

ブログにおける個人情報漏えいの状況調査

佐藤 和紀[†], 安井 良介[†], 針谷 友彰[†], 金井 敦[†], 廣田 啓一^{††}, 谷本 茂明^{††}

[†]法政大学大学院 工学研究科

^{††}日本電信電話株式会社 NTT情報流通プラットフォーム研究所

あらまし

近年、インターネットの普及率が飛躍的に高まると同時に、個人情報の漏えいが大きな問題となっている。それにつれて、個人情報保護法の施行にともなって、情報漏えい防止のための対策や技術の開発が行われ、個人情報の保護に対する意識が社会的にも高くなっている。その一方で、こうした対策や技術は企業における情報漏えい防止を目的としたものであり、個人が日記感覚で情報を発信するブログなどでの個人情報漏えいを対象とはしていない。我々はこうした個人の情報発信における個人情報保護の必要性を提案しているが、実際にどの程度個人情報が漏えいしているのか把握できていないのが現状である。そこで、本研究では実際にブログ記事を調査し、分析することにより、現状のブログにおける個人情報の露出状況の把握を行った。本報告では主に学生層を対象として一般層との比較を行い、学生層のブログにおける個人情報漏えいの傾向を分析する。

A Report on the State of Personal Information Disclosure for Blog

Kazuki SATO[†], Tomoaki HARIGAYA[†], Ryosuke YASUI[†], Atsushi KANAI[†],
Keiichi HIROTA^{††}, Shigeaki TANIMOTO^{††}

[†]Graduate School of Engineering, Hosei University

^{††}NTT Information Sharing Platform Laboratories, NTT Corporation

Abstract

In recent years, personal information disclosure has become a serious problem with the rapid spread of the Internet. Many companies become paying attention to protection of personal information disclosure, and have taken preventive measures against the problem. On the other hand, in the case of CGM, the possibility that the individual easily writes personal information on his/her blog or SNS become increasing. However, we don't recognize the situation and volume of personal information disclosure on blog. In this paper, we investigate actual blog articles and analyze the state of personal information disclosure on blog especially with secondary education students and general persons.

1. はじめに

近年、各家庭に複数台のパソコン、1人1台以上の携帯電話を所有することが当たり前になり、これらの機器を通じたインターネットの利用がとても身

近な存在となって、利用率が非常に高まっている。それと同時に、個人情報漏えいが大きな問題となっている。特に個人情報保護法の施行にともなって、情報漏えい防止の為の対策や技術の開発が行われ、個人情報

の保護に対しての意識が社会的にも高くなっている[1][2]。しかし、メディアなどによる情報漏えいのニュースではほとんどの場合、企業が所有するデータベース等が流出したことによる情報漏えいである。

一方、ブログやソーシャルネットワークサービス(SNS)の普及に伴い、個人が自分の情報を発信している場合もある。多くのユーザはブログを日記として利用している為、情報を発信しているという認識が薄い状態で記事を書いている。その結果、ブログの記事内容から住所や学校、生年月日などがわかり、個人情報が漏えいしていると考えられるケースがある。しかし、実際にどのような情報が漏えいしているのかといった調査は今まで行われていないため、ブログにおける個人情報漏えいの現状は把握されていない。

そこで本報告では一般層および学生層のブログを対象として、実際のブログ記事を調査、分析することにより、現状のブログにおける個人情報の露出状況を多角的に調査分析する。

2. 調査目的

本報告ではブログの利用者を一般層と学生層(中学生)で分類し、漏えい状況の比較を行う。一般層と学生層を分けた理由として、学生層は一般層に比べリテラシーが低いため、特に情報漏えいしやすい状況にあるのではないかと予想したからである。調査前に学生層の情報漏えいに関する以下のような予想を立てた。

まず、学生層の方が情報漏えいしやすいと考えた理由として、リテラシーの問題と共にブログの更新頻度が関係していると考えた。学生層は携帯電話から頻繁に更新するため、一般層に比べて更新頻度が高く、より多くの個人情報が書かれていると予想した。

次に、漏えいしている個人情報の種類について、一般層と学生層で漏えい状況に差がある情報があると考えた。リテラシーについての仮説から、たとえば本名は一般層に比べ学生層の漏えい度合いが大幅に高いと予想した。所属に関しても学生層は学校名などを書き込むため、一般層に比べ漏えい割合は高いと予想した。

一方、住所や生年月日については一般層と学生層ではあまり差は見られないと考えた。住所は個人の特定につながりやすいためどちらも漏えい度合いは低く、逆に生年月日はどちらもかなりの割合で漏えいしていると予想した。

その他の情報では、学生層と一般層という社会的立場の違いから、どちらかの層にしか存在しない情報が見られると予想した。例えば、結婚歴は学生層ではないと思われ、逆に一般層ではサークルなどの漏えいは

ないと予想できる。

また、学生層には特有の、一般層には見られない特徴や、ブログ形態があると予想した。

本報告では実際に調査を行った結果に基づいて一般層との比較を行い、これらの予想が学生層のブログにおける傾向として当てはまるのかを分析し、ブログにおける個人情報の漏えい状況を把握する。

3. 調査方法

3.1 調査・分析対象

本報告では社会的立場による比較分析を行うため、学生層(中学生)と一般層で分類して調査を行う。また、調査対象のブログとして、不特定多数による閲覧が可能なオープンブログを用いる。

3.2 ブログの調査手法

3.1で述べた対象とする層のブログ記事を収集し、実際に人が読んで個人に関する情報を抽出することで、表1で示す情報の中でどの情報が漏えいしているかを調査する。

まず対象とする層のブログ記事については、一般層はYahoo!ブログ[3]を対象にして調査を行った。学生層については、にほんブログ村[4]というポータルサイトに登録されているブログを対象に行った。このポータルサイトはYahoo!ブログやFC2ブログ[5]などを登録し、各カテゴリ別に分類してランキング形式で表示するものである。今回の調査では、学生層として中学生・高校生のカテゴリからブログを抽出した。

次に、対象のブログサイトから実際にサンプリングを行う個々のブログについては、公平性のためランダムに抽出した。抽出方法として、一般層についてはYahoo!ブログにある“ランダムブログ”という機能を使用した。学生層については、ランダム表示といった機能が無いため、乱数を発生させ、その乱数と同じランキングのブログを選択するといった方法を取った。

なお、にほんブログ村の性質上、登録をしていてもほとんど更新をしていないブログについてはランキング評価されていないものがあるため、抽出はランキング評価されているもののみとなる。さらに、男女別のランキングとなるため今回行った学生層の調査では男女それぞれの総数をほぼ同数で取得した。このため、単純な男女比較には、本データは適さない。

ランキング一般層と学生層におけるブログ選択方法のモデル図を以下に示す。



図1 一般層のブログ選択方法

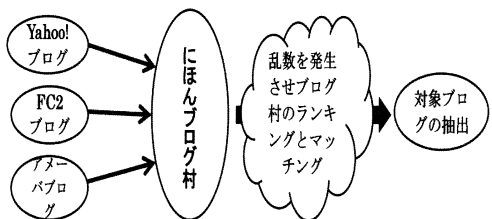


図2 学生層のブログ選択方法

3. 3 個人情報として採取した情報

個人情報は、個人を特定できる可能性がある重要な情報として、氏名、住所、電話番号、生年月日、所属などを主に採取した。表1に実際に採取した情報を示す。

また、採取した個人情報には完全な情報だけでなく、完全ではないもののある部分まで漏えいしている情報もある[6]。その場合は情報項目を階層化して採取する。図3に住所の階層化構造例を示す。

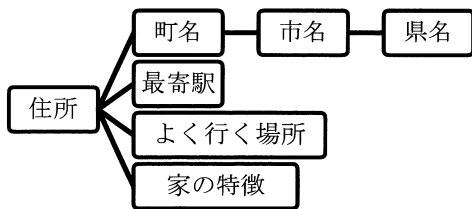


図3 住所の階層化構造例

住所の階層化構造では、町名など直接住所に結びつくものに加え、最寄駅など住所を特定するのに役に立つ情報も含むものとした。階層化構造において親情報より下の階層の情報をサブ情報項目とし、各情報におけるサブ情報項目を挙げ、それに該当する個人情報をサンプリングする。

表1 採取した個人情報

氏名	所属
苗字	学校名・勤務先

名前	学科・部署・役職
あだ名	学校所在地・勤務地
HN	顔写真
住所	その他
町名	性別
市名	趣味
県名	車
家の特徴	職業
最寄駅	アルバイト
よく行く場所	サークル
実家住所	家族構成
実家町名	結婚歴
実家市名	病歴
実家県名	好きなもの
出身地	血液型
電話番号	体格
メールアドレス	身長
生年月日	体重
年齢	資格
誕生日	

4. 調査・分析結果

4. 1 抽出ブログ件数

実際にサンプリングを行ったブログの件数を表2に示す。

表2 ブログの調査件数

一般層	105			
学生層	高校生	86	男子高校生	43
			女子高校生	43
	中学生	98	男子中学生	46
			女子中学生	52

4. 2 一般層・学生層の比較

一般層と学生層において、どのような情報が漏えいしやすい、しにくいかを調査・比較する。また、2章で述べた予想と実際の状況を検証していく。

4. 2. 1 更新頻度の比較

まず、図4にブログの更新頻度を比較したグラフを示す。

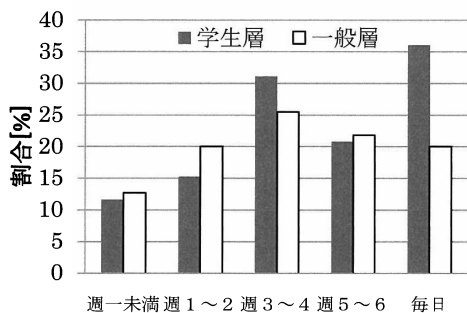


図4 ブログの更新頻度の比較

当初、学生層の方が一般層に比べ更新頻度が高いと予想したが、実際に比較してみると学生層と一般層であまり差は見られなかった。目立った点としては、学生層の方が毎日更新する割合が高いことである。この結果から、学生層の方が一般層に比べ情報漏えいしやすい要因の1つとして予想した更新頻度が高いという状況は当てはまらなかった。

4. 2. 2 氏名に関する漏えい状況

学生層の方が一般層に比べ漏えい割合が大幅に高いと予想した氏名に関する漏えい状況を図5に示す。

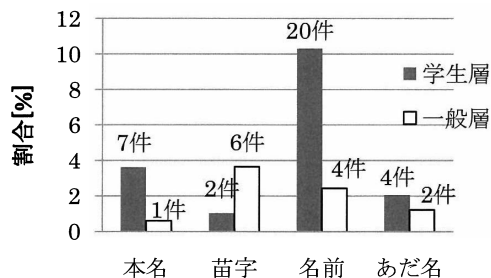


図5 氏名に関する漏えい状況

フルネームでの漏えいはやはり学生層の方が多いが、予想していたよりも学生層のフルネームでの漏えいは少なかった。また自分の名前に関して一般層は苗字を、学生層は名前を公開している割合が高かった。これは普段の生活において名前で呼んでいるか、苗字で呼んでいるかの違いによるものだと考えられる。

4. 2. 3 住所に関する漏えい状況

一般層、学生層ともに漏えいは少ないと予想した住所に関する漏えい状況を図6に示す。

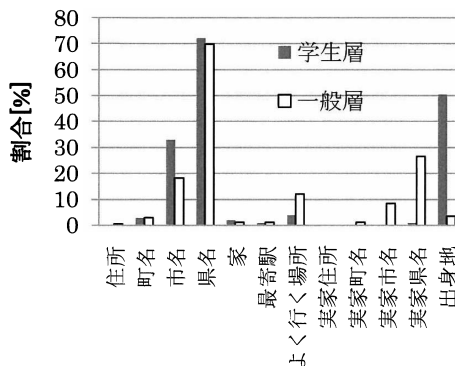


図6 住所に関する漏えい状況

住所に関しては市名の項目が、学生層が一般層に比べ割合が若干高いが、傾向としては大きな差はなかった。また、一般層は実家の市名や県名の記述が多いのに比べて、学生層は実家に関する情報の漏えいがほぼない。その一方で、出身地の項目が一般層に比べ漏えいしている割合が非常に高かった。

4. 2. 4 生年月日に関する漏えい状況

一般層、学生層ともかなりの割合で漏えいしていると予想した生年月日に関する漏えい状況を図7に示す。

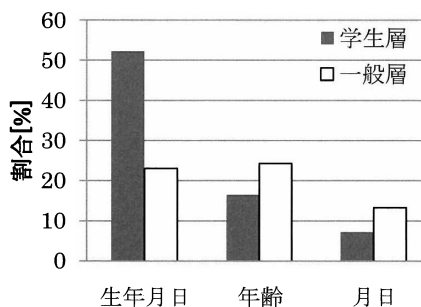


図7 生年月日に関する漏えい状況

生層の半数以上が完全な生年月日の書き込みをしている。しかし、一般層では2割程度と予想していたよりも少なく、逆に年齢やといった粒度を落とした情報の公開割合がやや高かった。

4. 2. 5 所属に関する漏えい状況

学生層が学校名などを書き込むため、一般層に比べ漏えい割合は高いと予想した所属に関する漏えい状況を図8に示す。

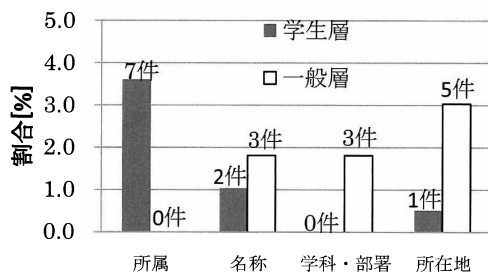


図8 所属に関する漏えい状況

予想とは異なり学生層・一般層共に、所属に関する漏えい割合は低かった。しかし、一般層が所属情報の漏えいが0件なのに対して、学生層は割合としては低いものの漏えいはしていた。

4. 2. 6 その他の情報の漏えい状況

氏名・住所・生年月日・所属以外の情報における漏えい状況を比較したものを図9に示す。

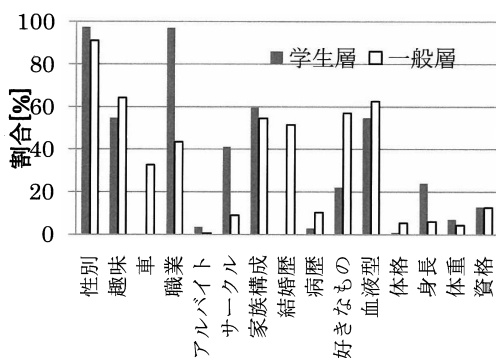


図9 その他の情報の漏えい状況

その情報では、学生層と一般層で社会的立場が関係してくる項目について差が見られた。たとえば学生では所有することができない車の項目ではやはり学生は0%である。結婚歴も同様のことが言える。学生層で職業の項目の漏えいが高いのは中学生、高校生と記載しているためであり、一般層の職業区分が多様でプライバシー性の高い場合があるのに比べて、こうした一般的な項目は割合が非常に高いと考えられる。

5. 学生層における状況と固有の特徴

4. 2では学生層と一般層での個人情報の漏えい状況を比較することで、学生層における傾向を主に調査

分析した。本項では分析対象を学生層のみに絞り込み、さらに男女別や中高別などに細分化し漏えい状況を分析する。これにより、一般層には見られないが学生層に固有に見られる特徴があるのではないかと予想について分析する。

5. 1 バトンにおける漏えい状況

学生層におけるブログ記事として“バトン”という特殊なジャンルの記事がある。バトンとは、特定のテーマにそって質問を列挙したリストの俗称で、知り合いなどからチェーン的に回って来たバトンに対し、受け取った側がバトン中の質問に対する回答をブログの記事とするものである。簡単に言うと、一昔前に流行ったチェーンメールのブログ版といえる。そして、回答した側が次は渡す側となって、知り合いなどに広めていくというものである。

学生層は特にこのバトンの記事としてブログに書き込んでいく割合が非常に高い。そこで、バトンの内容のみでどの程度情報が漏えいしているかを調査分析した。以下に、実際にブログ上で回っていたバトンの一例を示す。

私(僕)の名前は【 】です。
 性格は自分では【 】って思ってます。
 友達には【 】と言われるます。
 住んでるところは【 】で引越経験は【 】です。
 年齢は現在【 】で肌年齢は【 】です。
 誕生日は【 】で星座は【 】です。
 血液型は【 】です。
 そんな私(僕)の趣味は【 】です。
 特技は【 】です。 . . . etc

まず、図10にバトンを行っている中高生の割合を示す。

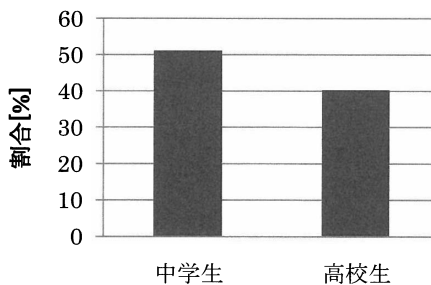


図10 バトンを行っている割合

図10からわかるように中高生ともに半数近くの学生がバトンを行っている。

次に、バトンの内容から氏名が漏えいしていた件数を図11に示す。

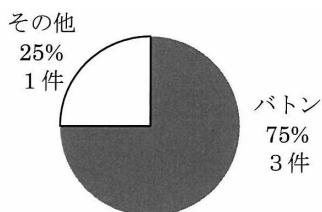


図11 中学生のバトンでの氏名漏えいの割合

元々の氏名の漏えいの件数自体が少ないが、中学生で氏名が漏えいしていたのは4人で、その内3人がバトン経由での漏えいであることがわかった。逆に高校生でバトン経由での氏名漏えいは0人だった。

次に、バトンの内容から生年月日が漏えいしていた件数を図12に示す。

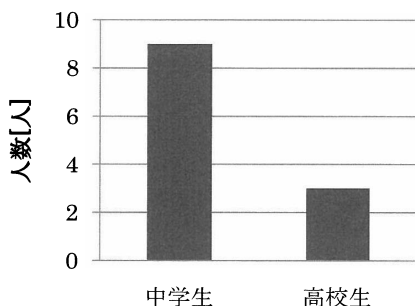


図12 バトンでの生年月日漏えい

生年月日が漏えいしている中学生の内、バトン経由での漏えいは9人で15.7%、高校生は3人で6.9%であった。

バトンに関しては、氏名や生年月日の漏えい状況から中学生は質問に対し回答をそのまま書いてしまい、高校生になると書き込む内容に対して慎重になる傾向が見られる。質問の内容にもよるが、それでもバトンでは通常の記事に比べ個人情報を公開しやすい傾向があると言える。また、生年月日については、半数以上漏えいしている学生層においてバトンでの漏えいがあり多くなかったことから、ほとんどが通常の記事やプロフィールから漏れていることになる。つまり学生層は生年月日を重要な個人情報だという認識があまりなく、かなりの割合で書き込んでいることがわかった。

5.2 男女中高生別による漏えい状況

次に、男女中高生別の漏えい状況について比較する。学生層を男子中学生、女子中学生、男子高校生、女子高校生に分類し漏えい状況を比較する。学生層をこの4つに分類し比較することで、男女や中高生での漏えい状況の違いを分析する。表3に氏名の漏えい状況、表4に生年月日の漏えい状況を示す。

表3 男女中高生別における氏名漏えい

	全 体 数	本名		苗字		名前	
		件 数	割 合	件 数	割 合	件 数	割 合
男子中学生	46	1	2.17	0	0.00	2	4.35
女子中学生	52	4	7.69	0	0.00	4	7.69
男子高校生	43	1	2.33	1	2.33	4	9.30
女子高校生	43	1	2.33	1	2.33	8	18.60

表4 男女中高生別における生年月日漏えい

	全 体 数	生年月日		年齢		月日	
		件 数	割 合	件 数	割 合	件 数	割 合
男子中学生	46	24	52.17	9	19.57	3	6.52
女子中学生	52	34	65.38	6	11.54	3	5.77
男子高校生	43	18	41.86	9	20.93	2	4.65
女子高校生	43	25	58.14	7	16.28	5	11.63

表3より、最も多くフルネームで書きこんでいた割合が高いのは女子中学生で、名前のみを書き込んでいたのは女子高校生であった。一方、男子は女子に比べ個人情報の書き込みが少ないが、中学生よりも高校生の方が情報を書き込みやすい傾向がある。女子高校生での下の名前だけの公開が多いのは、友達同士で下の名前と呼び合うことが日常的になり、それがブログの内容でも表れたためだと考えられる。

また、表4より生年月日についても女子中学生が最も書き込んでいる割合が高かった。生年月日をフルで書き込んでいる割合を見ると、氏名と同様に男子よりも女子が書き込みやすいという傾向がここでも見られた。また高校生よりも中学生の方が書き込みやすいという傾向も見られた。

5.3 顔写真漏えい状況

最近のブログでは、記事と一緒に写真を載せているものも少なくない。写真も個人情報の1つであり、載せられた写真から様々なことがわかる場合もある。本

項では、写真の中でも最も本人がわかる顔写真についての漏えい状況を男女中学生別に比較し分析する。表5に顔写真の漏えい状況を示す。

表5 男女中学生別における顔写真公開

	全体数	顔写真公開	
		件数	割合
女子中学生	46	5	10.87
男子中学生	52	2	3.85
女子高校生	43	9	20.93
男子高校生	43	1	2.33

表5より、男子よりも女子の方が顔写真を公開している割合が圧倒的に高い。さらに、女子の中でも中学生より高校生の方が公開している割合が高い。やはりここでも5.2で見られた男子よりも女子の方が情報を公開しているという傾向が見られた。また中学生よりも高校生が公開しているという傾向が見られた。これは携帯電話の所有率が高校生の方が高いためだと考えられる。

さらに女子学生の顔写真の公開状況を図13に示す。

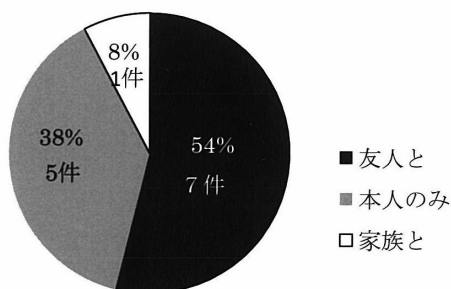


図13 女子の顔写真公開状況

顔写真を公開している女子学生の中でも、友人との写真を掲載している割合が最も高かった。また、この友人と写っている顔写真はその全てがプリクラで撮ったものをブログに使用しているものであった。

6. 考察

学生層と一般層での情報漏えい状況の比較から、最初の予想通り、一般層に比べ学生層の方が情報漏えいしていることがわかった。一般層が情報を全てではなくサブ情報項目の公開までに留めているのに対し、学生層は最上位の情報であっても公開してしまっている割合が高かった。また、学生層にはバトンといった固有の漏えい形態があることもわかった。さらに男子と女子での比較では男子より女子の方が情報漏えいして

いる傾向にあった。これは、女子がブログを全世界に公開しているページであるという意識よりも、友人同士でのコミュニケーションツールとして強く意識して使っているためではないかと思われる。このことは、女子学生の顔写真公開が友人とのプリクラを利用している割合が高い事からも考えられる。さらに、情報漏えいしている割合は、一般層よりも学生層、情報によっては学生層の中でも高校生より中学生が高いことから、年齢が低い方が情報漏えいしていることがわかった。これは年齢が低い方ほどインターネットリテラシーが不足しており、それが影響していると考えられる。

今回の調査結果全体を通して、危険な情報を完全な状態で漏えいしているブログの件数は、実際に言われているよりも少なかった。とは言え、個人情報漏えいが騒がれている今日においても、未だに個人情報を漏えいしているブログがあることについては楽観視できない。また、完全な状態でなくても個人情報が漏えいしているブログが多数あった事実も忘れてはならない。いくつかの情報や各情報が組み合わさった時、データベースとの照合が行われた場合などに、いつ誰が個人を特定され被害を受けるかわからない。これほどまでに多数の情報が漏えいしている様になった原因として、今と一昔前でのインターネット・ホームページの認識の仕方が変化したからだと思われる。一昔前はインターネットリテラシーを持った専門的な人たちしか使わなかったため安全だったインターネットが、環境が著しく進歩したためにリテラシーの有無に関係なくに誰もが使えるようになった。そしてブログなどに代表されるような誰でも簡単に自分のウェブページを持つことができるサービスの普及により、元々情報を発信する事が目的だったホームページにおいてより個人的な日記等を書く割合が増えたため、現在のような個人情報の漏えい状況になったと考えられる。今後、インターネットの普及率が低下することはないと考えられるため、個人情報保護の観点からインターネットを使用するにあたって最低限のリテラシーを身に付ける教育を徹底する必要があると言える。

7. おわりに

今回の調査では一般層と学生層の比較による分析を行った。その結果、一般層に比べ学生層の方が情報漏えいしやすいということがわかった。また、バトンや顔写真などの学生特有のブログの書き込み傾向があることもわかった。

今後の課題として、携帯電話から更新しているブログとPCから更新しているブログの比較や、オープンブログとクローズブログとの比較など今回とは違った対象での調査を行うことで、より様々な観点からブログにおける情報漏えいの現状の分析を行っていききたい。また、今回行った学生層の調査ではブログの抽出をほぼランダムという形式で行ったが、今後は完全にランダムで抽出できる方法も検討していききたい。

7. 参考文献

- [1] NRI セキュアテクノロジーズ株式会社, “情報セキュリティに関するインターネット利用者意識2006”
<http://www.nri-secure.co.jp/news/2007/pdf/vol3-1.pdf>
- [2] 谷本茂明, 廣田啓一, 山本太郎, 千田浩司, 畑島隆, 高橋克巳, 金井敦, “次世代プライバシー保護サービスのコンセプト提案” 情報処理学会論文誌, Vol.49, No.7, pp.2440-2455 (July 2008)
- [3] “Yahoo!ブログ”,
<http://blogs.yahoo.co.jp/>
- [4] “にほんブログ村”,
<http://www.blogmura.com/>
- [5] “FC2ブログ”,
<http://blog.fc2.com/>
- [6] 針谷友彰, 佐藤和紀, 安井良介, 金井敦, “ブログにおける個人情報漏えいモデル”, 情報処理学会研究報告, 2008-EIP-041, Vol.2008, No.91, pp.65-70 (2008)