社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
LOIS2011-12(2011-07)

Web 閲覧履歴情報に着目した ライフログによる本人認証に関する一考察

田村　健範† 鶴丸 和宏‡ 市野　将嗣* 小松　尚久†

† 早稲田大学理工学術院基幹理工学研究科
‡ 早稲田大学理工学研究所

〒169-8555 東京都新宿区大久保 3-4-1
*電気通信大学　大学院情報理工学研究科

〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: † ‡ {tamura, tsuru, nkomatsu}@kom.comm.waseda.jp

＊ ichino@inf.uec.ac.jp

あらまし　本稿では，ライフログの中でも Web 閲覧履歴情報に着目し，本人認証へ適用することの可能性を検討した結果を報告する．特に認証パラメータとしてアクセスドメイン，アクセス時刻，Web 接続時間などを取得することにより，ユーザの特徴を抽出し，それらパラメータの有効性を評価した．また，評価実験結果を通して，本人認証へログを適用する際に考慮すべき課題について述べる．

キーワードライフログ，Web 閲覧履歴，本人認証

A Study on a Person Authentication Method using User's Internet Access logs

Takenori TAMURA†, Kazuhiro Tsurumaru‡, Masatsugu Ichino*, and Naohisa KOMATSU‡

† Faculty of Science and Engineering, Waseda University
‡ Advanced Research Institute For Science And Engineering, Waseda University
3-4-1Ohkubo, Shinjuku-ku, Tokyo, 169-8555 Japan
* Faculty of Science and Engineering, University of Electro-Communications

1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 Japan

E-mail: † ‡ {tamura, tsuru, nkomatsu}@kom.comm.waseda.jp

＊ichino@inf.uec.ac.jp

Abstract　In this paper, we consider the effectiveness of person Authentication using life log such as web access logs. In our approach, we focus on Access Domain, Access time and connect time of the Internet from User's Internet Access logs to evaluate these parameter of life logs. Also, from experimental results, we consider problems of Person Authentication Method using User's Internet Access logs.

Keyword　lifelog, Internet log, authentication

1. 前書き

近年スマートフォンに代表されるネットワーク接続機器の高機能化に伴い，様々な個人の活動記録がデータとして蓄積させることが可能となっている．また，将来我々の生活には様々なセンサを搭載した機器が身の回りにあふれ，日常生活のあらゆる情報を記録，保

存するユビキタスネットワーク社会の実現が近づいている．これらの機器によって記録，蓄積されていく個人のディジタル化された活動記録はライフログと呼ばれ，収集されたライフログを有効活用すべく様々な検討がなされている[1][2][3]．その代表的な手段として，ライフログを用いたマーケティングがある．現在では収集されたライフログ，その中でも特にWebの閲覧履歴やインターネットショッピングでの購入履歴などを用いて個人の趣味を分析し，マーケティングやオススメ商品の広告などが盛んに行われている．このようにライフログは新たな活用方法が日々提案されている[4]．

こうした背景のもと本検討では，ライフログを本人認証に応用することを考える．将来高齢化社会が進むにつれ，現在使われているパスワードのような記憶に依存した本人認証方式では認証が困難になる人が増えてくる可能性がある．また端末の高機能化に伴い複雑な操作が要求されることも考えられ，将来的には記憶に依存せず，誰でも簡単に本人認証が行えるような手法の導入についても積極的に検討する必要があると考えられる．例えば，自動でデータベースにデータが渡され，それを自動で取得，認証を行う手段が利用できれば，以上の課題を解決することができる．そこで，自動でインターネットのアクセスログを収集，解析ができる機能の実現を念頭に置き，本人認証への適用にあたっての課題についての考察をした．

2. ライフログを用いた本人認証

2.1. ライフログの定義

現在，ライフログという言葉が示す意味について，明確な定義が存在しない．そのため，はじめに本検討で扱うライフログとはなにかを定義する必要がある．本検討におけるライフログとは「人の日々の行動(life)を自動で記録,蓄積(log)したディジタルデータ」とする[5]．

2.2. Web閲覧履歴の本人認証への適用

ライフログを本人認証に用いることを念頭に置き，本検討における認証システムの利用シーンとして，PCや携帯電話といった，データベースにアクセス可能な端末の利用中における本人認証を想定する．このような利用シーンにおいて，本稿では本人認証へ用いるライフログとしてWeb閲覧履歴を取り扱う．Web閲覧履歴を選択した理由として，今日ではPCのみならず，モバイル機器でも自宅，外出先など様々な利用環境においてWebにアクセスすることが一般化していることが挙げられる．このため，頻繁に利用するWebサイトから趣味や趣向といったユーザの特徴を抽出できるのではないかと予想した．

3. 実験

ユーザのWeb閲覧履歴情報を記録したデータより，ユーザごとのWebアクセス時刻，Webアクセスドメイン等を比較することでユーザの特徴を抽出する．以下の表1に実験緒元を示し，本検討で行った実験の手順について述べる．

1. ランダムに20人のユーザを選出する．
2. 選出されたユーザからそれぞれ，アクセスドメイン，アクセス時刻，サイト参照時間などを取得する．
3. 取得したパラメータを1週間分，1か月分と参照し，本人間での類似度から，本人の特徴を抽出できるパラメータであるか検討する．
4. 他者間での類似度から，その特徴がユーザごとに差別化することができ，本人認証に利用できるパラメータであるか検討する．

表1 実験緒元

データ購入元	Video Research Interactive Inc.
データ期間	2010/5/1～2010/6/30
データ数	20人
パラメータ	ユーザID，Webサイトアクセス日時 サイト参照時間，参照サーバURL 参照URL，参照ドメイン 参照元サーバURL，参照元URL 参照元ドメイン

3.1. 本人間におけるWeb閲覧履歴類似度

3.1.1. 平均Web接続時間における類似度

以下の図1にあるユーザの休日における平均Web接続時間を，図2に平日の平均Web接続時間を示す．本稿における休日とは土日および祝日とし，平日は休日以外の日とする．

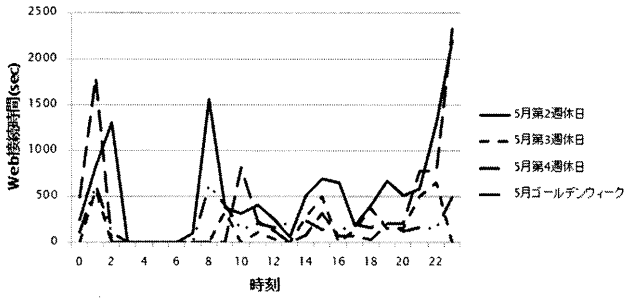

図1 休日の平均Web接続時間

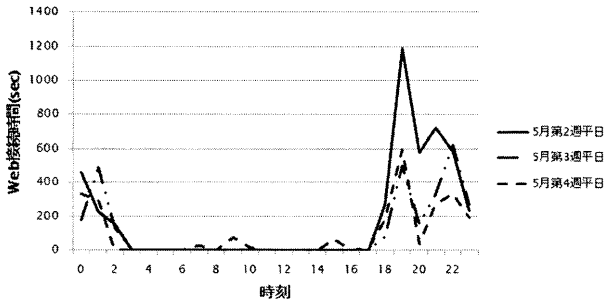

図 2 平日の平均 Web 接続時間

図1より休日のWeb 平均滞在時間は特徴的な傾向が見受けられず，各週で大きく異なったWeb アクセスとなっている．休日は基本的に時間が拘束されていないと考えられ，決められた生活規則は必ずしも見受けられない．このためWeb にアクセスする時間帯も分散し，ユーザの特徴が現れないことから，認証におけるパラメータとして利用するのは適当でないと考えられる．一方,図2より平日ではある特定の時間からWeb へのアクセスが集中する特徴が見受けられる．これは勤務先や通学先などから帰宅するなど日常生活における時間の規則性のためだと考えられる．この結果から，平均Web接続時間の場合，平均値を用いる場合は，自由な時間をとることができる休日に比べ，生活リズムが適度に定められた平日のほうが本人の特徴が見受けられるため，認証のパラメータとして採用できる可能性が高いと言える．

3.1.2. アクセスドメインにおける類似度

次に,ランダムに選択した1ユーザのログの取得期間を1週間としたときの休日のアクセスドメインの割合を図3に,平日のアクセスドメインの割合を図4に示す．また,同様に取得期間を1カ月としたときの休日のアクセスドメインの割合を図5に,平日のアクセスドメインの割合を図6に示す．

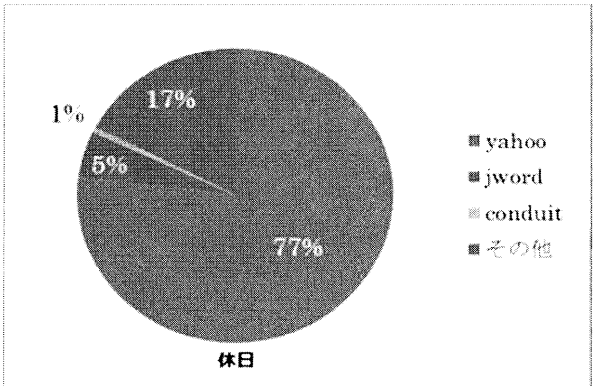

図3 1週間休日のドメインアクセス回数割合

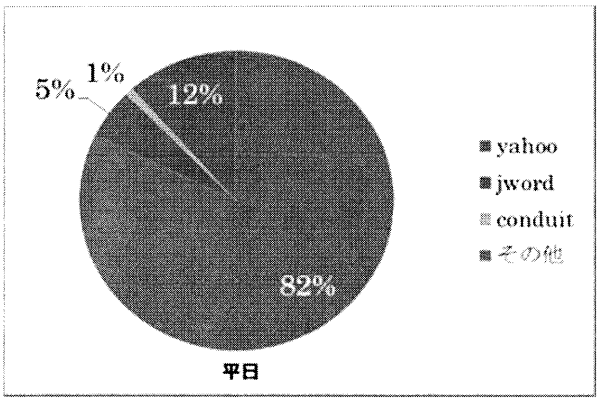

図4 1週間平日のドメインアクセス回数割合

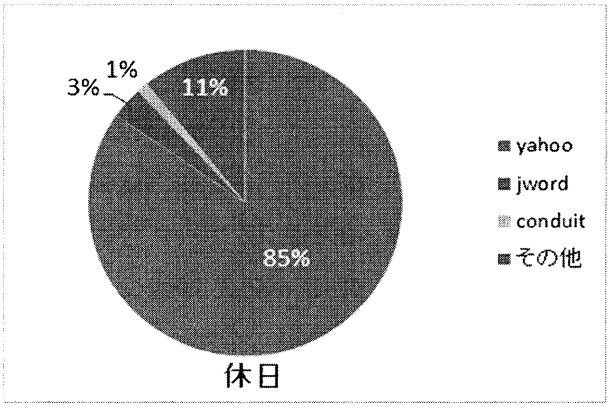

図5 1か月休日のドメインアクセス回数割合

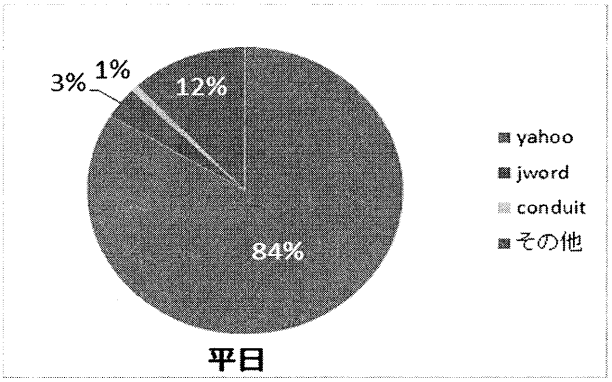

図6 1か月平日のドメインアクセス回数割合

図3，4，5，6から，アクセスドメインの場合，平均Web接続時間とは異なり休日，平日ともにアクセスドメインの傾向は同様であった.例えば今回のユーザの場合，参照期間，休日，平日問わずYahoo.co.jp へのアクセス割合が約80％となっている．これは休日や平日に左右されない，本人の特徴を示すパラメータになる可能性がある．すなわち，ユーザが頻繁にアクセスするサイトは，そのユーザの趣味や趣向が現れることが予想でき，本人認証において有効なパラメータとなる可能性がある．しかし本実験で対象としたユーザは，ポータルサイトへのアクセス割合が大部分を占めており，得られた結果がそのままユーザの趣味や趣向を反映しているとは言い難い．Web アクセスの多くがポータル

サイトである場合，そのサイトが提供するメールや検索，ショッピングといったサービスを細かく分析することでユーザの特徴を解析できることが考えられる．また図5，6から，参照期間を長期化してもアクセスドメインの傾向は変わらないことから，短い期間で本人の特徴を抽出できるパラメータとして扱うことができる可能性がある．

3.2. 他者間におけるWeb閲覧履歴類似度

3.2.1. 2者間における平均Web接続時間の類似度

3.1節で評価対象としたユーザとは異なるユーザの平日の平均Web接続時間を図7に示す．

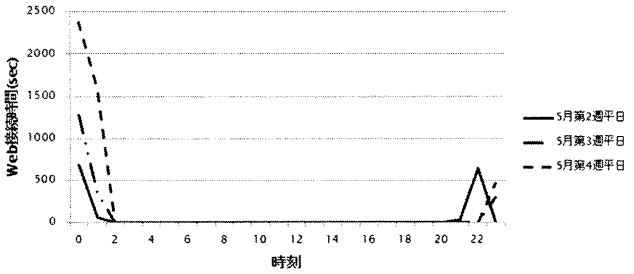

図7 ユーザBの平日の平均Web接続時間

3.1節で対象としたユーザをユーザA，本節で対象とするユーザをユーザBとすると，図2より，ユーザAでは19時付近に一度アクセスのピークが見られるのに対し，図5よりユーザBでは0時付近にアクセスのピークを確認する事ができる．平日ではユーザの生活規則が現れると考えられるため，両者の生活規則は異なることが予想できる．もちろんWeb 閲覧履歴のみを用いてユーザの生活規則を特定することはできないが，ユーザの生活規則の特徴を示すパラメータの一つとして有用なパラメータになり得ると言える．

3.2.2. 2者間におけるアクセスドメイン割合の類似度

次にユーザBのログの取得期間を1週間としたときの休日のアクセスドメインの割合を図8に，平日のアクセスドメインの割合を図9に示す．

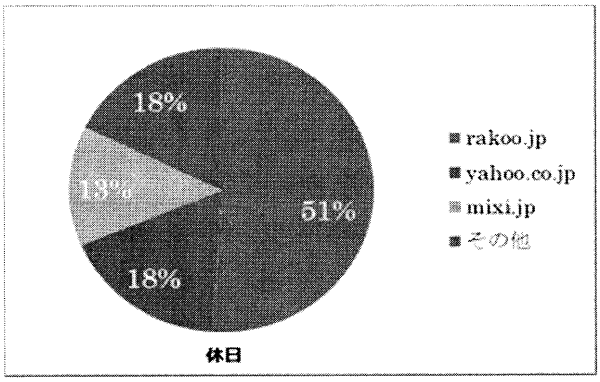

図8 ユーザBの休日のアクセスドメイン割合

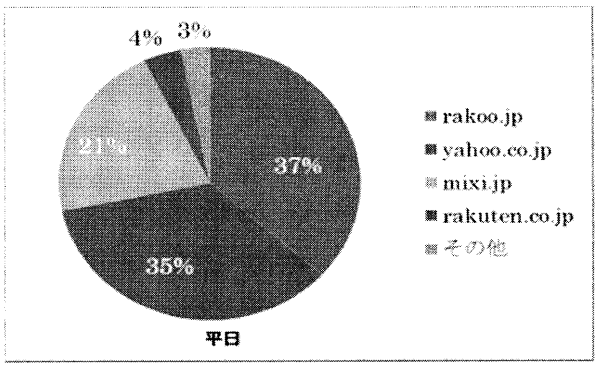

図9 ユーザBの平日アクセスドメイン割合

図3，4，8，9より，ユーザAとユーザBの間では頻繁にアクセスするドメインが大きく異なることがわかる．ユーザA ではアクセスの大部分がポータルサイトへ集中しているのに対し，ユーザB の場合，ポータルサイトだけでなく，SNS (Social Networking Service) へ多数アクセスする傾向にある．この結果，ユーザが頻繁にアクセスするドメインが本人認証に有効なパラメータとなる可能性があると言える．

4. 実験結果のまとめと課題の考察

4.1. 実験結果のまとめ

本検討で対象とした20名のユーザのアクセスピーク時刻、アクセスドメインを以下の表2に示す．ランダムに選んだユーザをユーザA～Tとし，それぞれのユーザに対して休日，平日におけるアクセスピークの時刻と頻繁にアクセスするドメインを示す．表中の「なし」の表記では，特徴を抽出可能なアクセスピーク時刻，アクセスドメインを得られなかったことを意味する．

Web 閲覧履歴情報のうち，平日の平均Web接続時間，平日，休日のアクセスドメインはユーザの行動特性，生活規則を抽出，本人認証に利用できる有用なパラメータとなる可能性があることが確認できた．しかし表2より，適切なパラメータであるといっても，特にWebに頻繁にアクセスしないユーザの場合，平均Web接続時間やアクセスドメインに特徴がまったく現れないケースがある．この結果から，頻繁にWebへアクセスするユーザの場合は平均Webアクセス時刻やアクセスドメインはユーザの特徴を示す有効なパラメータであるが，頻繁にWebにアクセスしないユーザの場合，Web閲覧履歴情報のみでは本人の特徴を抽出することが非常に困難であることがわかった．これは認証に用いるパラメータとして，Web閲覧履歴情報だけでなく，複数のライフログ情報を組み合わせる多要素認証を用いることで解決できる可能性がある．例えばGPSによる位置情報と組み合わせることを考えると，Webにアクセスしないユーザであっても，日々の通勤経路情報などからユーザを識別できることが考えられる．逆に通勤経路が重なるユーザであってもWeb閲覧履歴情報からユーザを識

別することが可能になることも考えられる．このように単一のライフログでは識別しきれない場合でも，複数のライフログを用いる多要素認証を適用することで対応可能になることが考えられる．

表2．実験対象ユーザのアクセスピーク時刻，アクセスドメイン

ユーザ	アクセスピーク時刻(時)		頻繁にアクセスするドメイン
	平日	休日	
A	19	なし	yahoo.co.jp
B	0	なし	rakoo.jp, yahoo.co.jp
C	なし	23	yahoo.co.jp
D	なし	なし	super-miracle.com
E	6～7	6～7	edita.jp
F	2	2	mixi.jp, yahoo.co.jp
G	21	なし	google.co.jp, yahoo.co.jp
H	22	23	rakuten.co.jp
I	なし	なし	なし
J	21	なし	rakuten.co.jp
K	22	なし	google.co.jp
L	なし	なし	なし
M	0	なし	yahoo.co.jp, google.co.jp
N	21	21	yahoo.co.jp, jword.jp
O	11	11	yahoo.co.jp
P	なし	なし	なし
Q	なし	なし	hirosefx.jp
R	2	2	mixi.jp,
S	なし	なし	yahoo.co.jp
T	10	3	yahoo.co.jp

以上の実験結果から，Web閲覧履歴情報を本人認証に適用するにあたり，課題が残されていることから，一般にライフログを本人認証に適用するにあたっても課題が残されていると考えられる．そこで，ライフログを認証に用いる際の要求条件，ライフログ自体に起因する課題，運用上の課題に着目し，考察した．

4.2 ライフログを本人認証に用いる際の課題

4.2.1. ライフログを用いる際の要求条件

ライフログを本人認証に用いるにあたり，考えられる要求条件として，情報の継続性がある．ライフログとは個人の生活に密着した情報であり，時々刻々と変化する情報である．このため，変化する情報を洩れなく取得する必要がある．これにはWeb閲覧履歴や位置情報のように自動で継続的に収集できるライフログが望ましい．これは食事情報[6]や睡眠情報のような手動で登録が必要なライフログ情報では，個人の特徴を抽出できる可能性があるが，情報の登録忘れや作業が手間になり，情報の登録を行わなくなってしまうことが考えられるためである．

4.2.2 ライフログ自体に起因する課題

4.2.2.1. 認証精度の問題

ライフログとは時々刻々と変化する情報であるため，これを本人認証に適用することを考えた場合，パスワードや暗証番号のような認証方式に対して頻繁な情報の更新という長所があるものの，それゆえに常に一定の認証精度を担保することができない可能性がある．例えば本検討で用いたWeb閲覧履歴情報の場合，普段動画サイトを利用しないユーザが，サッカーのワールドカップの際に見ることができなかった試合を動画サイトで閲覧するなど，何かしらの外的要因がもとで普段と異なるライフログとなることがある．結果，認証に失敗するケースが考えられる。このように単一のライフログを用いた認証では認証精度に問題が生じることが考えられる．

4.2.2.2. ライフログの初期値問題

ライフログとはそもそも記録，蓄積された過去のデータであるため，初期値が存在しない．このためライフログを用いた本人認証システムを導入しても，ライフログのみを用いた認証ではデータが収集できるまで認証が不可能になることが考えられる．

4.2.3. 運用上の課題

4.2.3.1. プライバシーの問題

ライフログとは個人の生活，あるいは行動特性によって記録されるデータである．これらは当然個人情報に含まれることは言うまでもなく，適切に運用，管理される必要がある．ユーザの所持する端末のみにライフログが保存されている分には，ユーザの管理に依存するが，認証に用いるライフログを外部のデータベースサーバに保存する場合，自身のライフログ情報を第三者に預けることに対して抵抗のあるユーザもいるだろう．このようなユーザに対して安全，安心な管理環境を提供する必要がある．また，断片的な個人情報を収集し，一元化することで完全な個人情報を復元する名寄せの犯罪対策も必要となる．本検討ではライフログを用いて本人認証を試みることから，この名寄せが行われてしまうと，ユーザの個人情報が漏えいしてしまうばかりか，認証アタックが行われる危険性もある．

4.2.3.2. ライフログの保存期間，参照期間の問題

様々なライフログデータが増え，ユーザの数が増えればそれに伴いライフログを蓄積するデータベースの容量が増える．これはリソースを大きく圧迫する可能性があり，最低でもどの程度の期間，種類のログを保存しておけば本人認証に有効であるか検証する必要がある．参照するライフログの期間が短く，種類が少なければリソースに与える負荷は少なくなるが，本人以外を認証してしまう他人受入率が上昇してしまう可能性がある．一方でライフログの参照期間が長く，種類が多すぎるとリソースを圧迫し，さらには識別精度を低下させることも考えられる．4.2.2節で述べたように

ライフログは時々刻々変化する情報であるため，参照するライフログが多いほど変化するライフログの数が増加する．その結果，一致するライフログが減少し，認証に失敗するケースが増加すると考えられるためである．このため適切な参照ログの期間，種類を検討する必要がある．

5. まとめと今後の検討

5.1. まとめ

本検討ではWeb閲覧履歴情報の中から,Web接続時間，アクセスドメインをパラメータとして選択し，ユーザの特徴を抽出，本人認証に適用することが可能であるか検討した．結果から，特に平日における平均Web接続時間，休日，平日のアクセスドメインからユーザの特徴を抽出でき，かつその特徴がユーザごとに異なる可能性があることも確認できた.しかしながらWeb閲覧履歴情報だけではWebに頻繁にアクセスしないユーザの場合，本人の特徴を抽出することができず，ユーザを識別することが困難であった．このことから，単一のライフログのみによる本人認証は困難であることが考えられる．そこで、ライフログを一要素とした多要素認証を適用することを検討する.多要素認証であれば，一部のライフログが変化した場合であっても他の要素で補うことで認証を成功させることが可能になることが考えられる．

また，本実験結果より，ライフログを本人認証に適用する際に考えられる課題を，ライフログを用いる際の要求条件，ライフログ自体に起因する課題，運用上の課題の観点から考察した．

5.2. 今後の検討

本検討より,Web閲覧履歴情報のみによる本人認証では，Webに頻繁に接続するユーザである場合，ユーザの特徴を抽出できる可能性があり，認証のパラメータとして有用であることが確認できたが,Webに頻繁に接続しないユーザであった場合，ユーザの特徴が抽出できない可能性がある．そのため，複数のライフログを組み合わせる多要素認証を用いることで，ユーザをより正確に識別することが可能であるか検討する.GPSによる位置情報など，より認証精度を向上させるためのそのほかのライフログについて検討を行う．

文献

[1] 茂木学，永徳真一郎，望月理香，八木貴史，武藤伸洋，“ライフログを活用した情報閲覧・アクセス方法の提案,”IEICE technical report 110(42), pp.35-40, 2010

[2] 角田雅照，伏田享平，三井康平，亀井靖高，後藤慶多，中村匡秀，松本健一，“位置と速度を利用した移動体向け認証方式の提案，”電子情報通信学会技術研究報告．MoMuC，モバイルマルチメディア通信，pp.11-16, 2006

[3] 垣正勝，小池誠，“ユーザの生活履歴を用いた認証方式-電子メール履歴認証システム，”情報処理学会論文誌，Vol 47, No.3, pp.945-956, 2006

[4] 相澤清晴，”ライフログとその展望，”ライフログ～役に立つために～，映像情報メディア学会誌，Vol.63, No.4, pp.445-448

[5] 保史生，”ライフログの定義と法的責任　個人の行動履歴を営利目的で利用することの妥当性，”情報管理，Vol53, No.6, 2010

[6] 橋克巳，廣田啓一，千田浩司，柴田賢介，五十嵐大，濱田浩気，山本太郎，畑島隆，間形文彦，”ライフログ活用への社会科学的アプローチ，”NTT技術ジャーナル，pp.24-28, 2010.7