information through the Internet [1]. These sites have raised privacy issues among many researchers [5,6,7,8] who have discussed the risks to individuals by becoming involved in these sites without considering privacy as a priority and taking steps to make their privacy secure and non-violated. Authors in [2] concluded that privacy risks can be classified into three types: security risks, reputation and credibility risks, and profiling risks. These risks arise either from users choosing inappropriate privacy settings, or from the social networking sites themselves disclosing information to third parties for commercial purposes.

What makes the problem even worse is that the average age of SNS users is becoming lower and lower, with a large number of users between 10 and 16 years old. These users are also actively participating in social networking activities which makes the situation even more serious. Young people have a lesser ability to determine what is right and what is wrong, and they are easily misled by unhealthy information such as violence and pornography [2,4]. Children at this age are not mentally mature enough to properly control their privacy. They often post sensitive information on their profile pages which can easily cause privacy disclosure. Hence, they are in need of guidance

from their parents regarding their use of SNS's and in dealing with their privacy settings. A privacy management tool named "PrivacyDoc" was presented in [3], the goal of which is to help users evaluate and adjust their privacy settings.

In this paper, we will further discuss the PrivacyDoc tool. Its system design is presented and explained in detail and then the system is tested and evaluated based on the responses from the participants who used the system.

2. PrivacyDoc

PrivacyDoc tool is a tool that helps SNS users manage their privacy settings by recommending a certain range of settings based on information stored on the database obtained from previous users. It was designed especially for parents who have children below 16 years old. Parents enter certain data about their children into the tool and the tool returns an appropriate privacy setting for their children's SNS account in order to protect them from any possible risks that could violate their privacy. The tool was developed using case-based reasoning which has not yet been used for designing any privacy tool or system [9]. Case-based reasoning makes direct use of past experiences or cases to solve a new problem by recognizing its similarity with a specific known previous case and then applying it to find a solution for the current situation. Each child's data is compared to the dataset of cases collected before deploying the tool, and the system gives the parents a recommendation to help them identify any inappropriate privacy settings of the child's SNS account and select better settings to prevent the child from any privacy risks.

The main issue in case-based reasoning is whether we can always trust the solutions suggested by it. In [11], the authors stated that the reliability of the system depends on three factors: case-oriented factors, algorithm-oriented factors and human-oriented factors. In case-oriented factors, we follow the foundation assumption "similar problems have similar solutions".

3. System Design and Implementation

3.1. System Design

The Java programming language was used to build PrivacyDoc's Graphical User Interface (GUI) and to write the program code for the system. The C# programming language was used to construct the internal code for PrivacyDoc because a web-based application allows a large number of users to create and store their data in the tool's SQL Express database.

Case-based reasoning was chosen for the system as it has a number of advantages compared to other methods [10, 11]. Case-based reasoning is a methodology for problem solving which was first used by [12]. It relies on using old experiences to understand and solve new problems based in their similarity to the old ones. Based on the data gathered from previous users, the system will propose and give a recommendation to new users for their SNS privacy settings by comparing the current situation with the cases stored in the database.

The database of the system is considered the core and the main mechanism that the tool depends on to give any recommendation. Without a database loaded with relevant cases, the tool is useless and cannot provide accurate recommendations to the users. So the method of acquiring the first set of data is very important. We conducted a survey in China before developing the system to gather parents' opinions on privacy protection.

The most essential item in that survey was asking the parents to define desired privacy levels (high, medium or low) for certain data, such as child's full name, home address, phone number and e-mail address. After collecting and analyzing the data, it was entered into PrivacyDoc's database. These cases serve as the basis for providing a recommendation to the first person to use this tool when it comes out.

As shown in Figure 1, the recommendation is arrived at using case-based reasoning, which means the recommendation may change if the data changes as the number of users increases. This is because every time a user selects a privacy level for an information item, his or her choices are stored in the database and combined with the choices and data which were collected and stored previously.

The tool performs comparisons in the database in order to return a recommendation to the user. Take "age" as an example, and assume that 10 users believe that age is a high level privacy item, 1 user thinks it is a medium level privacy item and 2 users think that it is a low privacy level item. At this stage, the tool will consider age as a high level privacy item because far more users (10) rank it high than medium (1) or low (2). So PrivacyDoc will suggest that a user set his or her age to be viewable by friends only. However, if in future the proportion of stored opinions change — for example, 10 users think that age is a high level privacy item, 20 think it is medium level, and 100 think it is a low level — the tool will follow the majority and recommend the user give everyone permission to view their age.

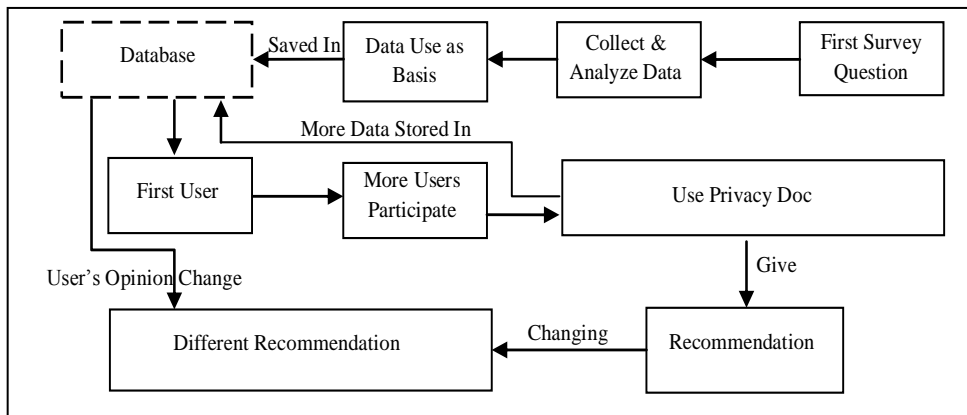


Figure 1. System Flowchart

This is how PrivacyDoc employs case-based reasoning to give appropriate recommendations to new users based on previous ones. More people using the tool will enrich the database and make the recommendations more meaningful and accurate.

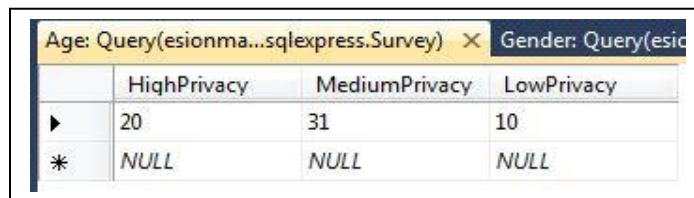
3.2. System Explanation

Based on the information drawn from many studies about privacy issues and risks related to social networking sites [5, 6, 7, 8] and the concept of a case-based reasoning system [12, 13], a prototype of the privacy management tool PrivacyDoc was developed and programmed. It was dubbed PrivacyDoc is because it will act like a real doctor in diagnosing users' privacy needs and providing a meaningful recommendations to help them improve their account's privacy level. PrivacyDoc uses case-based reasoning to make its recommendations based on data gathered from previous users. Previous users' choices are stored in the database and the system employs an algorithm to do

comparisons between old and new data in order to decide the appropriate and recommended settings for new users.

On the first screen of the system there are 11 options, captured from Facebook's privacy settings page and then modified into generic ones. The purpose of using these privacy items is because they are the most common data shared with others or stored in an SNS database. Moreover, these data are liable to result in online security risks such as identity theft, financial theft and other forms of privacy disclosure. Since PrivacyDoc at this stage is only a prototype, the collected data is only being used for testing purpose to see how well the system can calculate privacy levels using case-based reasoning. If most users can accept the concept of a privacy management tool like PrivacyDoc, then more cases will be added into the database and then the tool can be deployed more widely. The users make their selection based on their opinion of each item, evaluating whether they consider it as requiring a high, medium or low level of privacy, and submit their selections to the database. The data submitted this time are not actually performing any function. The system makes recommendations according to the previous users' choices. For each information item, the rankings for high-level privacy, medium-level privacy and low-level privacy are compared automatically within the system.

For example, as indicated in Figure 2, for "age" the number of users ranking that piece of information's privacy as High-level, Medium-level and Low-level is 20, 31 and 10 respectively. Obviously, most people considered age a medium level privacy item, so the system will recommend that a user set their age to medium privacy. But in SNS privacy settings, simply defining each item as high, medium or low privacy level is too general and hard to accept for most users. It is hard to accept because for age, users may not be able to decide the privacy's priority by just using the term low, medium and high privacy. So, it is necessary to be more specific and show the users the corresponding explanation of each privacy level. In this tool we decided to follow Facebook's guidelines, so a high level of privacy means share only with friends, medium level means share with friends of friends and low level of privacy means share with everyone.



	HighPrivacy	MediumPrivacy	LowPrivacy
▶	20	31	10
*	NULL	NULL	NULL

Figure 2. Number of Choices for Age

Through the programming code, the tool compares previous privacy level rankings and gives recommendations based on them. Images can help users better understand the appropriate choices for privacy level so, as shown in Figure 3, an image of a tick or checkmark appears in the recommendation page and suggests setting the privacy level of "age" to medium, or "friends of friends".

As the number of users increases the database will store more data. Selections for the privacy level of each information item in the database are updated after each user makes his or her choice and submits it.

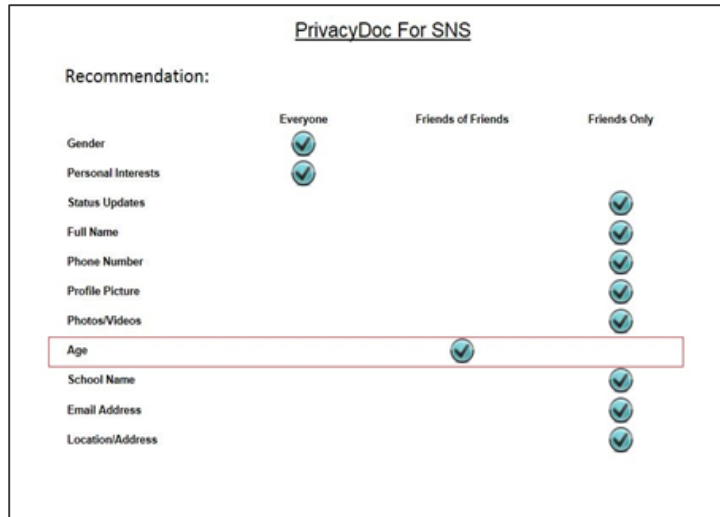


Figure 3. Recommendation Page

As mentioned above, it is possible that the recommended privacy level for one user's data item may change after a while. Referring to the example of "age" used before, more people may come to think that age should be a low privacy level information item. If the number of cases ranking it as low privacy exceeds the number of cases ranking it as high or medium privacy the recommendation will change, and the tick will move to recommend the user show their age to 'everyone'. The tool is thus a dynamic tool, in that the recommendations can change based on changes in the opinions stored in the database.

If two different privacy levels for one data item are chosen by same number of users, then the tool will consider the more private option as the optimal choice. For example, if 30 users consider one's full name as high privacy level information, but 30 other users think full name is low privacy level information, the preferred ranking will stay at the high level and the tool will automatically recommend that users should display their full name to friends only. This bias towards higher privacy will protect users' sensitive data more effectively.

4. System Evaluation

4.1. System Evaluation Design

In order to get responses from SNS members about the usability of PrivacyDoc, one more survey was done to collect feedback from people who were willing to test the tool and give their opinions about its recommendations.

The survey contained 8 questions and focused on the accessibility and usability of the tool and on users' reactions about privacy protection. The users stated their opinions about the ease or difficulty of using the tool; based on the results collected, the tool could be improved to better serve SNS users and provide more appropriate and accurate recommendations.

4.2. Data Collection and Analysis

A total of 35 participants — 12 parents and 23 students — used the tool and filled out the survey. The parents had children whose ages ranged from 11 to 21 years old, and the 23 students were between 19 and 23 years old.

The first few questions asked how easy or difficult the tool was to use, how the interface looked and how understandable the language was, as shown in Fig. 4.

All the participants had agreed that the PrivacyDoc is easy to understand and use. This is likely due to the system itself, which is relatively simple and provides only two functions (enter data and get recommendation). This tool does not include any scoring features; it proposes recommendations for SNS privacy settings directly based on previous users' data. However, 6 of the 35 people believed that the default SNS settings provide enough protection for their privacy; they think managing privacy settings is not that important as they are protected by SNS privacy rules and regulations.

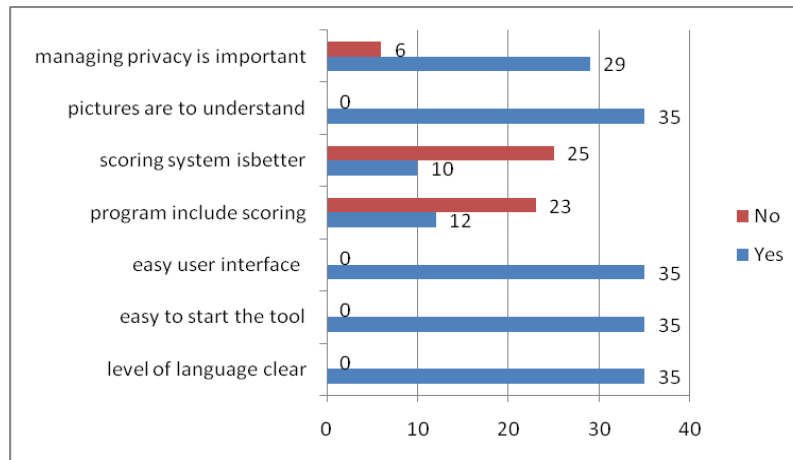


Figure 4. Participants' Responses regarding privacyDoc

The next question asked participants what method they were using currently to protect their privacy from being violated. As indicated in Fig. 5, the majority of participants said that they only added friends whom they knew personally. Eight participants know how to customize their privacy settings in order to prevent any unauthorized access to their account. Seven of them said that they are not actively participating in SNS, which means that they can only respond based on others' sharing choices, so they share only limited information with the public and don't post a lot. Finally, 2 of them stated that they do not really care about their privacy.

Some people believe that SNS is a platform where everyone should share as much information as they can, and that this is the rule for living in a virtual community. If people refuse to share their status and information about their daily life, the SNS will lose its meaning of communication and information sharing. Someone who holds this thought must be taught and educated about the threats to their privacy and the serious risks of privacy disclosure [5]. SNS users should exercise self-limiting commonsense both in reviewing what they post, and in periodically reviewing what is available online about them as suggested by [5]. Privacy breaches may not only result in harassment or embarrassment, it may even lead to mental and physical damage, a serious danger which many users do not expect.

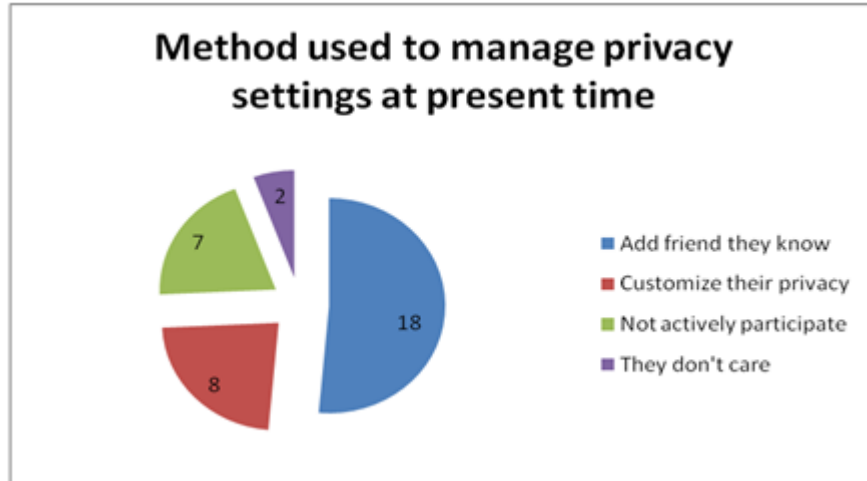


Figure 5. Privacy Management Methods

Another question asked whether or not the user was satisfied with the recommendation they received from the system. Of the participants, 28 were satisfied with the recommendation provided by PrivacyDoc. However, 7 participants said that they were not convinced by the recommendations proposed by the tool. Among the 28 people who found the recommendation acceptable, 25 of them would like to adopt this recommendation for their SNS privacy settings but 3 people would not like to do so. Those participants who were not satisfied with the recommendation and the people who refused to apply the recommendation to their privacy settings mostly think that PrivacyDoc is very simple and basic and that it does not provide enough functions that will really protect their privacy. The simple GUI also makes it hard for them to believe strongly in the recommendations that PrivacyDoc gives.

Case-based reasoning can use normal or expert users to build our database, but for PrivacyDoc we had only one type of user, which is normal users. So new users find it difficult to accept the previous users' opinions. They think that most of the users are just the same as themselves and therefore their opinions are not as useful as if the recommendation had come from expert users. Actually, one key to the success of case-based reasoning is the number of experts involved in building the case database. The experts have professional knowledge in a specific area and their recommendations are more powerful and trustworthy, which makes it easy to convince other users to believe in the system's advice and adopt it for their accounts.

An awareness of the potential risks of privacy does not mean that users always do the right things to protect their privacy settings. This has been proven by [14] who conducted a survey among SNS users. She found that although the participants selected stalking and identity theft as their primary privacy concerns, it did not stop them from posting their phone numbers, real names or profile pictures in the SNS.

PrivacyDoc currently is only a prototype but it may have potential market value if a professional team develops it further. As mentioned above, the cases stored in the database should ideally come from experts who have the knowledge and ability to properly define information items' privacy level, so the user will feel safe in relying on their opinions and will be more likely to apply them. Furthermore, a beautiful and friendly GUI design might lead to better acceptance of the tool; it currently has a poor GUI design which makes users less inclined to accept or use it. If PrivacyDoc were developed by a professional team, about 85% of participants reported they would be glad to use this tool and adopt it to help them to improve their level of privacy protection. The limited functionality also makes users think the

tool is not trustworthy, so the tool could also be expanded to include more functions instead of just storing the data and giving the recommendation.

5. Conclusion

Social networking sites play an important role in our daily life, but they also present privacy risks especially for young users who are not aware of the potential risks or mature enough to protect themselves. In this paper, the privacy management tool PrivacyDoc was presented, tested and evaluated based on real users' perspectives. PrivacyDoc received good responses from the users but has some limitations. A good user-friendly GUI design and experts willing to donate their professional knowledge to the development of a database of cases would lead more users to consider PrivacyDoc as a reliable tool, and make them willing to apply its recommendations to their own SNS privacy settings.

References

- [1] A. Acquisti and R. Gross, ACM Workshop on Privacy in the Electronic Society (WPES), (2005).
- [2] A. Ho, A. Maiga and E. Aimeru, "Privacy Protection Issues in Social Networking Sites", Proceedings of IEEE/ACS International Conference on Digital Object in Computer Systems and Applications, (2009), pp. 271 – 278.
- [3] M. Meng, N. Zakaria and S. Bindahman, Universiti Malaysia Terengganu International Annual Symposium UMTAS, (2011).
- [4] A. Felt and D. Evens, In Web 2.0 Security and Privacy W2SP: Oakland, California, (2008).
- [5] D. Rosenblums, Published by the IEEE Computer Society, (2007) May/June, pp. 40-49.
- [6] E. Craig Wills and K. Balachander, Published in WOSP '08 Proceedings of the first workshop on online social networking, (2008).
- [7] H. Richter and K. Strater, Published in SOUPS' 07 of the 3rd symposium on Usable Privacy and Security, (2007).
- [8] X. Chen and S. Shi, "A Literature Review of Privacy Research on Social Networking Sites", Proceedings of International Conference on Multimedia Information Networking and Security, vol. 1, (2009), pp. 93-97.
- [9] L. Linru, S. Jieli and Z. Rongmei, "Industrial and Information Systems", IIS' 09, (2009), pp. 178-181.
- [10] S. H. Rubin, "Learning in the Large: Case-Based Software Systems Design", In Proceedings of IEEE International Conference on Decision Aiding for Complex Systems, vol. 3, (1991) October 13-16, pp. 1833-1838.
- [11] K. Wang, J. Liu and M. Wei, "A Study on the Reliability of Case-Based Reasoning Systems", Proceedings of IEEE International Conference on Data Mining Workshops, ICDMW '08, (2008) December 15-19, pp. 60-68.
- [12] G. SongJie, Second International Symposium on Electronic Commerce and Security, ISECS '09., vol. 1, (2009) May 22-24, pp. 40-42.
- [13] H. Xue and F. Yu-Qiang, "Research on Negotiation Support System Based on Case-Based Reasoning", Proceedings of 2005 International Conference on Machine Learning and Cybernetics, vol. 1, no 5, (2005) August 18-21, pp. 2850-2854.
- [14] H. Pashley and T. Govani, <http://lorrie.cranor.org/courses/fa05/tubzhlp.pdf>, (2007).