# ソーシャルメディアのプライバシー保護と攻撃に関する情報処理モデルの研究

片岡　春乃

電気通信大学大学院　情報理工学研究科

博士（工学）学位申請論文

2015 年 9 月

# ソーシャルメディアのプライバシー保護と攻撃に関する情報処理モデルの研究

博士論文審査委員会

主査　吉浦　裕　　教授

委員　市川　晴久　教授

委員　兼子　正勝　教授

委員　坂本　真樹　教授

委員　高橋　裕樹　准教授

Copyright (C) 2015 Haruno KATAOKA All Rights Reserved.

# Abstract

Computational Model for Protection and Infringement of Privacy in Social Media

Social media are forming an important infrastructure for communications in modern society but are also causing revelations of private information. Although countermeasures such as privacy settings are being taken against this problem, these countermeasures often impede the natural flow of communication. Another problem is that information disclosed in social media is linked with different kinds of information, such as information in a database, leading to privacy infringement. However, the reality of this risk has not yet been clarified. The objectives of our research are therefore to protect privacy without impeding communication and to clarify the risks of privacy infringement by linking different kinds of information.

Representing and detecting private information are central issues in both protection and infringement of privacy. It is because private information depends on individuals and is ambiguous while it has a wide variety of expressions in social media.

To protect privacy without impeding communication, we proposed a system that checks sentences that users may have posted in social network without considering privacy, detects revelation of private information, and warns the user or modifies the sentences. This detection is word-wise, not sentence-wise, to localize modification so that the original sentence is maintained and communication is not disturbed.

Because private information depends on each user, the user has to define his/her private information so that the system can detect it. However the private information is often ambiguous and difficult to define. Therefore burden on users in defining private information should be minimized while private information in a wide variety of expressions, such as combinations of words, should be detected. To meet these two requirements, DCNL lets users define their private information with phrases called NG words and infers the relation between various expressions in the user's posts and these NG words. The system uses words in the posts as keywords for Web retrieval and infers relation to the NG words on the basis of appearances of the NG words in the retrieval results. We evaluated the proposed detection method using 11,000 tweets from 11 student volunteers (1000 tweets/volunteer) and 7047 sentences from the mixi diaries of an adult worker. The system detected 90% of sentences where humans detected revelation of private information. Moreover 10% of revealing

sentences were detected by the system and not by the humans. Thus the system is not only effective in preventing revelation of private information but also effective in complementing human ability.

To clarify the risk of linking different kinds of information, we designed a system to identify tweets posted by a specific person by linking them with that person's resume. Terms in a resumes are rarely used in tweets of the corresponding person but are indirectly implied in a wide variety of expressions. The Web-based inference method proposed above is used to detect hidden relations between tweets and a resume by taking the resume as a set of NG words. This identification system was evaluated using the resumes and 12000 tweets from 12 volunteers. Tweets of 8 volunteers were correctly identified out of tweets of 101 persons including each volunteer and 100 twitter users. Tweets of 3 volunteers were included 4 candidates. Though these results, we have shown the risks of privacy infringement by linking different kinds of information.

Through designs and evaluations of these two systems, we have established a computational model for detecting private information and have shown its effectiveness in both protection and infringement of privacy.

# 和文要旨

ソーシャルメディアは，現代社会のコミュニケーション基盤としての地位を確立しつつある一方，プライバシー情報や機密情報の漏洩問題を引き起こしている．これに対して，公開範囲の設定などのプライバシー保護対策が実施されているが，これらの対策はソーシャルメディアの本来の目的であるコミュニケーションの活性化と相反している．また，ソーシャルメディアのプライバシーに対する攻撃として，近年，メディアに公開された情報がデータベース等の異種の情報と照合されることで，個人特定等に至ることが懸念されているが，そのリスクの現実性は明らかになっていない．そこで，本研究では，ソーシャルメディアにおいてコミュニケーションとの両立が可能なプライバシー保護および，異種情報の照合による攻撃リスクの明確化を目的とする．

上記のプライバシー保護と攻撃の両面において，プライバシー情報すなわち個人特定などのプライバシー侵害に至る可能性のある情報の扱いが共通課題となる．特に，プライバシー情報は個人に依存し範囲が曖昧である上，ソーシャルメディアで多様な表現形態を有するため，これを計算機で定式的に処理することが大きな問題となる．

コミュニケ―ションとの両立が可能なプライバシー保護については，プライバシー保護の新しい形態として「自然言語情報の開示制御 DCNL（Disclosure Control of Natural Language Information）」を提案する．DCNL は，ユーザがプライバシーを考慮せずにメディアに投稿した文章を検査して，プライバシー情報の漏洩個所を検知し，警告あるいは該当箇所の言い換えを行う．その際，漏洩個所を局所化することで，ユーザの文章をできるだけ変更せず，コミュニケーションを妨げないようにする．

DCNL の中核は，投稿文からのプライバシー情報の検知であるが，その実現には以下の問題に対応する必要がある．

(1)プライバシー情報はユーザによって異なるので，ユーザが定義する必要があるが，プライバシー情報の範囲は曖昧であるため，厳密な定義は困難であり，ユーザへの負担となる．

(2)投稿文においてプライバシー情報は多様な表現形態を有する．勤務先名称などの固有名詞の直接的な記載以外に，複数の単語の組合せによる表現や間接的な示唆もある．

上記の問題により，ユーザによるプライバシー情報の定義の簡易化と，システムによる多様な表現の検知を同時に満たすことが課題となる．そこで，ユーザが自分のプライバシーを NG ワードと呼ばれる単語によって定義し，投稿文中の多様な表現と NG ワードとの結びつきをシステムが推定する手法を提案する．この推定では，投稿文中の単語およびその組み合わせをキーワードとして Web 検索を行い，検索結果に NG ワードが含まれる位置と回数に基づいて，結びつきを定量化する．11 人の学生の Twitter のつぶやき各 1000 件および 1 名の社会人の mixi の日記 7047 文をサンプルとして，通学・通勤先および職種

情報の検知精度を評価した結果，人間が情報漏洩と判断した文章の約 90%を検知できる一方，人間が見逃していた情報漏洩を検知したケースが，情報漏洩した文章数の 10%程度に達した．このことから，提案システムは，本評価実験の範囲では，プライバシー情報の漏洩防止手段として実用的であるだけでなく，ユーザの注意力を補えることが明らかになった．

異種情報の照合によるプライバシーリスクの明確化については，Twitter のつぶやきと個人の履歴書との照合を通じて，多数のつぶやきの中から特定個人のつぶやきを抽出する技術を検討した．つぶやきでは，履歴書の情報（たとえば勤務先の固有名詞）が直接表現されることは少ないが，関連情報（仕事の内容や地名）は表現される．そこで，つぶやきと履歴書の間接的な結びつきを検知するために，履歴書を NG ワードの集合とみなし，上述した Web 検索による多様な表現への対応技術を拡張して用いた．学生 10 人，社会人 2 人の履歴書とつぶやきをサンプルとして，履歴書の人物を含む 101 人の Twitter ユーザのつぶやきから，履歴書の人物のつぶやきを特定できるか評価したところ，12 人の被験者のうち 8 人については 101 人の中から 1 人に特定でき，3 人については 101 人中 4 人に絞り込むことができた．また，被験者のつぶやきの数が増える程，特定精度が高まることが明らかになった．これにより，たとえば就職希望者や社員等の履歴書を用いることで，当該人物のつぶやきを特定し，交友関係や日頃の言動を調査できることを示し，異種情報の照合による社会的リスクの現実性を明らかにした．

以上の研究を通じて，プライバシー情報という曖昧性，個人性の大きい情報を検知するための計算モデルとして，個人毎のプライバシー情報を NG ワードによって表現し，Web 検索を通じてプライバシー情報の多様な表現と NG ワードを照合するモデルを提案し，その有効性を明らかにした．また，この計算モデルがプライバシーの保護と攻撃の両面に利用できることを明らかにした．

# 目次

目次 ..... v
第 1 章　序論 ..... 1
1.1　ソーシャルメディアのプライバシー問題 ..... 1
1.2　先行研究の概要 ..... 2
1.3　本研究の目標と論文の概要 ..... 4
第 2 章　コミュニケーションとの両立が可能な プライバシー保護技術 DCNL ..... 5
2.1　研究背景と目的 ..... 5
2.2　先行研究 ..... 5
2.3　研究目標 ..... 7
2.4　実例日記の分析 ..... 7
2.5　システム要件とシステム構成 ..... 10
2.6　システムイメージ ..... 12
2.7　まとめ ..... 13
第 3 章　ソーシャルメディアへの投稿文からの プライバシー情報の検知 ..... 14
3.1　研究背景と目的 ..... 14
3.2　先行研究 ..... 14
3.3　プライバシー情報 ..... 14
3.4　実例投稿文の調査・分析 ..... 17
3.4.1　調査対象 ..... 17
3.4.2　調査方法 ..... 17
3.4.3　調査結果 ..... 17
3.4.4　人間によるプライバシー情報検知モデル ..... 19
3.5　技術課題 ..... 22
3.6　システム設計と実装 ..... 22
3.6.1　基本方針 ..... 22
3.6.2　アルゴリズム ..... 24
3.6.3　想起検知アルゴリズム ..... 26
3.6.4　実装 ..... 28
3.7　評価 ..... 30
3.7.1　評価方法 ..... 30
3.7.2　学生サンプルに関する評価結果 ..... 31
3.7.3　False Positive の再分析 ..... 36

3.7.4 False Negative の分析 ..... 38
3.7.5 社会人サンプルに関する評価結果 ..... 39
3.7.6 ユーザインタフェースについて ..... 41
3.8 関連研究 ..... 43
3.9 まとめと今後の課題 ..... 43
第 4 章 履歴書との照合を通じた ソーシャルメディア上の注目者の発言の特定 ..... 46
4.1 研究背景と目的 ..... 46
4.2 先行研究 ..... 47
4.3 履歴書からの発言特定システム ..... 48
4.3.1 提案システム概要 ..... 48
4.3.2 発言特定のモデル ..... 49
4.3.3 類似度算出の課題 ..... 50
4.3.4 類似度算出アルゴリズム ..... 51
4.4 発言特定方式 ..... 51
4.4.1 基本的な方式 ..... 51
4.4.2 Web 検索の利用 ..... 55
4.4.3 改良方式 ..... 56
4.5 予備評価 ..... 59
4.5.1 概要 ..... 59
4.5.2 実験 1 ..... 59
4.5.2.1 サンプルデータ ..... 59
4.5.2.2 評価 ..... 59
4.5.3 実験 2 ..... 62
4.5.3.1 サンプルデータ ..... 62
4.5.3.2 手法 ..... 63
4.5.3.3 評価 ..... 63
4.5.4 考察 ..... 70
4.5.5 評価のまとめ ..... 70
4.6 社会的リスクに関する考察 ..... 71
4.7 まとめと今後の課題 ..... 72
第 5 章 結論 ..... 74
5.1 まとめ ..... 74
5.2 今後の課題 ..... 76
謝辞 ..... 78
参考文献 ..... 79

関連論文の印刷公表の方法および時期 ..... 83
その他の研究業績 ..... 85

# 第1章

# 序論

## 1.1　ソーシャルメディアのプライバシー問題

人の関係を情報ネットワーク上で構築し，日記やつぶやきなどのテキスト情報や画像，動画，位置情報などを流通させるソーシャルメディアの利用および社会における役割が拡大している．例えば，世界最大規模のソーシャルネットワークサービス（以下，SNS）である Facebook の月間アクティブユーザ数（以下, MAU）は 12.8 億人[1], Twitter の MAU は 2.41 億人となっている[2]．また，インスタントメッセンジャーアプリとして普及した LINE は，国内登録ユーザが 5000 万人を超え，日本の人口の 37%以上をカバーしたと発表している[3]．サービスの拡大と共に，ソーシャルメディアの利用目的も広がっている．友人とのコミュニケーションや情報交換だけでなく，企業内コミュニケーション，就転職活動，マーケティングやブランド形成，さらには政治や民主化運動にも活用されている．

このように，ソーシャルメディアは重要なコミュニケーションインフラとなっているが，一方では，プライバシー情報や組織の機密情報の漏洩，誹謗中傷などの多くの問題を引き起こしている．

この問題に対して，サービスによっては，ユーザによる公開範囲の設定や，事業者による不適切な書き込みのチェックが行われている．しかし，公開範囲の設定はユーザの負担となり，情報発信のたびに適切な設定を行うことは難しい上，ソーシャルメディアの本来の目的である「コミュニケーションの楽しみ」を損なう．また，事業者によるチェックはコストや人手および通信の自由との関わりもあるため限界がある．

従来の情報セキュリティ技術であるアクセス制御技術の利用も考えられるが，ソーシャルメディアで流通する個人情報は日記やつぶやき，ユーザ同士の会話，画像や動画，位置情報，それらの関係の中に，直接あるいは間接的に表現され，形式化，構造化されていないため，アクセス制御ルールの設定が困難である．さらに，ソーシャルメディア上の情報が他のメディアの異種情報と統合されることで，単独のサービスでは問題とならなかった情報がプライバシー侵害につながる危険性が懸念されている．

## 1.2　先行研究の概要

ソーシャルメディアのプライバシーに関する先行研究は，(1)ソーシャルメディアにおけるプライバシー情報の開示状況調査，(2)プライバシー情報の漏洩防止対策，(3)ソーシャルメディアから個人情報を抽出する攻撃手法の 3 つに分類することができる．ここで(3)の攻撃手法の研究は，個人情報漏洩のリスクを示し，事業者やユーザに対策を促すことを目的としている．

(1)のソーシャルメディアにおけるプライバシー情報の開示状況調査としては，例えば，Gross が Facebook の 4000 人のユーザプロフィールを分析し，99%が公開範囲を設定せず，89%が実名を，88%が誕生日を，51%が現住所を公開していることを明らかにした[4]．その後，Acquisti，Baatarjav らが Facebook について異なる観点からの調査を行い[5][6]，Viegas が Blog について[7]，Meeder が Twitter について[8]の調査を行った．

(2)のプライバシー情報の漏洩防止技術については，各サービスにおいて，テキストや画像の公開範囲をユーザが設定することができる．しかし，公開範囲の設定はユーザの負担となり，情報発信のたびに適切な設定を行うことは難しい上，ソーシャルメディアクの本来の目的である「コミュニケーションの楽しみ」を損なう可能性がある．図 1.1 に Facebook のコンテンツ投稿における共有範囲の設定画面を示す．Gurses は，公開範囲の設定機能を分析し，初級ユーザにとって分かりにくく，煩雑であることを示している[9]．さらに，1 回の開示単位で公開範囲を設定するため，開示するコンテンツのなかに 1 つでも隠したい箇所があれば全てを隠すことになり，残りの内容を相手に伝えることができなくなってしまう．その意味でも，「コミュニケーションの楽しみ」を損なう可能性がある．情報開示を制御する従来技術としては，データベースや OS を対象に開発されたアクセス制御技術がある．しかし，ソーシャルメディアで流通する個人情報は日記やつぶやき，ユーザ同士の会話，画像や動画，位置情報，それらの関係の中に，直接あるいは間接的に表現され，形式化，構造化されていないため，アクセス制御ルールの設定が困難である．


図 1.1 Facebook のコンテンツ投稿における共有範囲の設定画面

(3)のソーシャルメディアから個人情報を抽出する攻撃手法については，Novak らが，2004 年に，著者推定技術を用いて，ブログ上の 2 つの投稿が同一著者のものであることを推定可能であることを示した[10]．その後，Lam が台湾のソーシャルネットワークにおいて，一言メッセージという部分からユーザの本名が推定できることを示した[11]．攻撃の手法研究は，最近では，ソーシャルメディアの情報を他のソーシャルメディアの情報と照合する手法が主流になってきた．たとえば，Narayanan は，ユーザ同士のつながりを表したソーシャルグラフを他のメディアのソーシャルグラフと比較することで，両メディアを利用する同一ユーザを特定できると報告している[12]. Goga は位置情報，タイムスタンプ，writing-style（文の長さ，前置詞の使用率などの文章の特徴）の情報を解析し，その結果を統合することによって異なるソーシャルメディアを利用する同一ユーザの特定を行っている[13]．これらの先行研究においては，ソーシャルメディア同士の照合を行っている．また，これらの先行研究の示すプライバシーリスクは比較的軽微であり，真の社会的リスクとまでは言えない．これに対し，ソーシャルメディアとデータベース（例えば医療データベース）あるいは組織内情報（警察や企業の人事の情報）が照合される場合，さらに大きなリスクが生じると考えられるが，そのようなリスクはいまだ明らかになっていない．

## 1.3　本研究の目標と論文の概要

以上述べた背景と先行研究の状況から，本研究では下記を目標とする．

(1) コミュニケーションとの両立が可能なプライバシー保護技術の検討

ソーシャルメディアの本来の目的である「コミュニケーションの楽しみ」との両立が可能なプライバシー保護技術を確立する．ユーザによる煩雑な事前準備や設定を不要とし，ソーシャルメディアにおけるコミュニケーションを阻害することなく，利用の面白さを失わない方式とする．

(2) 異種の個人情報の照合による攻撃可能性の検討

同一人物に関する異種の情報を照合する攻撃手法を開発し，社会的リスクの存在を明らかにする．これにより，今後のソーシャルメディアのプライバシー対策を行うための知見を得る．

上記のプライバシー保護と攻撃の両面において，プライバシー情報すなわち個人特定などのプライバシー侵害に至る可能性のある情報の扱いが共通課題となる．特に，プライバシー情報は個人に依存し範囲が曖昧であり，ソーシャルメディアで多様な表現形態を有するため，これを計算機で定式的に処理することが大きな問題となる．本研究では，保護と攻撃の検討を通じて，プライバシー情報を扱う情報処理モデルの一例を明らかにしたい．これにより，プライバシー保護の他の研究への波及効果が期待できる．

以下，2 章では，コミュニケーションとの両立が可能なプライバシー保護技術として，自然言語情報の開示制御技術（Discloser Control of Natural Language information: DCNL）の構想を述べる．DCNL は，ユーザがソーシャルメディア上に開示しようとするテキストを検査して，プライバシー情報の漏洩の可能性を検知する．その結果，ユーザへの警告あるいは情報漏洩個所の言い換えを行う．

3 章では DCNL の実現に向け，テキストからのプライバシー情報の漏洩検知方法について述べる．ソーシャルメディアへの実例投稿文を調査，分析したうえで，システム設計し，実装，評価を行う．

4 章では，組織内の履歴書の情報と照合することで，ソーシャルメディア上の特定個人の投稿文を抽出する技術を提案し，実装，評価により，異種情報の照合による社会的リスクの存在を明らかにする．なお，この照合技術の中核として，DCNL のプライバシー情報検知技術を用いている．

5 章では，研究目標に照らして本研究を総括すると共に，今後の課題について述べる．

# 第2章

# コミュニケーションとの両立が可能なプライバシー保護技術 DCNL

## 2.1　研究背景と目的

ソーシャルメディアは重要なコミュニケーションインフラとなっているが，一方では，個人のプライバシー情報や組織の機密情報の漏洩，誹謗中傷などの多くの問題を引き起こしている．ユーザが投稿した文章から犯罪につながる情報が漏洩した事例として，Twitter に旅行中であることを示唆する書き込みをしたユーザが空き巣の被害に遭ったという事件があった[14]．この事例ではユーザが居場所を絶え間なくソーシャルメディアに投稿していたため，泥棒が犯行におよぶことができた．また，SNS の投稿からストーカーが住所や生活ぶりをさぐっていた事件[15]も起きている．さらに，GPS 情報を付与した投稿から，自宅住所を割り出すサービス[16]や，外出先からの投稿から，自宅が不在状態であることを検知するサービス[17]もあり，英国の空き巣の 8 割がソーシャルメディアを活用して犯行の事前準備をしているという調査もある[18].

これに対し，多くのサービスでは，テキストや画像の公開範囲をユーザが設定できるようになっている．しかし，情報発信のたびに適切な公開範囲を設定することは難しく，設定ミス，自分が想定していないユーザからの漏えいの危険性もある．また，公開範囲の設定はユーザの負担となり，ソーシャルメディアの本来の目的である「コミュニケーションの楽しみ」を損なう可能性がある．

そこで，本章では，コミュニケーションとの両立が可能なプライバシー保護技術として，自然言語情報の開示制御技術（DCNL）の構想を述べる．

## 2.2　先行研究

ソーシャルメディアにおけるプライバシー情報の開示状況やユーザのプライバシー意識の調査が多数行われている．Gross は Facebook の 4000 人のユーザプロフィールを分析し，99%が公開範囲を設定せず，89%が実名を，88%が誕生日を，51%が現住所を公開していることを明らかにした[4]．また，Viegas は，ブログの投稿における人々のプライバ

シー意識を調査し，55%のブロガーが実名を，21%が友人の名前を，66%が無断で友人のことを書き，76%が公開範囲を設定していないことを明らかにしている[7]．Acquisti は Facebook のユーザの 84%がプロフィールに実名を，39%が携帯電話の番号を，28%がパートナーの氏名を記載していると報告している[5]．Baatrajav は，ソーシャルメディアにおいてユーザの出身地，政治観，知り合いなどのプライバシー情報の開示状況を調査している[6]．Meeder は，Twitter の 27 億のつぶやきと 8000 万人のユーザプロフィールを分析し，フォロワーのつぶやきを retweet したつぶやきが増加傾向にあると発見した[8]．元のつぶやきがユーザのフォロワーのみに開示されていても，フォロワーの retweet によってその範囲を超えて公開されてしまうという問題がある．インターネット白書 2011 によると，Twitter への書き込みの内容においてアンケート調査を行った結果，39.3%のユーザが「ほとんどがプライベートな内容である」と，また 37.8%のユーザが「プライベートな内容の方が多い」と回答している[19].

情報開示を制御する従来技術としては，データベースや OS を対象に開発されたアクセス制御技術がある[20]．しかし，ソーシャルメディアで流通する個人情報は日記やつぶやき，ユーザ同士の会話，画像や動画，位置情報，それらの関係の中に，直接あるいは間接的に表現され，形式化，構造化されていないため，アクセス制御ルールの設定が困難である．アクセス制御の対象をテキスト情報だけに限定したとしても，人間のコミュニケーションでは，相手や状況に合わせて様々な言葉とその組み合わせを使用する．そのため，ソーシャルメディアにおいてもすべての言葉の組み合わせ，相手や状況との組み合わせに対応したアクセス制御ルールを網羅的に定義する必要があるが現実的に困難である. さらに，将来話題になる内容やそこで使われる言葉を事前に予測することは難しく，事前に定義するのは困難である．以上の理由により，アクセス制御は，ヒューマンコミュニケーションにそのまま適用できないといえる．

多くのサービスでは，テキストや画像の公開範囲をユーザが設定できるようになっている．2011 年に Liu らによって行われた調査では，公開されるコンテンツのうち 36%がデフォルト設定の「全てのユーザに公開」(to All Facebook Users) であり，半分近く（49%）の設定範囲が「友人のみ」(All Friends)または「一部の友人のみ」(Selected Friends)であると報告している[21]．しかし，公開範囲の設定は，プライバシー保護が不十分である．例えば，芸能人であるユーザが公開範囲の設定を誤ったことで，個人情報が漏えいした事件[22]などの事件が起きている．2013 年には IPA（情報処理推進機構）が，公開設定の確認が不十分だったことにより省庁，教育機関の機密情報が漏えいした報道が相次いだことを受け，注意喚起を行っている[23]．このように，公開範囲の単純な設定ミスだけではなく，公開範囲を友人の友人までに限定した場合であっても，それぞれのユーザに平均 53 人の友人がいる[24]と言われているため，実際には約 2800 人のユーザに公開されること

になる．公開される 2800 人のユーザがそれぞれどのような人物であるか，自分の投稿をどのように扱うかを網羅的に把握することは難しく，自分が想定していないユーザの挙動が漏えいの原因となっている．

また，公開範囲は，コンテンツの開示毎に設定するため，コンテンツのなかに 1 つでも隠しておきたい箇所があれば，すべてを隠すことになり，残りの内容を相手に伝えることができなくなってしまう．その上，コミュニケーションの楽しみとの両立が困難である．つまり，プライバシーを優先させればソーシャルメディアの目的であるコミュニケーションが断絶される可能性が高い．

## 2.3　研究目標

ソーシャルメディアにおけるコミュニケーションとの両立が可能なプライバシー保護技術として，ユーザが十分配慮せずにソーシャルメディアへ投稿した文章を自動的にチェックし，文章の意味を維持しながら相手によって必要な開示制御を行う DCNL（Discloser Control of Natural Language information）を提案する．すべての語句について事前に網羅的にルールを設定していなくても，高い確率で投稿文からプライバシー情報の漏えいを自動的に検知して，ユーザに警告，言い換えをする．さらに，単語単位で制御することで，投稿文章全体を見せる，見せない，の二者択一にすることを避ける．これにより，ユーザに手間をかけず，複数の相手と 1 つの話題を共有してコミュニケーションを図るというメディア特有の面白さを維持することができる．

## 2.4　実例日記の分析

実際に SNS に投稿された日記を分析し，DCNL のシステム要件を明らかにする．

ここでは SNS の例として mixi[25]を用いて分析する．mixi には，ユーザ毎にユーザページがあり，ページの閲覧者を，自分，友人，友人の友人，全体の 4 つのクラスに分けている．例えば，ユーザのプロフィールページにある氏名や所属のような基本的な個人情報は，それぞれどのクラスの閲覧者までアクセス可能か設定できる．また，投稿される日記については，投稿時に公開範囲を設定することが可能である．しかし，日記ごとに公開範囲を設定することは相手によってメディアを変えることと同義であり，ユーザの手間となりメディアの利用を阻む恐れがある．

次の文章は mixi ユーザである電気通信大学人間コミュニケーション学科 4 年生の女子学生「さやか」が 2006 年 12 月から 2007 年 2 月に投稿した日記からの抜粋である．

(文 1)

「来週の就職説明会は西 6 号館でやるらしい.」

(文 2)

「昨日，調布駅で聡子に会った．やっぱり卒業研究が大変みたい.」

(文 3)

「恭輔とお台場に『鉄コン筋クリート』を観に行った.」

文 1 について分析する．さやかは自身の所属情報で｛電気通信大学｝,｛人間コミュニケーション学科｝等を公表していないので，日記文章中にも大学名を明記していない.「西 6 号館」は一見すると問題の無い言葉である．それは，西 6 号館と呼ばれる建物が多数存在し，そこから大学名が推測されると考えにくいからである．しかし「西 6 号館」をキーワードとして Web 検索エンジン Google[26]で検索すると，図 2.1 に示すように，検索結果上位に電気通信大学に関わるサイトが現れる．このため，所属が明らかになる可能性が高い．つまり，ユーザがプライバシー情報の漏えいにつながると思っていなかった言葉であっても，Web 検索により得られた知識を介して，プライバシー情報を想起させる可能性があるといえる. この問題は, 原理的には, 制御ルールを事前に定義することで避けられる. しかし，プライバシー情報を洩らす可能性がある全ての語句を予測するのは難しい上，ユーザの手間が増えれば，SNS を使用する楽しみが損なわれる.

文 2 の「調布」と「卒業研究」の組み合わせは，調布で卒業研究を行う唯一の大学，電気通信大学を想起させる．つまり，それぞれの単語は比較的安全でも，組み合わせることで問題表現になる場合がある．全ての組み合わせに対して事前に制御ルールを決めることは更に困難である.

文 3 からは彼女の人間関係が明らかになる可能性がある.「お台場」は，デートスポットとして有名である．そのため，さやかと「恭輔」がデートに行くような親密な関係にあることを想起させる可能性がある．これは，語句が間接的に表す事象まで考える必要があることを示している.

これらの例文から，開示制御ルールを事前に定義することが非常に困難であるといえる.


図 2.1　西 6 号館の Web 検索結果

## 2.5　システム要件とシステム構成

以上を踏まえ，DCNL のシステム要件を見出した．

(1) ソーシャルメディアに投稿された文章を検査し，プライバシー情報を洩らす可能性があるか検知する．その際，単語単位でプライバシー情報を洩らす可能性があるか検知する．
(2) 単語の組み合わせによってプライバシー情報を洩らす可能性があるか検知する．
(3) 間接的な言葉の意味によってプライバシー情報を洩らす可能性があるか検知する．
(4) プライバシー情報を漏らす可能性を検知した場合，ユーザへの警告，元の意味を維持しより安全な言葉をサジェスト，または自動的に置換する．
(5) ユーザへの負担を最小限に抑える．例えば，友人関係ごとに制御ルールを詳細に定める，異なる文章を書くといった負担をかけない．

DCNL のシステム構成案を図 2.2 に示す．DCNL はユーザとソーシャルメディアの間に位置し，ユーザがソーシャルメディアへアクセスするときに動作する．DCNL は，ソーシャルメディアに投稿された文章からプライバシー情報を検知する処理と，検知後にプライバシー情報の漏えいを防止する処理に大きく分かれる．

システム構成の概要を図 2.3 に示す．プライバシー情報検知処理は，自然言語処理によって原文の語句を認識する．前節の文 1 を例にすると，「来週」「就職説明会」「西 6 号館」などである．次に，プライバシー情報の定義を用いて，これらの語句がプライバシー情報を直接表すか，間接的に想起させるか，無関係かを判断し，プライバシー情報を検知する．このとき，言葉の組み合わせ，間接的な意味を考慮した検知のために「外部情報」として Web 検索やシソーラスを利用する．例えば，西 6 号館は Web 検索の結果上位 10 件に電気通信大学に関わるサイトが出現したため，NG ワードを間接的に想起させる語句として判断する．

漏えい防止処理は，プライバシー情報検知処理の結果を受けて動作する．検知結果をユーザに警告する，言い換える語を提示する，あるいは，自動的に語句を置換する．例えば，西 6 号館の代わりとしては，西地区の建物や学内が候補に挙がる．これを，閲覧するユーザとの関係によって開示制御する．

以上の動作によって，ユーザが十分配慮せずにソーシャルメディアに投稿しても，システムが自動的にチェックし，文章の意味を維持しながら相手によって必要な開示制御を行う．これにより，ソーシャルメディアによるコミュニケーションの活発化を目指す．

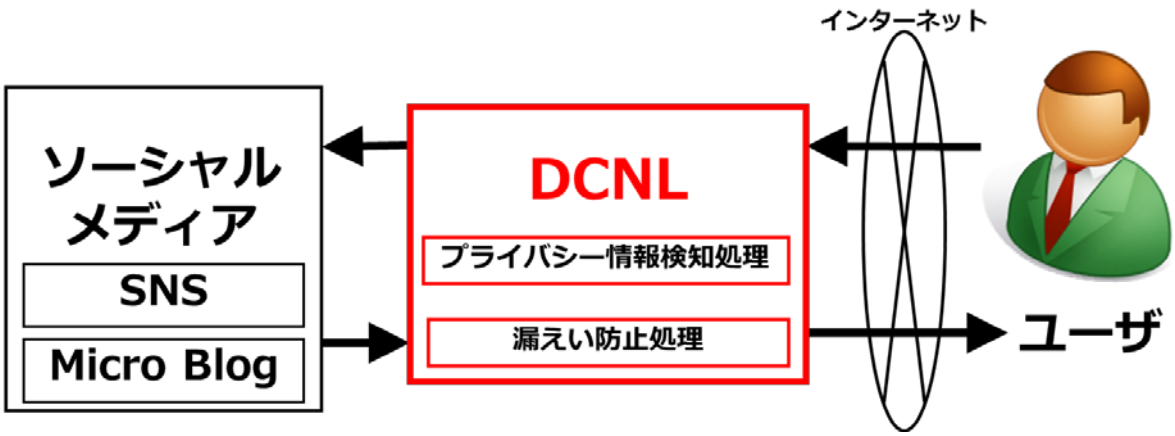


図 2.2　システム構成案

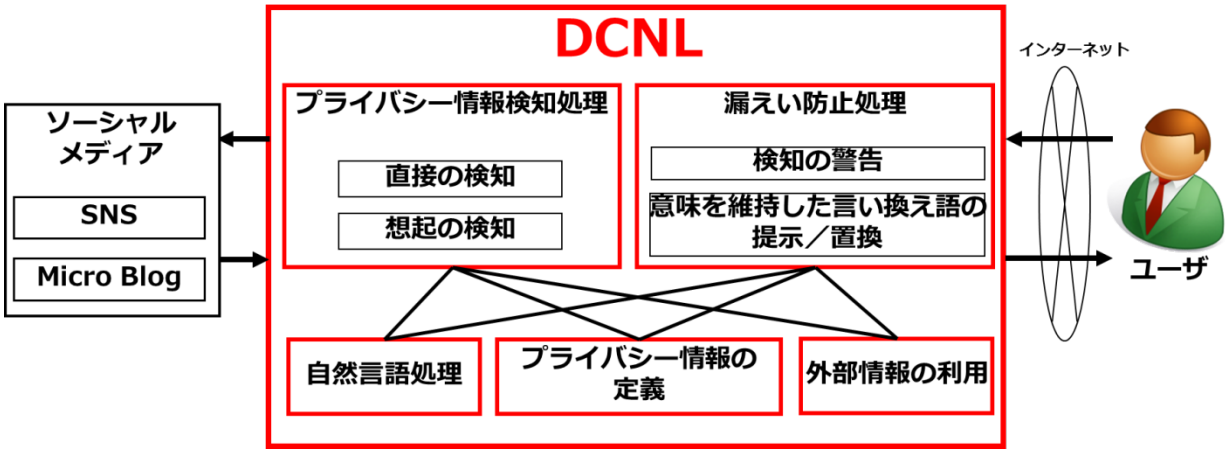


図 2.3　システム構成の概要

## 2.6　システムイメージ

DCNL が実現した際の動作イメージを図 2.4 に示す．ソーシャルメディアに投稿された文章は，DCNL によってプライバシー情報が漏えいする可能性のある言葉が検知され，表示する相手によって，意味を維持したまま適切な語へ言い換えられる．例えば，2.4 節で例示した「来週の就職説明会は西 6 号館でやるらしい」は，2.5 節で示したとおり「西 6 号館」から電気通信大学が想起される可能性が高いため，自分と通学先を知る親しい友人には原文を，それ以外の友人には「西地区」，全体には「学内」にそれぞれ言い換えて表示するのが望ましい．


図 2.4　システム動作イメージ

## 2.7　まとめ

本章ではソーシャルメディアのプライバシー漏えい問題を背景に，公開範囲を投稿単位に設定する従来の対策ではソーシャルメディアの本来の目的であるコミュニケーションを阻害する可能性のあることを指摘し，コミュニケーションとの両立が可能なプライバシー保護技術として，テキスト情報の開示制御技術（DCNL）を提案した．ソーシャルメディアに投稿された実例日記の分析から，DCNL のシステム要件を明らかにし，システム構成を提案した．DCNL は，ユーザがプライバシーに十分配慮せずにソーシャルメディアに投稿しても，システムが自動的にチェックし，文章の意味を維持しながら相手によって必要な開示制御を行う．これにより，ソーシャルメディアによるコミュニケーションの活発化を目指す．

なお，近年相次ぐ SNS，Blog を介した略取誘拐等の事件を背景に，2009 年には青少年インターネット環境整備法が施行された．これは，保護者が不使用の申請をしない限り 18 歳未満の青少年者に対してフィルタリングサービスを提供する義務や，利用者が求めた場合にフィルタリングサービスを提供する義務などを定めている[27]．政府主導でフィルタリングの導入促進を図っている[28]ものの，フィルタリングに対する保護者の無理解，青少年者が規制を免れた別のサイト・メディアを通じて犯罪者とコミュニケーションする可能性が残る．この問題の本質は，コミュニケーションにおける，プライバシー情報の取り扱いにあると考える．青少年者はリテラシーが低く，どの情報がプライバシー情報で，どの表現がプライバシー情報を表すか，正しく判断し難く，自らプライバシー情報を公開する可能性が高い．そこで，コミュニケーションにおける問題表現を検知し，安全に言い換える開示制御 DCNL が，この種の問題に対しても適用可能と考える．

# 第3章
# ソーシャルメディアへの投稿文からのプライバシー情報の検知

## 3.1　研究背景と目的

2章で提案した自然言語情報の開示制御技術DCNLでは，投稿した文章からユーザのプライバシー情報を検知する処理が重要となる．本章ではDCNLの実現に向けて，ソーシャルメディアに投稿された実際の日記からどのようなプライバシー情報が漏洩するか調査し，人間によるプライバシー情報の検知モデルを考察したうえで，システムで検知するための方法を検討する．

## 3.2　先行研究

ユーザが Web へ投稿する文章をチェックする技術として,コンテンツフィルタリングが広く知られている．コンテンツフィルタリングは，あらかじめ不適切な言葉がリスト化されており，ユーザがサイトにアクセスする際や，文章を投稿する際，不適切な言葉が含まれるかどうかリストとのキーワードマッチを行い，アクセスや投稿の可否を決定する．企業内のプロキシサーバや,子ども向けにブラウザの拡張機能として提供,利用されている．

しかしながら，原則としてあらかじめ設定された不適切な言葉のみを検知するため，ソーシャルメディア上のコミュニケーションで用いられる多様な表現に対応しているとは言えない．

## 3.3　プライバシー情報

プライバシーは，一人でいる権利[29]，自分自身または所属するグループへのアクセスコントロール[30]，自分自身の情報へのコントロール[31]など，時代と共に様々な定義が行われている．本研究は，ソーシャルメディアにおけるユーザのプライバシー情報保護，プライバシー情報漏えいの防止を目的としている.そのため,プライバシーの定義として，自分自身の情報へのコントロールを採用する．

NEC 総研『現代人のプライバシー』[32]では，1134 人の生活者に対してアンケート調査し，32 種類の情報を「個人情報に該当すると感じるもの」,「プライバシーに該当すると感じるもの」に分類している（図 3.1）．本研究では，このうち画像情報（顔写真，自分が映った写真や映像，指紋）を除いた 25 種類の情報を自分自身の情報，すなわちプライバシー情報とする．また，これら 25 種類をプライバシーカテゴリと呼ぶ．プライバシーカテゴリには，氏名，住所などの基本的情報，職歴などの社会的属性情報，家族などの個人的属性情報，支持政党や信教などの内的情報が含まれている．

また，プライバシー情報の範囲は個人によってそれぞれ異なるといえる．たとえば，電気通信大学の教員は,自分の大学の学生に対して職業を非公開にする必要がない.しかし，ある教員は休日に参加する趣味の集まりにおいては，自分が電気通信大学の教員であることを隠しているかもしれない．あるいは，隣家の住人に対して，教員であることを公開しても電気通信大学に所属していることを隠している教員もいるかもしれない．実世界ではこのように，個人によってコミュニティに応じたプライバシー情報の範囲をそれぞれ設定し振る舞っている．これはソーシャルメディアにおいても同様である．実世界での面識の有無の関わらず，職業や本名をオープンにする人もいれば，実世界で面識がある人に対しても個人が特定されないよう振る舞いたい人もいるかもしれない．

このように，プライバシー情報は様々な種類があり，その範囲は個人によってそれぞれ異なると言える．そのため，検知するべきプライバシー情報の範囲をユーザ毎にカスタマイズ可能なシステムとして開発するべきである．

なお，OECD8 原則[33]をもとに策定された個人情報保護法では,「個人情報」を生存する個人に関する情報(識別可能情報)として定義している．個人に関する情報を判例で見ると「思想，宗教，意識，趣味等に関する情報，心身の状態，体力，健康等に関する情報，資格，犯罪歴，学歴等に関する情報，職業，交際関係，生活記録等に関する情報，財産の状況，所得等に関する情報など，個人に関する全ての情報が含まれる」となっている[34].

| 分類①：個人情報、プライバシー双方に該当するという意識が高い情報 |
| --- |
| ID番号（住民票コード・免許証番号・保険証番号・社員番号等） |
| 年収や貯金の情報 |
| 指紋（画像） |
| 携帯電話番号 |
| 病歴 |
| 顔写真（画像） |
| 電話やメールの通信履歴 |
| アドレス帳の内容 |
| 家族に関する具体的な事柄 |
| 学歴職歴 |
| 身長や体重 |
| 電話やメールの通信の内容 |
| 家族構成 |
| 分類②：個人情報という意識は高いが、プライバシーという意識は低い情報 |
| 氏名 |
| 住所 |
| 自宅の電話番号 |
| 生年月日 |
| メールアドレス |
| 職業 |
| 出身地 |
| 分類③：プライバシーという意識は高いが、個人情報という意識は低い情報 |
| 自分の部屋の様子 |
| ホームページの閲覧記録 |
| スケジュール←時相表現の意味が理解できないので不可 |
| 知人や友人に関する具体的事柄 |
| 自分が写った写真や映像（画像） |
| 商品やサービスの購入記録 |
| ブックマークの中身 |
| 支持政党や宗教 |
| 分類④：個人情報・プライバシー双方に該当するという意識が低い情報 |
| 趣味 |
| 好きな本や音楽や映画 |
| 好きな歌手やタレント |
| よく読んでいる雑誌 |

図 3.1 『現代人のプライバシー』による情報の分類

## 3.4　実例投稿文の調査・分析

### 3.4.1　調査対象

50代男性ユーザが2005年10月から2007年5月に投稿した日記(149件56,044文字)と日記に対するコメント(本人の記述375件55,127文字，閲覧者の記述618件82,181文字)を対象に，プライバシー情報がどの程度含まれているか調査し，分類した．ユーザは都内理工系私立大学の50代男性教授である．なお，ユーザ本人の希望により所属等の実名称は全てユーザとは無関係の名称（電気通信大学等）に置き換えて表記する．

### 3.4.2　調査方法

筆者らが対象となる日記とコメントを読み，50代男性ユーザに関わるプライバシー情報を認識し，分類・計数した．その際，プライバシー情報が日記スレッド中に直接記述されている場合と，直接記述されてはいないが，記述された情報をヒントとして認識できる場合を区別した．以下では，直接記述されている場合を「直接」，直接記述されていないがWeb検索や年代早見表の参照によりプライバシー情報への手がかりとなる場合を「想起」と呼ぶ．

### 3.4.3　調査結果

結果を表3.1に示す．調査対象149件中，75.8%（113件）に509件のプライバシー情報が記述されていた．そのうち直接は79.2%，想起は20.8%あり，記述の2割が想起によるプライバシー情報であった．また，直接のうち本人が記述したものは90.8%（366件），閲覧者が記述したものは9.2%（37件），想起のうち本人が記述したものは99.1%（105件）であり，ほとんどは本人が記述したものであった．25種類のプライバシー情報のうち抽出されたプライバシー情報は表3.1より17種類あり，抽出されなかったプライバシー情報は，ID番号，年収や貯金の情報，携帯電話番号，自分の部屋の様子，支持政党や宗教，メールアドレス，ブックマークの中身，自分の部屋の様子，自宅の電話番号であった．

抽出したプライバシー情報の上位5件はそれぞれ職業(137件)，スケジュール(102件)，学歴職歴（51件），家族に関する具体的な事柄（35件）であった．表3.2に職業の直接と想起，スケジュールの直接の文例を示す．

表3.1　プライバシー情報調査結果

| プライバシー情報 | 直接 | | 想起 | | 合計 |
|---|---|---|---|---|---|
| | 本人 | 閲覧者 | 本人 | 閲覧者 | |
| 病歴 | 14 | 2 | 1 | 0 | 17 |
| 電話やメールの通信履歴 | 3 | 0 | 1 | 0 | 4 |
| アドレス帳の内容 | 12 | 0 | 3 | 0 | 15 |
| 家族に関する具体的な事柄 | 26 | 3 | 6 | 0 | 35 |
| 学歴職歴 | 39 | 1 | 11 | 0 | 51 |
| 身長や体重 | 3 | 0 | 0 | 0 | 3 |
| 電話やメールの通信の内容 | 3 | 0 | 3 | 0 | 6 |
| 家族構成 | 38 | 8 | 5 | 0 | 51 |
| 氏名 | 5 | 1 | 10 | 0 | 16 |
| 住所 | 7 | 1 | 4 | 1 | 13 |
| 生年月日 | 13 | 0 | 10 | 0 | 23 |
| 職業 | 86 | 9 | 42 | 0 | 137 |
| 出身地 | 7 | 5 | 2 | 0 | 14 |
| ホームページの閲覧記録 | 1 | 0 | 0 | 0 | 1 |
| スケジュール | 91 | 6 | 5 | 0 | 102 |
| 知人や友人に関する具体的事柄 | 8 | 0 | 2 | 0 | 10 |
| 商品やサービスの購入記録 | 10 | 1 | 0 | 0 | 11 |
| 合計 | 366 | 37 | 105 | 1 | 509 |

表3.2　職業の直接と想起，スケジュールの直接の文例

| プライバシーの種類 | 文例 | 検知されたプライバシー |
|---|---|---|
| 職業（直接） | 今日は大学のPRに高校訪問してきました。大学の先生も色々な仕事があります。 | 大学の先生である |
| | 4月から学科長をやることになり（後略） | 学科長になる |
| 職業（想起） | 私の大学でやっていたICTトライアングルフォーラムが今日で終了します（2006年8月9日） | 「ICTトライアングルフォーラム　2006」をWeb検索<br>→電気通信大学で開催 |
| | 私の所属する人間コミュニケーション学科は、（後略） | 学科名をWeb検索<br>→電気通信大学の学科 |
| | 土曜日は、授業がないため学生の研究指導日にあてています。 | 学生に研究指導をする<br>→教員 |
| スケジュール（直接） | 鉄道を利用してフランクフルト経由でミュンヘンに来ています。 | ミュンヘンに出かけている |
| | 土曜日は、授業がないため学生の研究指導日にあてています。 | 土曜日は研究指導日 |

### 3.4.4　人間によるプライバシー情報検知モデル

人間は日記文章の内容を理解し，自らの知識と照らし合わせることで，文章中に直接記述されているプライバシー情報を認識する．例えば，「電気通信大学に通っています」という文章が記述されていた場合に，プライバシーカテゴリ｛職業｝のプライバシー情報が電気通信大学であることを理解する（図 3.2）．また，自らの知識だけで足りない場合は，書籍や Web 検索など必要に応じて資料を調べる．例えば，上記の例において，「電気通信大学」が「UEC」と記述されていた場合に，Web 検索を行うと「UEC」が電気通信大学の英語略称である事がわかりプライバシー情報が直接記述されている事を認識する（図 3.3）．さらに，文章中の言葉の組み合わせから連想，推論する場合もある．例えば，「土曜日は，授業がないため学生の研究指導日にあてています」という文章の場合，言葉そのものだけを理解するならば「学生」という言葉によって｛職業｝のプライバシー情報が学生であると認識する．しかし，人間は「学生」「研究指導日」という組み合わせから，ユーザは学生ではなく教員ではないか，と推測することができる．また，前述の 2 例では通学先がプライバシー情報となっていたが，本例文では職業がプライバシー情報となる．つまり，人によってプライバシーの範囲は異なることが想定できる（図 3.4）．

加えて人間は，文章中には直接記述されていないプライバシー情報を連想や推論によって認識する事ができる．その際には，必要に応じて書籍を調査したり，Web 検索を行ったりする．例えば，「私の大学でやっていた ICT トライアングルフォーラムが今日で終了します」という文章からは，「ICT トライアングルフォーラム」の開催する大学が分かれば，筆者の大学が明らかになる可能性が高いと推測できる．そこで，キーワード「ICT トライアングルフォーラム」を Web 検索する．そして，検索結果から今年の開催場所が電気通信大学である事を発見し，日記の筆者が電気通信大学に通っている事を推定できる（図 3.5）．この例もまた，1 つの単語から明らかになるとは限らない．2.4 節で挙げた文 2 のように「調布駅」と「卒業研究」との組み合わせから，調布駅付近で卒業研究を行う大学を調査することで電気通信大学を推定することもできる．また，認識したプライバシー情報を記憶し，以後日記文章を閲覧した際の参考にする．

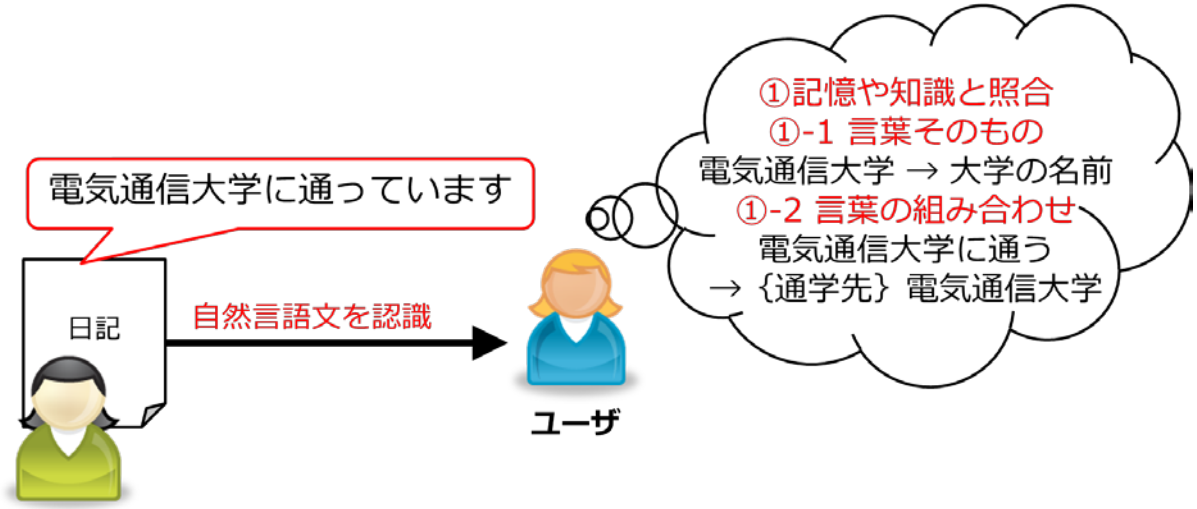


図 3.2　人間による記憶や知識と照合したプライバシー情報の認識

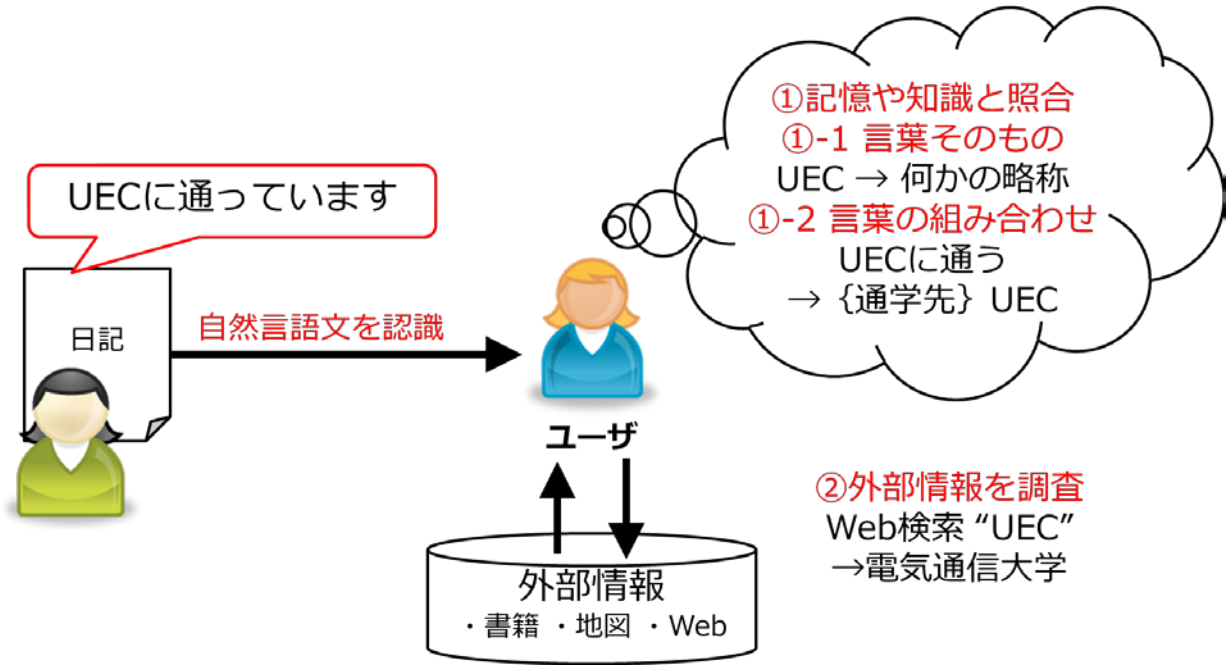


図 3.3　人間による外部情報を調査したプライバシー情報の認識（直接）

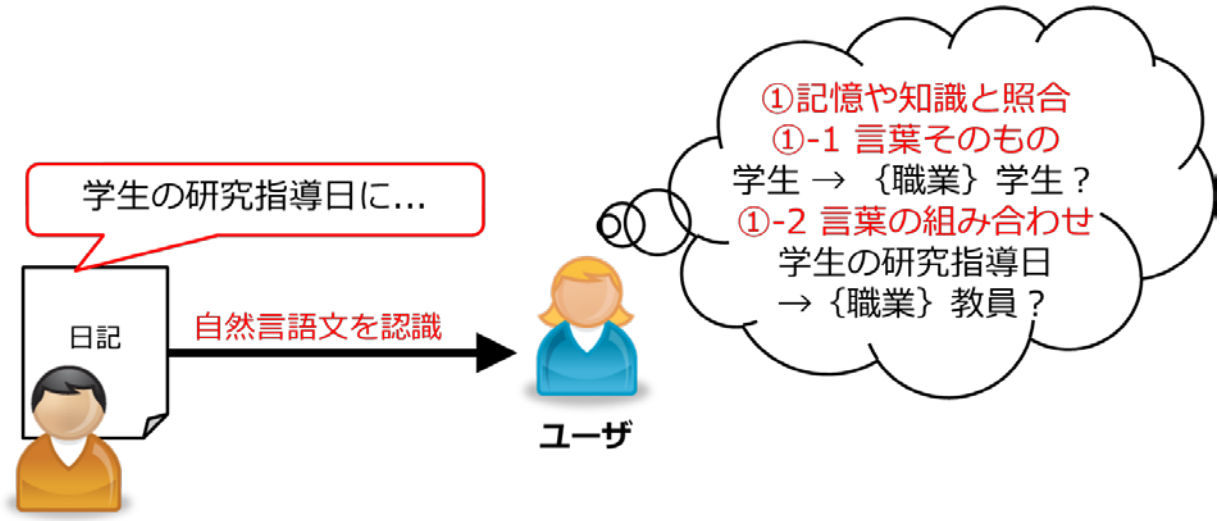


図 3.4　人間による言葉の組み合わせから連想，推論したプライバシー情報の認識

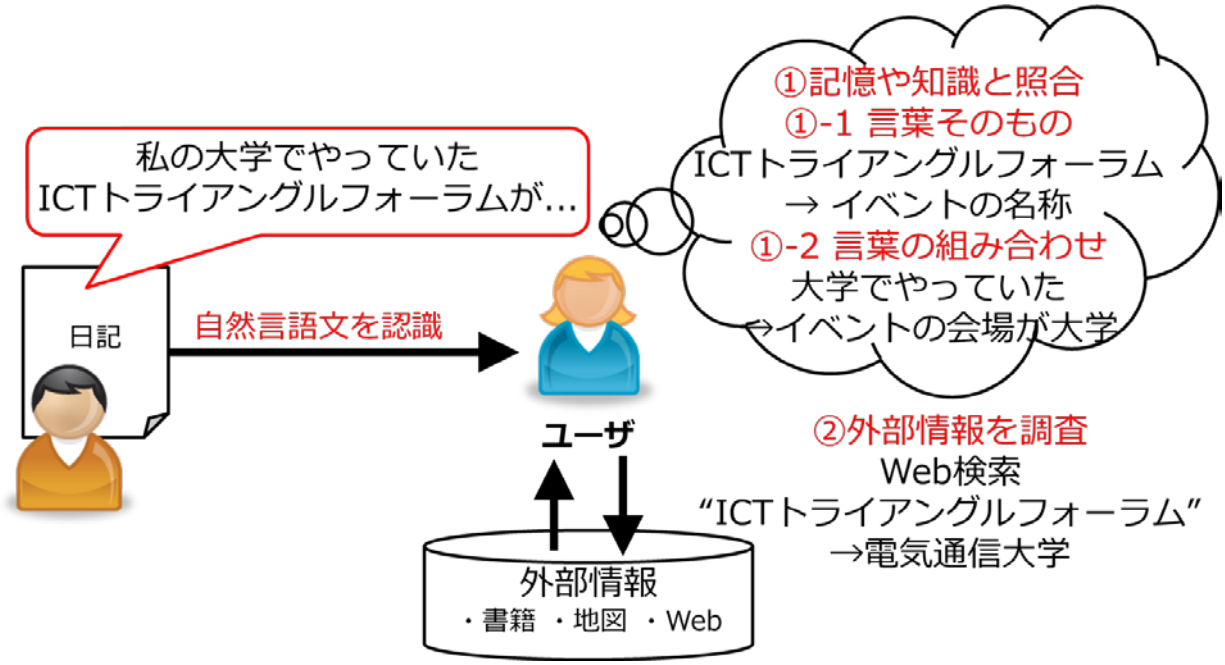


図 3.5　人間による外部情報を調査したプライバシー情報の認識（想起）

## 3.5　技術課題

前節のように実例文章を調査分析した結果，システムがプライバシー情報を検知するための技術課題を以下の通り明らかにした．

**課題 1**　プライバシー情報をユーザが無理なく定義可能とすること

人間は記憶や知識を参照して文章にプライバシー情報が含まれるか判断する．そのため，システムも検知すべきプライバシー情報を把握する必要がある．しかしプライバシー情報は，3.3 節で述べたように，個人によって範囲も内容も異なるため，個々のユーザがそれぞれ定義する必要がある．しかも，プライバシー情報は曖昧であるため，その厳密な定義は難しく，ユーザへの負担となる可能性がある．そこで，ユーザに負担をかけないでプライバシー情報の定義を可能にする方法が必要である．

**課題 2**　多様な表現形態によるプライバシー情報をシステムが高精度で検知可能とすること

3.4.4 節で述べたように，プライバシー情報は，名称などの直接的な表現だけでなく，単語の組み合わせや略称表現による表現を有する．「ICT トライアングルフォーラム」から電気通信大学が明らかになったように，外部知識と組み合わせるとプライバシー情報に到達するような表現もある．このような多様な表現形態によるプライバシー情報を高精度で検知可能とする方法が必要である．

**課題 3**　ソーシャルメディアに適した自然言語処理に対応すること

ソーシャルメディアに投稿された文章は，コミュニケーションを前提としているため様々な新語や略語，特定のコミュニティに所属する人の間のみで通じる語句などが含まれており，これらの語句は辞書には掲載されていない．そこで，辞書に掲載されない語句からもプライバシー情報を検知できるようにする必要がある．

## 3.6　システム設計と実装

### 3.6.1　基本方針

前節で挙げた技術課題に対して，以下の方針を立てた．

**(方針 1)**　NG ワードによるプライバシー情報の定義

プライバシー情報はユーザ毎に異なるので，ユーザに定義してもらう必要がある．しかし，ユーザに複雑なルールを書かせることは現実的でない．そこで，課題 1（プライバシー情報をユーザが無理なく定義可能とすること）を解決する方針として，知られたくないプラ

イバシー情報ごとに原則として 1 つの標準的な単語をユーザに定義してもらうことにする．ユーザに定義してもらうプライバシー情報を NG ワードと呼ぶ．プライバシー情報の多様な表現と NG ワードとのつながりをシステムが推定することでプライバシー情報を検知する．

例えば，ユーザが電気通信大学に勤務していることを開示したくない場合には，「電気通信大学」を NG ワードとして定義し，「UEC」や「西 6 号館」などの表現と「電気通信大学」との関連は，システムが推定する．これにより，ユーザにかける手間を抑える．なお，ユーザが改めて NG ワードを記載しなくても，すでに記載済みのソーシャルメディアのプロフィール情報から転用できる場合もある．

**(方針 2)** Web 検索による多様な表現への対応

課題 2，3 に対して，外部知識として Web 検索を利用する．これにより NG ワードの略称や別表現，NG ワードを間接的に表す言葉や，単語の組み合わせによって NG ワードを想起するものを検知する．例えば，“UEC”で Web 検索すると検索結果に電気通信大学のサイトが挙がり，UEC が電気通信大学の略称であることが明らかになる．また．“リサージュ”は図形の名称として知られているため，NG ワード「電気通信大学」と関係ない表現であるように思えるが，“リサージュ”と“1 階”の組合せで Web 検索すると検索結果から電気通信大学の建物の名前であることが明らかになる．そこで，文章中の単語および単語の組み合わせをキーワードとして Web 検索を実行し，検索結果の中に NG ワードがどれだけ含まれるかによってプライバシー情報を検知する．これにより，ユーザは多種多様な表現を考慮してプライバシー情報を定義する必要がなく，標準的な NG ワードのみ定義すれば良い．さらに，常に更新される Web を利用することで新語や略語などに対応できる．

3.5 節で明らかにした，システム化に向けた技術課題と方針を図 3.5 に，課題と方針との対応関係を図 3.6 にそれぞれ示す．

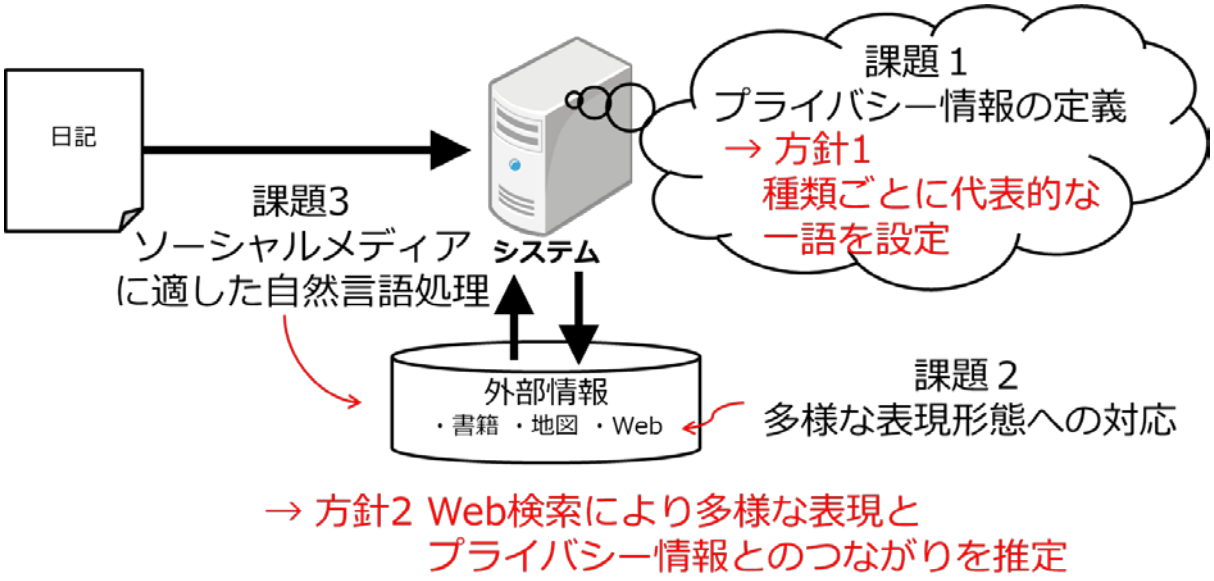


図 3.5　システム化に向けた技術課題と方針

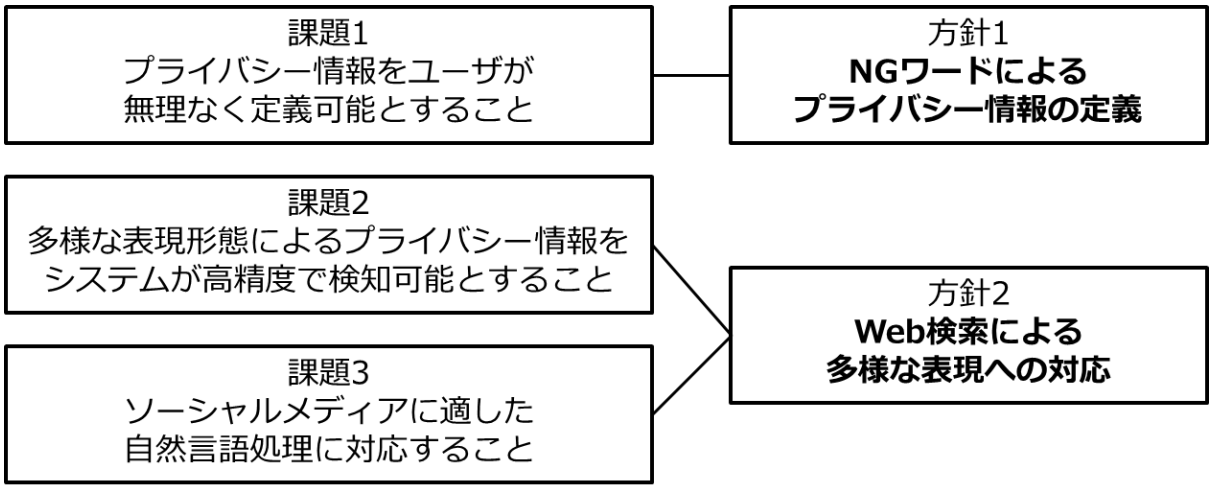


図 3.6　課題と方針との対応関係

### 3.6.2　アルゴリズム

前節の方針の通り，プライバシー情報は NG ワードとして定義されている．検知処理は，まず，自然言語処理によって投稿文章から単語を抽出する．その後，定義された NG ワードそのものを検知する処理（直接検知）と，略称や別表現，間接的に表す言葉，単語の組み合わせによる NG ワードの想起を検知する処理（想起検知）とを行う．検知システムの概要を図 3.7 に示す．

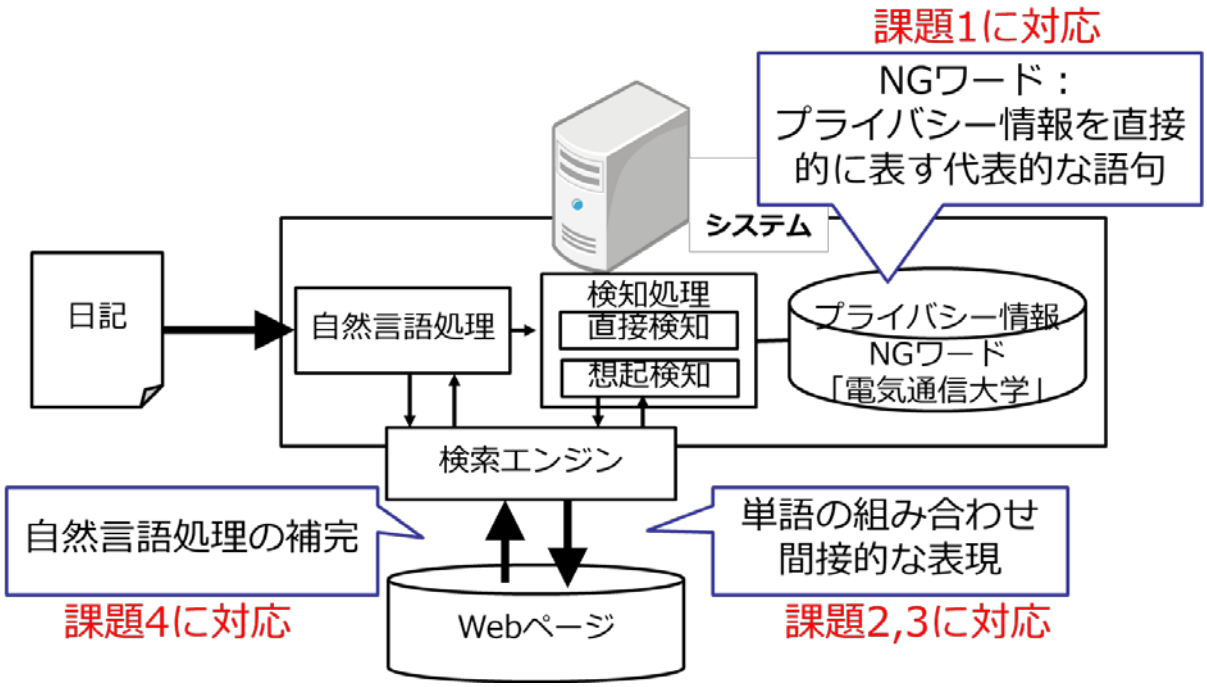


図 3.7　検知システムの概要

(1)　直接検知

投稿文章から抽出した単語と，NG ワードとの文字列マッチングを行い検知する．図 3.8 に直接検知の概要を示す．

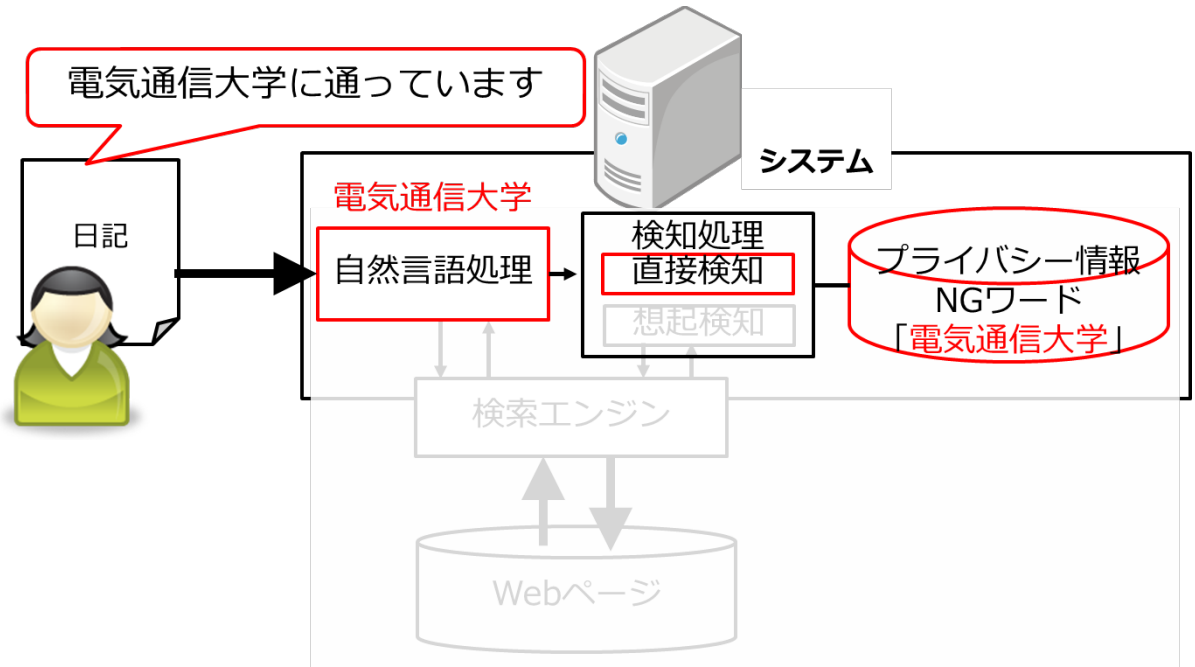


図 3.8　直接検知の概要

（2）　想起検知

3.4.4 節で述べたように，投稿文章中の単語および単語の組み合わせによる NG ワードの想起を検知する．その方法として投稿文章中の単語の最大 m 語までの組み合わせをクエリとして Web 検索を行い，検索結果における NG ワードの出現頻度によって検知する．

図 3.9 に想起検知の概要を示す．

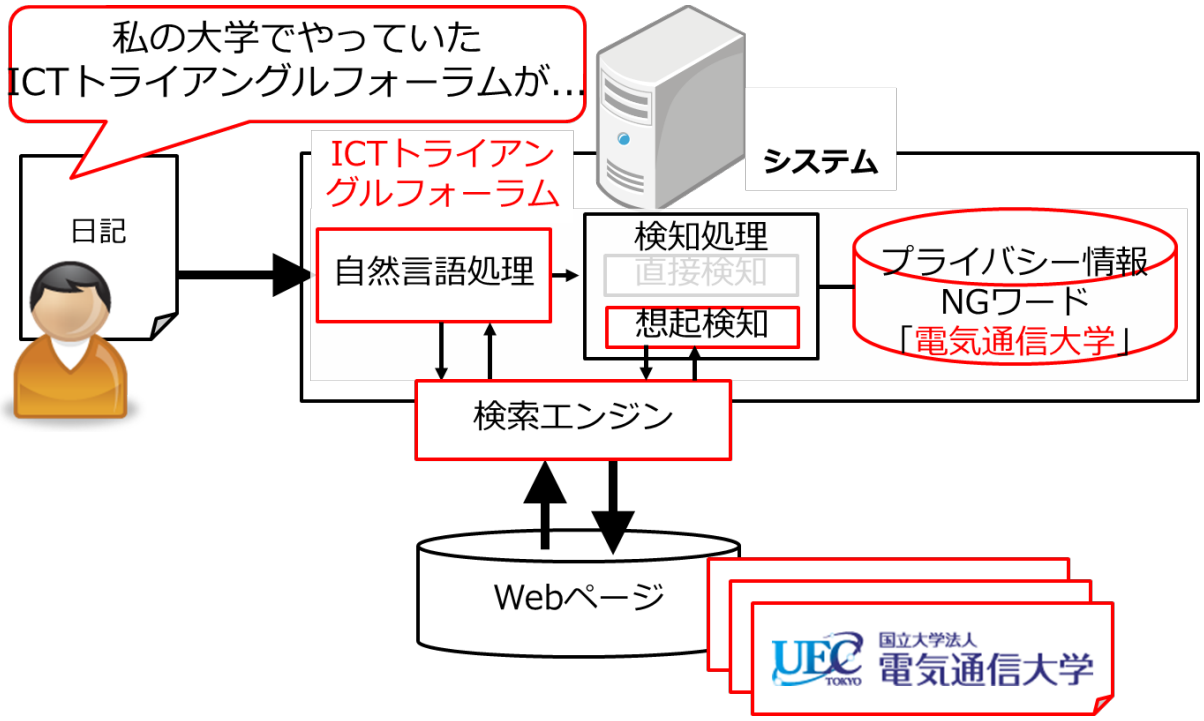


図 3.9　想起検知の概要

### 3.6.3　想起検知アルゴリズム

想起検知の具体的なアルゴリズムを述べる．形態素解析器や構文解析器を用いることで文章に含まれる名詞および複数の名詞からなる複合語を取り出し，これらの名詞および複合語をキーワードとする．

キーワードからなる最大$m$ 語までの組み合わせを生成し，これらをクエリ$q$ とする．このとき，キーワードの総和を$n$ とすると，クエリの総数$S$ は$\sum_{r=1}^{m} nCr$となる．そして，各クエリ$q_i(1 \leq i \leq S)$の想起度$score(q_i)$を，以下の手順で算出する．

① クエリ$q_i$によってWeb検索を実行し，検索結果の上位$k$ 件のページタイトルおよびサマリを取得する．

② 検索結果の$j$番目$(1 \leq j \leq k)$のタイトルおよびサマリにNGワードが存在するか探索し，NGワードの出現情報$f_{ij}$を式3.1のように返す．

$$f_{ij} = \begin{cases} 1 & (\text{ページタイトルまたはサマリにNGワードが存在する}) \\ 0 & (\text{ページタイトルまたはサマリにNGワードが存在しない}) \end{cases} \quad \cdots \text{ 式3.1}$$

③ 検索結果の順位$j$ によって，式3.2のように重み$w_{ij}$を定義する．

$$w_{ij} = k - j + 1 \quad \cdots \text{ 式3.2}$$

④ NGワードの出現情報$f_{ij}$と重み$w_{ij}$により，式3.3のように想起度$score(q_i)$を算出する．なお，想起度を$(0 \leq score(q_i) \leq 1)$の範囲にするために$\sum_{j=1}^{k} w_{ij}$で割り正規化を行う．

$$score(q_i) = \frac{\sum_{j=1}^{k} w_{ij} f_{ij}}{\sum_{j=1}^{k} w_{ij}} \quad \cdots \text{ 式3.3}$$

$(1 \leq i \leq S)$における$score(q_i)$の最大値を選定し，これを当該投稿文章の当該 NG ワードを示唆する想起度とする．ここでクエリ $q$ の単語数を $m$ =3，検索結果件数を $k$ =24 とした．これは，ユーザが Web 検索を行う際，入力する検索キーワードの平均単語数が 2.21 語であり，検索結果上位 23.5 件を閲覧することが知られているためである[35][36]．文章に含まれるキーワードおよびクエリの例を図 3.10 に示す．

④で選定した想起度が 0 でない場合に想起検知する．本アルゴリズムによると，検索結果上位 24 件のうち，1 件でもページタイトルおよびサマリに NG ワードが含まれていると想起検知する．これは，プライバシー情報の漏えいを防ぐという意味で適切である．

日記やつぶやきなど1文単位で形態素解析し名詞，名詞複合語を抽出

私の大学でやっていたICTトライアングルフォーラムが…

キーワードを最大3語を組み合わせた場合のクエリの例

| クエリ |
| --- |
| 私 |
| 大学 |
| ICTトライアングルフォーラム |
| 私　大学 |
| 大学　ICTトライアングルフォーラム |
| 私　ICTトライアングルフォーラム |
| 私　大学　ICTトライアングルフォーラム |

図 3.10　文章に含まれるキーワードおよびクエリの例

### 3.6.4　実装

文章からキーワードを抽出するため，形態素解析器として Mecab[37]，Yahoo! Japan が提供する日本語形態素解析 API[38]の 2 つを使用する．Mecab の解析用辞書に IPA が提供する IPAdic[39]を使用した．また，「電気通信大学」を「電気」「通信」「大学」と分割せず，一つの複合語として認識するために，専門用語ツール termExtract[40]を使用した．さらに，Wikipedia のカテゴリ名[41]とはてなキーワード[42]を名詞として登録し，新語に対応した．Wikipedia のカテゴリ名を用いることで，複合語の認識も行っている．Web 検索エンジンには，Yahoo! Japan が提供するウェブ検索 API[43]を使用した．

以上述べたアルゴリズムおよびツールを用いて，スクリプト言語 Python と Ruby で実装した．モジュール図を図 3.11 に示す．

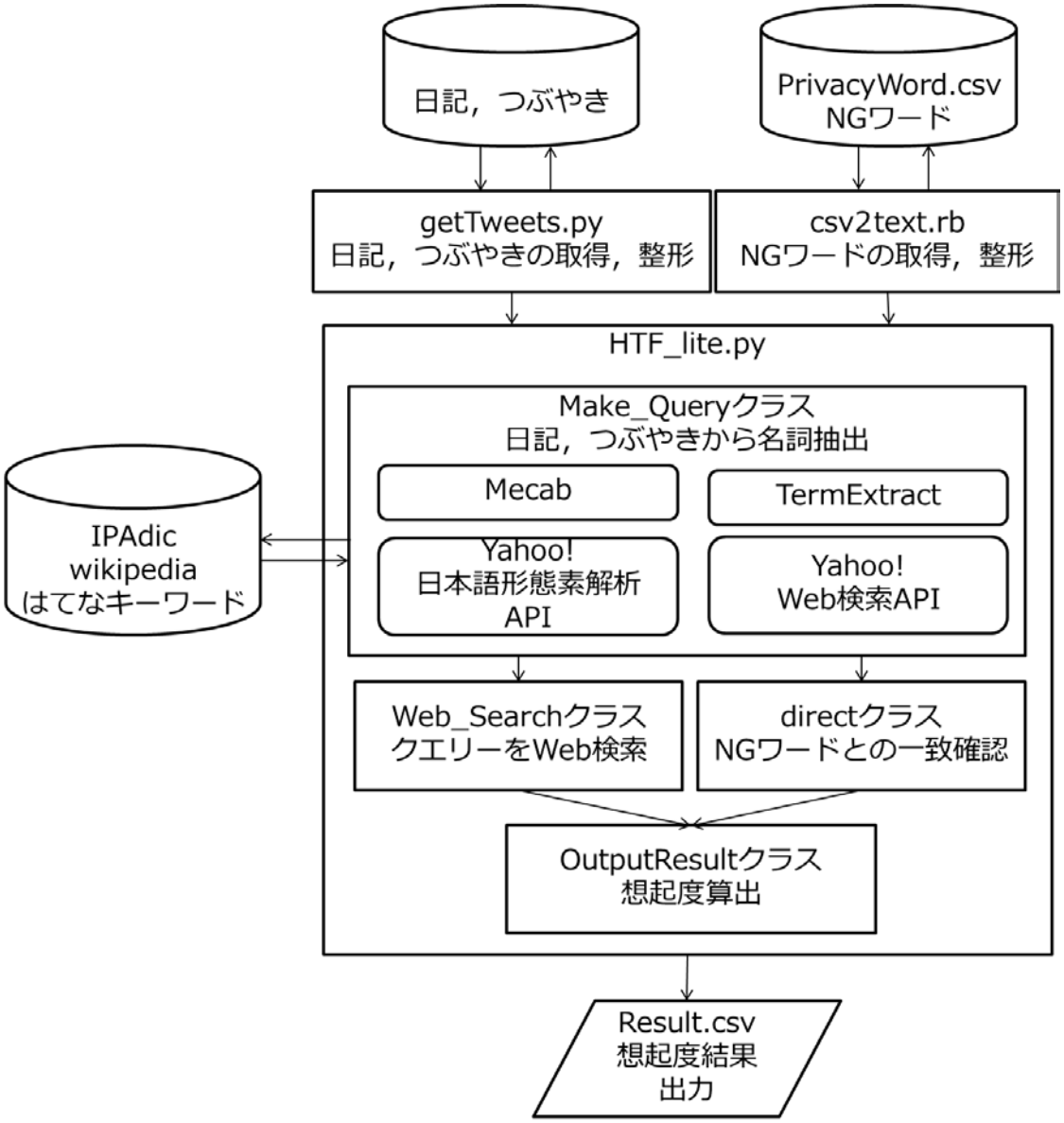
日記，つぶやき
PrivacyWord.csv
NGワード
getTweets.py
日記，つぶやきの取得，整形
csv2text.rb
NGワードの取得，整形
HTF_lite.py
Make_Queryクラス
日記，つぶやきから名詞抽出
Mecab
TermExtract
IPAdic
wikipedia
はてなキーワード
Yahoo!
日本語形態素解析
API
Yahoo!
Web検索API
Web_Searchクラス
クエリーをWeb検索
directクラス
NGワードとの一致確認
OutputResultクラス
想起度算出
Result.csv
想起度結果
出力

図 3.11　モジュール図

## 3.7　評価

### 3.7.1　評価方法

本評価では，人間が文章を読み，個人情報を含むか判断し，人間の判断を真実とみなすことにした．検知精度を調べるために，本研究では，再現率（Recall），適合率（Precision）F値（F-measure），真陽性率（True Positive Rate：TPR）および真陰性率 （True Negative Rate：TNR）を指標として用いる．

表3.3　混同行列

| | | システム検知 | |
|---|---|---|---|
| | | Positive | Negative |
| 人間が個人情報を含むと判断 | Positive | TP (True Positive) | FN (False Negative) |
| | Negative | FP (False Positive) | TN (True Negative) |

表3.3に示した混同行列より，再現率および適合率は以下の式の通り求められる．

$$再現率 = \frac{TP}{TP + FN} \quad \cdots 式3.4$$

$$適合率 = \frac{TP}{TP + FP} \quad \cdots 式3.5$$

再現率が高く適合率が低い状態は，人間が個人情報の漏洩を含むと判断する文章を，提案手法が高い確率で検知できることを意味する．しかし，人間が個人情報の漏洩を含まないと判断する文章を誤って検知する率も高い状態である．それに対して，適合率が高く再現率が低い状態は，提案手法が個人情報の漏洩を含むと判断した文章は，高い確率で人間が個人情報の漏洩を含むと判断した文章と一致することを意味する．しかし，その一方で見落としも多くなる状態である．

一般的な検知システムでは検知精度の評価にF値を用いることが多い．F値は以下の式3.6によって求められ，再現率と適合率の双方の値が高いほど高くなる．

$$\text{F値} = \frac{2}{\frac{1}{再現率} + \frac{1}{適合率}} \quad \cdots \text{式3.6}$$

しかし，Positiveに比べてNegativeが非常に多い場合には，システムは稀にしか警報を発しない．その場合は，適合率が低く警報の誤りが多くても，ユーザにとっての煩わしさは大きくならない．そのため，適合率を犠牲にしても再現率を高めることが望ましい．この観点の精度は真陽性率TPRと真陰性率TNRで評価される．TPRは適合率に等しく，TNRは，以下の式3.7で表される．

$$\text{TNR} = \frac{TN}{FP + TN} \quad \cdots \text{式3.7}$$

Positiveに比べてNegativeが多い場合には，FPが大きくてもTNRは高くなる．TPRとTNRの両者が高いほど検知精度は高い．

### 3.7.2　学生サンプルに関する評価結果

学生ユーザ 11 名の Twitter のつぶやき各 1000 件（合計 11,000 件）を対象に評価を行った．ユーザは電気通信大学情報理工学部総合情報学科 9 名，情報・通信工学科 1 名，知能機械工学科 1 名の 3 年（2012 年当時）男子学生である．NG ワードとして「電気通信大学」を設定し，人間およびシステムでプライバシー情報を含むか判断した．評価結果を表 3.4 に示す．

表 3.4　学生サンプルに関する検知精度

| ユーザID | TP | FP | TN | FN | 再現率 | 適合率 | F値 | TNR |
|---|---|---|---|---|---|---|---|---|
| 1 | 15 | 16 | 966 | 3 | 0.83 | 0.48 | 0.61 | 0.98 |
| 2 | 3 | 12 | 983 | 2 | 0.60 | 0.20 | 0.30 | 0.99 |
| 3 | 50 | 67 | 878 | 5 | 0.91 | 0.43 | 0.58 | 0.93 |
| 4 | 6 | 16 | 977 | 1 | 0.86 | 0.27 | 0.41 | 0.98 |
| 5 | 12 | 5 | 976 | 7 | 0.63 | 0.71 | 0.67 | 0.99 |
| 6 | 3 | 30 | 967 | 0 | 1.00 | 0.09 | 0.17 | 0.97 |
| 7 | 12 | 29 | 955 | 4 | 0.75 | 0.29 | 0.42 | 0.97 |
| 8 | 3 | 9 | 988 | 0 | 1.00 | 0.25 | 0.40 | 0.99 |
| 9 | 3 | 9 | 988 | 0 | 1.00 | 0.25 | 0.40 | 0.99 |
| 10 | 6 | 23 | 971 | 0 | 1.00 | 0.21 | 0.34 | 0.98 |
| 11 | 2 | 11 | 987 | 0 | 1.00 | 0.15 | 0.27 | 0.99 |
| 合計 | 112 | 227 | 10636 | 22 | | | | |
| 平均 | 10.5 | 20.6 | 967 | 2 | 0.87 | 0.30 | 0.42 | 0.98 |
| 標準偏差 | 13.9 | 17.4 | 31.3 | 2.5 | 0.15 | 0.17 | 0.15 | 0.018 |

再現率は，最も低いユーザ 2 において 60%，11 人中 8 人で 80%以上となった．適合率はユーザ 6，ユーザ 11 が低く，0.09，0.15 であった．F 値としても，ユーザ 6 および 11 が低く，それぞれ 0.17，0.27 であった．これは，適合率の低さに起因していると考えられる．表 3.5，表 3.6 にユーザ 6，ユーザ 11 においてシステムのみが検知した FP の全つぶやきとその想起度を示す．

表 3.5　ユーザ 6 の全 FP と想起度

| | つぶやき | 想起度 |
|---|---|---|
| 1 | 新年と同時に情報が押し寄せてきたので自分はちょっち混乱気味ですwww個人的には？はいつごろ公開なのかが気になるところです。一応2013年となると卒研とかが絡んできますので。。。。。。 | 0.10 |
| 2 | 神田朱未、田村ゆかり、堀江由衣、植田佳奈、白石涼子、能登麻美子、水樹奈々、喜多村英梨、井上麻里奈、花澤香菜、石原夏織、井口裕香が声優をつとめ、河森正治が監督で、岡田麿里が脚本で、大月Pがエグゼクティブプロデューサーをつとめるアニメが来季あるらしい。ホントかな？？ | 0.08 |
| 3 | バーチャルアイドル学／講師：くさぱん／開講時間：水6／難易度：鬼畜／成績評価：受講者の魅力で判断する／受講者の声：「死ぬほど眠い授業」アイマスの魅力に始まり、錦織さんのキャラの素晴らしさ、制作会社トリガーの昨今など盛りだくさんの内容 | 0.07 |
| 4 | ブログ更新ですー。「ACEに行ってNERV特別仕様スマートフォンを見てきた。」 やや淡白な文体になっているけど、これはちょっち眠いから。大学の新歓期は忙しいー。 | 0.07 |
| 5 | 現役大学生のエヴァオタな日常：此迄ノ荒筋。… 公開月はじめにつき、PLAY BACK。そして再始動へ。 | 0.07 |
| 6 | でも、エヴァには庵野秀明って人間が良く作品に表れている。アニメという映像作品に、あれほどまでに一人の人間の苦悩や考えが反映されているものは他に無い。エヴァをドキュメンタリーと呼び、登場人物は全て私自身だと言った庵野さんの言葉の裏にはそういった事情がある。 | 0.06 |
| 7 | 大学関係、片付きました！以後はブログを通常更新できると思います！ | 0.06 |
| 8 | ルパン、めだか、アクセル、つり球、氷菓、AKB、アポロン、モーパイ、FateZero。 | 0.06 |
| 9 | 今日はカヲルフェアに行ってきます！大学関係がちょっちだけ落ち着いたので、今日はブログ更新できそうですよー！ | 0.06 |
| 10 | ちなみに私はニコ生を観るために大学のPCルームへ急行中です。14時40分から講義なんですよねー。。。。。 | 0.06 |
| 11 | エヴァストアのバイト募集に応募しようと思ってだだけど、週3日4時間が確保できずに断念。うちの大学きっついなー。 | 0.06 |
| 12 | 「ダサい」www樋口さんさすがハッキリ言うなぁww王立のリイクニも「ブス」っていってたしwlive at | 0.06 |
| 13 | 庵野、鶴巻、摩砂雪、平松、鈴木、松原、本田、吉成、すしお、今石、貞本、錦織、林、樋口、黄瀬、前田、山賀、榎戸、鷺巣、増尾、山下、okama、大月、芳垣（敬称略）。ここらへんの名前があると大騒ぎする。 | 0.05 |
| 14 | すべてはエヴァのために。その次にバイト、勉強と続く。大学二年生くらい好きな事すればいいじゃんという考えで生活してきた。逆に言えば、嫌なことからは徹底的に逃げてきた。社会人になったらこうは好き勝手に行かないだろう。来年はもっと遊びまくろうと思っている。 一年はこうでした | 0.05 |
| 15 | 動画こそ、最終的なアニメのクオリティを決定する大事な役職だと思うのですが。。。。。やはり、あなた様のおっしゃるように現場との意識のズレがあるのでしょうね。 | 0.05 |

| | | |
|---|---|---|
| 16 | バイトが長引いた影響で11時30分ギリギリになりそうだなぁ。抽選会開始に間に合えっ！ | 0.05 |
| 17 | 今理系の人工知能系目指す人たちはみんな電王戦観てるんだろうか。 | 0.05 |
| 18 | スタジオカラー、生体認証キーなんて付いてるんだ。秘密基地化してるとは聞いていたがそれ程とは。 | 0.05 |
| 19 | うーん。大学の方が忙しくてなかなかブログ更新できないな。。。3月のカヲルフェアのレポートは絶対書きたいところ。 | 0.04 |
| 20 | ケチというか、ここは？と思った部分はというと、よくわからない伏線とその回収があったってことと、ややストーリーが断片的というか、個別のメッセージがそれぞれ微妙にだけ繋がった状態で(うーむ。うまく表現できん)投げかけられてる気がした。まぁ、素人意見ですけど。 | 0.04 |
| 21 | いや、ミラノ1に座席予約なんて高度なシステム無いから、先行もオンラインチケもないはず。あそこはやはり徹夜組が出るだろうね。初回は並ばなきゃ。問題は、初回観た後にさらに観る分のチケットを前もって買っておかなきゃいけないってこと。しかも、先行販売は一週間分。 | 0.03 |
| 22 | あのシンプルな特報で、3DCGのレベルの高さを示し、様々な推測をさせるメッセージ性を盛り込んでくるあたりは、さすがとしか言いようがない。 | 0.03 |
| 23 | そうですね。ここまで情報統制してたのに、今更日テレでそんなに重要な情報は開示しないだろうとは思ってましたw個人的には、あのピアノのCGにもものすごい驚かされました。Qでのデジ部の活躍を感じさせてくれたと思っています。 | 0.02 |
| 24 | ノイタミナ枠の二作の最終話を視聴。BRSは少数話ながらよくまとまっててとても良かったなぁ。CGを金田的に見せるということを、今石さんとサンジゲンが上手くやってのけてたあたり、映像面も素晴らしかったと思います。アニメスタイルで取り上げられてたように、まさしく先鋭的な作品でした。 | 0.02 |
| 25 | 新宿駅北側、デモの音響いてる。 | 0.02 |
| 26 | 棋士達だけでなく、開発者たちも人間の勝利を願っているんだからね。 | 0.02 |
| 27 | :「動くプラットホーム」、列車と併走　究極の高速鉄道システム さんから | 0.02 |
| 28 | ぴあ観客満足度調査98.9パーセント好評につき、劇場側の希望で半年以上のロングラン上映 第33回日本アカデミー賞 優秀アニメーション作品賞 第9回東京アニメアワード個人部門・音楽賞『鷺巣詩郎』第15回リヨン・アジア映画祭 アニメ部門第1位 | 0.01 |
| 29 | 「先の地震によって日本の迎撃システムは大きなダメージを受け、現在までの普及率は26%。実践における稼働率はゼロといっていいわ。したがって今回は、上陸直前の目標を水際で一気に叩く！非リアは交互に目標に対し波状攻撃。近接戦闘でいくわ。クリスマスを、なんとか食い止めるのよ。」「了解！」 | 0.01 |
| 30 | はい。フォロワーの方、TL荒らしちゃってすいません。フツーのエヴァオタに戻ります。 | 0.01 |

表 3.6　ユーザ 11 の全 FP と想起度

| | つぶやき | 想起度 |
|---|---|---|
| 31 | 下級戦士の俺でも上級科目がとれるってのがありがたい | 0.23 |
| 32 | あのカマぶっころして明日は休講だッ！ | 0.14 |
| 33 | 休講だな | 0.14 |
| 34 | 【数論アルゴ】よく考えたらこの講義単位とるのは楽かも【サニーサイドアップ】 | 0.13 |
| 35 | さーてついたぜ、電車のドアがボタン式なんて、エキサイティングだな | 0.12 |
| 36 | いまね、情報基盤で実験のレポート書いてる | 0.07 |
| 37 | 学内の女の子？たぷん全員抱いたわ | 0.07 |
| 38 | おい！確率論のテスト中止で、レポートだってよ！！！！！掲示出てるぞ！ | 0.06 |
| 39 | 調布駅で記念乗車券の徹夜組がいたわ | 0.06 |
| 40 | 履修希望者がいっぱいいるのは、自分の講義が面白いからだわウフフとか思ってるー？なめられてるんだよ！おめーの講義がぬるいってな！ | 0.05 |
| 41 | エンジニアリングデザイン | 0.03 |

これらの FP であったつぶやきは，学生であることが想起できるものが多いが，一般的な言葉が使われているため，人間には電気通信大学を特定するのは困難である．また，つぶやき 2，つぶやき 13 のように人名を列挙しているつぶやきでは，大学の関係者に何名か同姓がいることから電気通信大学を検知していると考えられる．また，つぶやき 25 の「デモ」は，文脈として「デモ活動」のことを指しており，電気通信大学とは全く関係ないと考えられるが，システムは「デモンストレーション」の略語として，デモ展示に関わるサイトから想起していると考えられる．

一方，本システムでは Positive に比べて Negative が非常に多いので，TPR と TNR を用いた評価も行った．前述のように，TPR は再現率と等価であり，TNR は式 3.7 で示したとおりである．

TPR は前述のように高い値である．TNR は，すべてのユーザについて 93%以上となった．したがって，Negative が多いという状況を考慮すると，提案方式の精度は高いと言える．

見方を変えると，システムの検知（ユーザへのアラート）は，最も適合率の低いユーザ 5 でも 1000 文中，33 文程度であり，適合率が低くとも煩わしくはない．さらにユーザインタフェースを工夫することでアラートの煩わしさを軽減することは可能である．例えば，文書ソフトのスペルチェックや文法チェックのように該当する文字の色を変える，下線を引くといったユーザインタフェースが考えられる．

### 3.7.3　False Positive の再分析

システムによってプライバシー漏えいを検知したが，人間は漏えいを判断しなかった False Positive（FP）のつぶやきを改めて精査し，本当に FP であるかを分析した．結果を表 3.7 に示す．

表 3.7　再チェックし漏えいの可能性が高いと判断した文章

| ユーザID | つぶやき | | 想起度 |
|---|---|---|---|
| 1 | 25 | IEDのメイヤー奪われてた | 0.05 |
| 2 | 26 | 実験のレポートの物理的提出場所は東3の6階だっけ？ | 0.29 |
| | 27 | OYM研かNSN研入れなかったら他大院かIS行きますはい。 | 0.17 |
| | 28 | @各位 明日の四限が暇なんでシミュレーション理工学第一に潜り込んでみようか悩んでいるのですが潜り可能な感じでしょうか？ | 0.08 |
| 4 | 29 | 明日はIEDだっけか | 0.05 |
| | 30 | よっぽど細かく線が並んでないとモアレらないでしょ(元群青副編集長) | 0.01 |
| 8 | 31 | 図書館、IED、docomo、・・・他に寄る所あったかな？ | 0.05 |
| 9 | 32 | ビジュアル情報処理の成績まだですか！？！？！？ | 0.09 |
| | 33 | IED果てしなく使いにくい | 0.05 |

改めてチェックしてみると，確かに漏えいの可能性が高い文章があった．それらを表 3.7 に示す．例えば，ユーザ 4 のつぶやき「よっぽど細かく線が並んでないとモアレらないでしょ(元群青副編集長)」（つぶやき 30）の「群青」という言葉は，色の名前や，曲の名前で使われる一般的な名前であり，一見すると電気通信大学とは関係ないように思える．しかし，「編集長」という言葉と組み合わさることで，電気通信大学の学友会が発行する冊子である「群青」に関するサイトが検索結果に出現し，電気通信大学を想起する結果になった．

ユーザ 2 のつぶやき「実験のレポートの物理的提出場所は東 3 の 6 階だっけ？」（つぶやき 26）は，東 3（号館）という名称の建物は世の中に多数存在すると考えられ，特に電気通信大学を想い起させる文章には見えない．しかし，Web 検索をすると，実験のレポート提出場所として東 3 号館の 6 階を指定する電気通信大学のサイトがヒットし，電気通信大学を想起する結果になった．

ユーザ 9 のつぶやき「ビジュアル情報処理の成績まだですか！？！？！？」（つぶやき 32）は，ビジュアル情報処理が講義，単位の名前であることが想像できるものの，一見す

ると電気通信大学と関係あるかは分からない．しかし，Web 検索を実施すると，「ビジュアル情報処理」という講義を開講している大学は，電気通信大学，茨城大学，福井工業大学など数校に限られることが分かる．

これらのつぶやきは，一見すると安全に見えるため，システムが過剰にプライバシー情報の漏えいを検知しているようにも見える．しかし，プライバシー情報として NG ワードを用意し Web 検索を用いることで，本人さえも気づかなかったプライバシー情報の漏えいが明らかになったと言える．また，ユーザの素性を突き詰めて調査したい人間がいれば，上述のような Web 検索を行うことで「電気通信大学」を想起できてしまうため，一概に過剰検知とは言えない．一見すると人間では漏えいに気づかない言葉であっても，システムが NG ワードと Web 検索を用いることにより，プライバシー情報漏えいの可能性を検知することができるため，ユーザがプライバシー漏えいを回避可能となる．このように，本システムはプライバシーを保護するという点で有意義であると言える．

以上の分析を考慮して，検知精度を修正したものを表 3.8 に示す．

表 3.8　FP の再分析を踏まえた検知精度

| ユーザID | TP´ | FP´ | TN | FN | 再現率´ | 適合率´ | F値´ |
|---|---|---|---|---|---|---|---|
| 1 | 15 | 16 | 966 | 3 | 0.83 | 0.48 | 0.61 |
| 2 | 4 | 11 | 983 | 2 | 0.67 | 0.27 | 0.38 |
| 3 | 53 | 64 | 878 | 5 | 0.91 | 0.45 | 0.61 |
| 4 | 6 | 16 | 977 | 1 | 0.86 | 0.27 | 0.41 |
| 5 | 14 | 3 | 976 | 7 | 0.67 | 0.82 | 0.74 |
| 6 | 3 | 30 | 967 | 0 | 1 | 0.09 | 0.17 |
| 7 | 12 | 29 | 955 | 4 | 0.75 | 0.29 | 0.42 |
| 8 | 3 | 9 | 988 | 0 | 1 | 0.25 | 0.40 |
| 9 | 4 | 8 | 988 | 0 | 1 | 0.33 | 0.50 |
| 10 | 8 | 21 | 971 | 0 | 1 | 0.28 | 0.43 |
| 11 | 2 | 11 | 987 | 0 | 1 | 0.15 | 0.27 |
| 合計 | 124 | 218 | 10636 | 22 | | | |
| 平均 | **11.3** | **19.8** | **967** | **2** | **0.88** | **0.34** | **0.45** |
| 標準偏差 | 14.6 | 16.9 | 31.3 | 2.45 | 0.14 | 0.2 | 0.16 |

このように，改めて検知結果をみると，TP が 10%程度増加していることが分かる．これらは，最初は人間が気付かなかったがシステムの検知を受けて読み直した結果，漏えいに気づいたケースである．このことから，提案システムにより人間の知識や能力を超えた検知が可能となり，ユーザの注意を補うことができるようになったと言える．

### 3.7.4　False Negative の分析

人間はプライバシー情報の漏えいを検知したがシステムが検知せず False Negative（FN）となった全てのつぶやきを表 3.9 に示す．これらの原因は全て，形態素解析に失敗し，クエリの生成に必要となるキーワードを適切に抽出できなかったことであった．具体的には，専門用語解析において，複数の名詞の連続を一つの複合語としていることに起因していた．例えば，つぶやき 39 の「@電通女子 ごめんなさい」は「電通女子」がキーワードとして抽出されていれば Web 検索結果 9 位に電気通信大学に関わるサイトが出現し，想起検知が出来る．しかし実際には，「@電通女子」がキーワードとして抽出され，システムは検知することが出来なかった．また，つぶやき 50「ダバディさん 10J 科かなんか？？」も同様に，「J 科」がキーワードとして抽出されていれば，Web 検索結果 1 位に電気通信大学に関わるサイトが出現し想起検知が出来る．しかし実際には，「ダバディさん 10J 科」がキーワードとして抽出されたため，このつぶやきもまたシステムが検知しない結果となった．

これらは，つぶやきに特有の記号の使い方や表現によって，形態素解析が適切に行えなかったことが原因であると考える．「@電通女子」は，「@」以下にユーザ名を指定すると，そのユーザ宛のつぶやきになる，という Twitter のルールから転じて，『電通女子のみなさんへ』という意味として使われたと考えられる．また，「ダバディさん 10J 科かなんか？？」は，口語表現と同様に「は」や「を」といった助詞が省略されているつぶやきである．このように，記号や助詞の省略により名詞が複数連続したため，形態素解析に失敗し，キーワードを適切に抽出できなかったと考えられる．

これらに対しては，複数の名詞が連続していた場合，すべての組み合わせで複合語を作りクエリとする方法が改善策として挙げられる．これにより，正しい区切りでキーワードを生成できる可能性が高まり，システムの精度向上が期待できる．しかしながら，必然的にキーワード数が増加し，それに伴い Web 検索回数も増加すると考えられる．他には，例えば過去に 1 度でも想起を検知したことのある単語はシステムが学習し，次回以降の形態素解析の際に活用するなど，形態素解析の精度を向上する工夫が考えられる．

表 3.9　FN となった全てのつぶやき

| ユーザID | | つぶやき |
|---|---|---|
| 0 | 34 | 聞いている人の電通の割合が知りたいw |
| 0 | 35 | 電通生？ |
| 0 | 36 | 電通 |
| 1 | 37 | J科勢こわい |
| 1 | 38 | J科「W9-135は我らがいただいた」 |
| 2 | 39 | @電通女子 ごめんなさい |
| 2 | 40 | S科は全部ブラックだろ目を覚ませ |
| 2 | 41 | 西9 5階の先生方はC言語の系統がお嫌い |
| 2 | 42 | J科10生特定冊子、当時はテンションがおかしかったので痛いこと書いてた記憶が |
| 2 | 43 | やっぱり10生J科クラスタで飲み会したい！ |
| 3 | 44 | 昨日会った電通生のアカウント見つからない |
| 4 | 45 | 平均的な電通男子、電通女子とは。 |
| 4 | 46 | いま西9一回のでかい教室 |
| 4 | 47 | せってくん、さっき非公式でリプ飛ばしてきた王国民、J科？？ |
| 4 | 48 | J科にようじょなんかいたら、誰かしらの餌食になってるだろ。ゆのすけとか、まゆしぃとか。 |
| 4 | 49 | ゆう君フォローした(J科カップルコンプリート！！) |
| 4 | 50 | ダバディさん10J科かなんか？？ |
| 4 | 51 | 彼氏彼女の関係ってどんなものなのか、気になります。(J科の目立たない方のカップルを観察しながら) |
| 6 | 52 | 【こんな孫はいやだ】電通生 |
| 6 | 53 | 東5-241の黒板はUserStream対応！！ |
| 6 | 54 | ポケカGB電通杯は無いのか |
| 6 | 55 | 電通は？ |

### 3.7.5　社会人サンプルに関する評価結果

3.4 節において分析した 50 代男性ユーザによる日記 149 件と日記に対するコメント 993 件の計 7047 文を対象に評価を行った．NG ワードとして通勤先大学名を設定し，人間およびシステムでプライバシー情報を含むか判断した．評価結果を表 3.10 に示す．

表 3.10　社会人サンプルに関する検知精度　NG ワード：通勤先大学名

| NGワード | TP | FP | TN | FN | 再現率 | 適合率 | F値 | TNR |
|---|---|---|---|---|---|---|---|---|
| (通勤先大学名) | 53 | 377 | 6617 | 0 | 1 | 0.12 | 0.22 | 0.95 |

また，NG ワードとして「教員」を設定し評価した．評価結果を表 3.11 に示す．

表 3.11　社会人サンプルに関する検知精度　NG ワード：「教員」

| NGワード | TP | FP | TN | FN | 再現率 | 適合率 | F値 | TNR |
|---|---|---|---|---|---|---|---|---|
| 教員 | 53 | 1652 | 5342 | 0 | 1 | 0.03 | 0.06 | 0.76 |

NG ワードとして「教員」を設定した場合の適合率が，通勤先大学名を設定した場合と比較し，FP が多く適合率が低い．また，NG ワード「教員」は，前節の学生サンプルによる評価と比較しても適合率，TNR が低い．また，再現率は，NG ワードが通勤先大学名，「教員」ともに 1 であり，学生サンプルと同様に高い値となった．

表 3.12 に NG ワード通勤先大学名のときに FP となった文例と想起度，表 3.13 に NG ワード教員のときに FP となった文例と想起度をそれぞれ 5 文抜粋して示す．

文 7 の「今日は大学の入学式が武道館で行われた」は，武道館が入学式の会場である大学は多数存在し，特に大学名の想起につながるとは考えられないが，Web 検索をすると勤務先大学のサイトが出現し，想起する結果になった．文 11 の「大学の 3 年生に対するグループスタディという科目があります」は，グループスタディが実は勤務先大学特有の科目名であるため，Web 検索すると当該大学のサイトが出現する結果となった．また，文 8 は，符号理論，情報理論など，学術分野の名称が列挙されているため，関連する分野を研究する大学として想起している．

これらの文は大学名だけでなく，教員も想起している．大学，入学式，科目，情報理論，業績といった言葉が起因している．他にも，文 4 の「私たちの大学は，従来から就職が良いのですが，今年は特に企業が熱心のようです」は，人間には大学関係者であることを推測することができるものの，教員であるかは分からない．しかし，大学，就職，企業など教員と共に出現するサイトが存在するため，システムは教員を想起すると判断している．大学名と比較し，教員は一般的な言葉であることから，より多様な言葉と共に Web 空間に存在していると考えられる．そのため，より多くの文で教員を想起検知する結果になったと考察できる．教員の場合 7047 文のうち 1705 文でアラートを発し，そのうち約 97%が誤検知となった．しかしながら，例えば文 3 や文 4 のように，日記文のみでは教員である

ことを特定することができない文章も，システムによるアラートを受けることで確かに教員を想起する可能性があることに気づくことができる．そのため，一概に過剰検知であるとはいえず，想起の可能性を指摘するという意味では有益な検知になると考える．

表 3.12　FP となった文例と想起度　NG ワード：通勤先大学名

| | 日記文 | 想起度 |
|---|---|---|
| 7 | 今日は大学の入学式が武道館で行われた。 | 0.09 |
| 8 | 符号理論、情報理論、暗号理論、情報セキュリティの分野で優れた業績を上げるとともに、非常に優秀な弟子たちを育て上げた方です。 | 0.08 |
| 11 | 大学の3年生に対するグループスタディという科目があります。 | 0.06 |
| 12 | 現在、研究室のサーバ上にある自分のホームページで、高校の昭和41年卒業の同窓会の広報担当をやっています。 | 0.04 |
| 10 | 卒業式が、今日、千鳥が淵の武道館で実施されました。 | 0.07 |

表 3.13　FP となった文例と想起度　NG ワード：教員

| | 日記文 | 想起度 |
|---|---|---|
| 7 | 今日は大学の入学式が武道館で行われた。 | 0.03 |
| 8 | 符号理論、情報理論、暗号理論、情報セキュリティの分野で優れた業績を上げるとともに、非常に優秀な弟子たちを育て上げた方です。 | 0.21 |
| 11 | 大学の3年生に対するグループスタディという科目があります。 | 0.06 |
| 4 | 私たちの大学は、従来から就職は良いのですが、今年は特に企業が熱心のようです。 | 0.04 |
| 3 | 実は大学の中にも、奉祝神田祭の垂れ幕が掲げられています。 | 0.03 |

### 3.7.6 ユーザインタフェースについて

前節までで述べたように，提案手法は FP が多い傾向があり，ユーザに不必要なアラートを発することになる．そのため，不必要なアラートがユーザに煩わしさを与えないように，ユーザインタフェースを工夫する必要がある．その一案として，文書ソフトの校正機能（スペルチェックや文法チェック）のようにテキストの該当箇所の色を変える，下線を引くことが考えられる．文書ソフトの校正機能，特に英文法のチェック機能は，以前は精度が低かった．しかし，ユーザインタフェースが煩わしくなかったので，多くの人に受け入れられ，長い間に技術が進歩した．提案手法も同様な戦略で実用化を目指すことが可能と考える．提案手法のユーザインタフェースのイメージを，図 3.12 に示す．

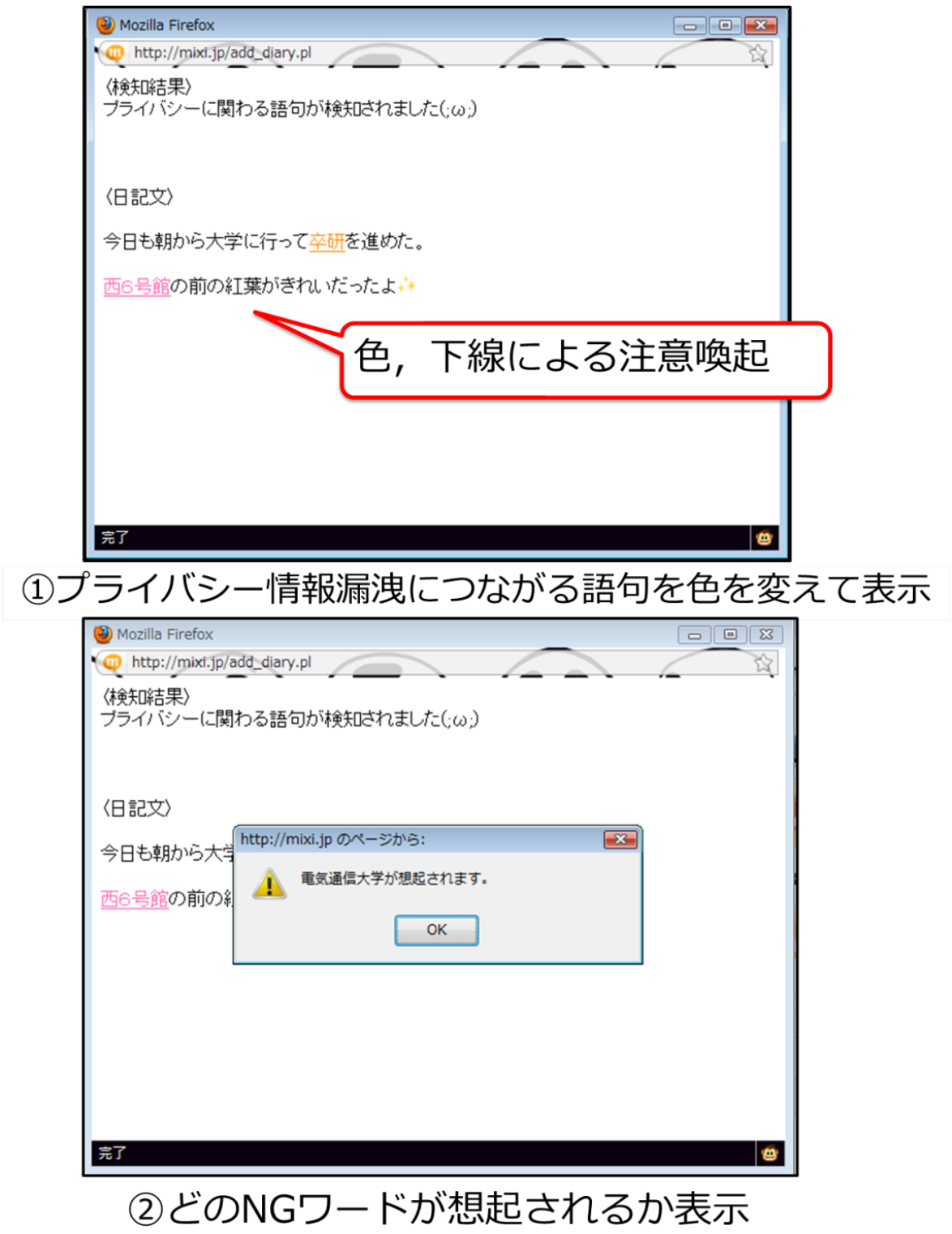


図 3.12 ユーザインタフェースイメージ図

## 3.8 関連研究

我々が本研究を最初に発表した 2007 年以降に発表された関連研究を挙げる．Lam らは台湾のソーシャルネットワーク Wreck を取り上げ，友人からの一言コメントと呼ばれる Wreck 固有の書き込み行をキーワードマッチにより解析することで 72%のユーザのファーストネームを，また 30%のユーザのフルネームを明らかにできると報告している[11]. Kótyuk らはハンガリーのソーシャルネットワーク「iwiw」を取り上げ，ユーザプロフィールの一部の属性（たとえば年齢）が非開示の場合に，開示された属性，友人関係，ユーザグループへの所属から非開示属性を推定する手法を提案し，たとえば年齢が公開されていなかったとしても 4.76 歳の誤差で年齢を推定できると報告している[44]. 同様に, 74.3%の確率で性別を，67.5%の確率で marital status（結婚しているかどうか）を推定できると報告している．Banks らは，ソーシャルネットワーク上でのユーザ間の交流状況に基づいて親密度を推定し，親密度に基づいて適切な公開範囲を設定する手法を提案している[45].

Mao らは，病気や飲酒などのセンシティブ情報が漏えいするつぶやきについて，漏えいの原因や経緯を調査し，またそのようなつぶやきの分類手法を提案している[46]．それによると，旅行や病気に関するつぶやきを 76%の精度で，また飲酒運転のつぶやきを 84%の精度で検出できると述べている．また検知の結果に基づいて，ユーザにプライバシー漏洩を警告するとしている．町田らは，公文書の公開基準を参考にして，プライバシー情報を内容と重要性によって分類する一方，開示された文章の中の単語からプライバシー情報を検知する手法を提案している[47]．さらに，交流頻度などに基づいて友人を分類し，適切な公開範囲を設定する手法を提案している[48]．Ngoc らは，ソーシャルネットワークへの投稿文からプライバシー情報の漏洩を検知し，言い換える処理の研究を進めている[49]. これらの研究は，本論文の 2 章で提案した自然言語情報の開示制御技術 DCNL および 3 章で提案したプライバシー情報の検知技術の後続研究と考えられる．

さらに，本研究から派生した研究として[50]が挙げられる．これは，SNS などに投稿した写真などの画像を対象とした開示制御である. SNS からのプライバシー情報の漏えいを防ぐという我々の考え方を引用した研究である．

## 3.9　まとめと今後の課題

本章では，自然言語の開示制御技術DCNLの実現に向けて，投稿した文章からユーザのプライバシー情報を検知する手法を検討した．SNSに投稿された投稿文の実例を分析して，どのようなプライバシー情報が漏洩するか調査し，人間によるプライバシー情報の検知モ

デルを考察した．その上で，技術課題として，（1）プライバシー情報は個人に依存し曖昧であるが，これをユーザが無理なく定義できること，（2）プライバシー情報は投稿文において多様な表現を有するが，これらを高精度で検知可能とすること，の2点を明らかにした．課題を解決するための基本方針として，NGワードと呼ばれる単語によってユーザがプライバシー情報を簡単に定義し，投稿文に現れるプライバシー情報の多様な表現とNGワードとの結び付きをシステムが推定することとした．この推定では，投稿文中の単語およびその組み合わせをキーワードとしてWeb検索を行い，検索結果にNGワードが含まれる位置と回数に基づいて，結びつきを定量化した．

提案したアルゴリズムを実装し，学生11名のTwitterのつぶやき各1000件（合計11,000件）および社会人1名のmixiの日記149件（7047文）をサンプルとして，通学・通勤先および職業情報の検知精度を評価した．人間によるプライバシー情報漏えいの検知とシステムによる検知結果を基に混同行列を作成し，再現率（漏洩した件数のうちシステムが検知した率)，適合率（システムが検知した件数のうち真に漏洩している率)，F値（再現率と適合率の調和平均)，真陰性率(システムが検知しなかった件数のうち真に漏洩していない率)を求めた．その結果，再現率（漏洩した件数のうちシステムが検知した率）は平均して90%程度であった．残り10%のFalse Negativeについては，キーワードの抽出に失敗したため検知できなかった．原因は，複数の名詞が連続した場合に適切に形態素解析ができなかったことである．そこで，改善策として，複数の名詞が連続していた場合，すべての組み合わせで複合語を作りキーワードとする方法が挙げられる．また，過去に一度でも想起を検知したことのある単語はシステムが学習し，次回以降の形態素解析の際に活用するなど，形態素解析の精度を向上する工夫が必要となる．また，情報漏えいした文章のうち約10%は，最初は人間が気付かなかったがシステムの検知を受けて読み直した結果，漏えいに気づいたケースである．このことから，提案システムにより人間の知識や能力を超えた検知が可能となり，ユーザの注意を補うことができるようになったと言える．

一方，適合率（システムが検知した件数のうち真に漏洩している率）は平均して30%程度であり，ユーザに不必要なアラートを発することが明らかになった．しかしながら，文書ソフトの校正機能のように，不必要なアラートであってもユーザに煩わしさを感じさせることの少ないインターフェースの提供は可能と考えられる．

今回，1つのプライバシーカテゴリに対して1つのNGワードを設定し評価を行った．今後の課題として，ユーザに負担のない形でNGワードの集合を集める方法を検討する必要がある．例えば，履歴書など，ユーザが普段書く経験があるものを活用する方法が挙げられる．NGワードの集合として履歴書を利用しソーシャルメディア上のコンテンツとの繋がりを計算する手法について4章で示す．

今後の課題は以下の通りである．

１．形態素解析を改良し，キーワード抽出の精度向上を行う．

２．ユーザに煩わしさを感じさせないインターフェースを検討する．

３．NGワードの収集方法として履歴書などを利用する方法を検討する．

４．通学・通勤先や職業のみならず，様々な種類のプライバシー情報について評価を行う．

# 第4章

# 履歴書との照合を通じたソーシャルメディア上の注目者の発言の特定

## 4.1 研究背景と目的

個人に関する様々な情報がインターネットなどの情報網に開示されるようになってきた．これらの情報の開示において，プライバシー侵害を避けるためにアクセス制限や匿名化など，多様な対策が実施されている．ソーシャルメディアを例に挙げると，発言や写真などのコンテンツの公開範囲の設定が各サービスによって提供されている．またユーザ自身においても，登録する氏名を友人だけに通じるニックネームに変更するなど，プロフィール情報を匿名化し対策している．特に日本人は匿名性を好む傾向にあり，日本の大手ソーシャルネットワークサービスである「mixi」では，半数以上のユーザが実名以外の名前で登録している[25].

これらの対策にも関わらず，同一人物に関する複数の情報の照合を通じて，その人物のプライバシーが侵害される危険性が懸念されている．特に，各メディアで匿名化されていた情報が，別情報と照合されることでその匿名性を失うことが，具体的な手法とともに指摘されている．たとえば，Narayanan は，ユーザ同士のつながりを表したソーシャルグラフを他のメディアのソーシャルグラフと比較することで，両メディアを利用する同一ユーザを特定できると報告している[12]．Goga は位置情報，タイムスタンプ，writing-style（文の長さ，前置詞の使用率などの文章の特徴）の情報を解析し，その結果を統合することによって異なるソーシャルメディアを利用する同一ユーザの特定を行っている[13]．このように，複数の情報の照合を通じて単独の情報からでは分からなかったことが明らかになり,その結果,個人の特定などのプライバシー侵害につながるリスクが懸念されている．しかしながら,これらの先行研究の多くは,ソーシャルメディア同士の照合を行っている．また，これらの先行研究の示すプライバシーリスクは比較的軽微であり，真の社会的リスクとまでは言えない.

そこで，本研究では，同一人物に関する異種の情報を照合する攻撃手法を示し，社会的リスクの存在を明らかにする．本研究の目的は，複数の情報の照合によるプライバシー侵害について，より現実的なリスクを明らかにし，プライバシー保護の意識向上と対策の必

要性を示すことである．その一つのステップとして，ここでは，個人の履歴書情報とソーシャルメディア上に開示された発言内容との照合のリスクを明らかにする．具体例として，企業などの組織が保有する個人の履歴書情報とマイクロブログ Twitter の発言である「つぶやき」を照合し，多数のつぶやきの中から履歴書の人物が発信したつぶやきを特定あるいは絞込める可能性を示す．その際，Twitter のユーザプロフィールの情報を使用せず，つぶやきの内容だけから照合を行う．このような照合が可能であることは，企業などが，履歴書情報を用いて，就職希望者や社員のつぶやきを特定できる可能性を意味する．つぶやきの特定ができれば，履歴書の人物について，交友関係や思想，日常の言動など様々な情報を入手することが可能となる．Twitter のユーザプロフィールを使用しないので，プロフィールをニックネームにするといった匿名化を行っていても，つぶやきの内容から特定される可能性がある．また，4.4 節，4.5 節で述べるように，氏名や職場名，所属学会名といった履歴書中の単語がつぶやきの中に書かれていなくても，特定される可能性がある．

履歴書に相当する情報は，Facebook などの実名指向のソーシャルメディアにおいて，ユーザプロフィールとして公開されているものもある．そのため，本論文の報告内容は，Facebook のユーザプロフィールと匿名で発言された Twitter のつぶやきとの照合を通じて，匿名のつぶやき者が特定される可能性を示唆する．さらに，ソーシャルメディアに限らず，大学や企業の Web サイトには，教員や役員の履歴書情報を個人プロフィールとして公開しているところがある．そのため，これらの公開された個人プロフィールとソーシャルメディア上の発言が照合され，個人の発言が特定される可能性も考えられる．

## 4.2 先行研究

ソーシャルメディアには個人に関する様々な情報が蓄積されているため，ソーシャルメディアからプライバシー情報を取り出す攻撃手法が多数報告されている．3.8 節で述べたように，Lam らは台湾のソーシャルメディアを対象として，72%のユーザのファーストネームを，また 30%のユーザのフルネームを推定した[11]．開示された情報から非開示の情報を推定する手法を提案し，非開示の年齢，性別，marital status を推定できると報告している[44]．Mao らは，旅行や病気に関するつぶやきを 76%の精度で，また飲酒運転のつぶやきを 84%の精度で検出できるとしている[46]．

近年，対象とするメディア以外からの情報を活用した非匿名化手法(De-anonymization)が提案されている．Narayanan らは，データベースの照合で用いていた手法[51]をソーシャルメディアに拡張し，2009 年にネットワークトポロジーの照合により 2 種類の異なるソーシャルメディアを利用する同一ユーザを特定する手法を提案している[12]．その中で，Twitter と Flickr を利用する同一ユーザを特定することができると述べている．また，2012

年に Goga らは位置情報，タイムスタンプ，writing-style（文の長さ，前置詞の使用率などの文章の特徴）の情報を解析し，その結果を統合することによって異なるソーシャルメディアを利用する同一ユーザの特定を行っている[13].

また，著者推定の技術もこのような攻撃によく用いられている．2004 年に Novak らは著者推定の技術を応用することにより，文章レベルの類似度を算出することで，同一人物により Web 上に投稿されたテキストを照合することが可能であると報告している[10]. 最近の研究では，2012 年に Narayanan がブログに投稿された匿名の発言の著者を，SVM（Support Vector Machine），ナイーブベイズ分類器，線形判別器など様々な手法によって推定することができると報告している[52].

また, 2010 年に Polakis らはソーシャルメディアのユーザ名と実在するメールアドレスを対応付けることが可能であると報告している．その中で，実際に Twitter と Facebook の情報を活用することでユーザプロフィールと実際のメールアドレスとの対応付けを行ったとしている[53]. 他にも，2011 年に Acquisti らは Facebook の写真を基に，出会い系サイトの写真に対して顔認識を行ったところ，出会い系サイトのユーザの氏名を特定することができたと報告している[54].

## 4.3 履歴書からの発言特定システム

### 4.3.1　提案システム概要

同一人物に関する異種の情報を照合する攻撃手法を示し，社会的リスクの存在を明らかにするために，以下のシステムの構築を行った.

個人の履歴書と，その個人（以下,「当該個人」と呼ぶ）の発言を含むソーシャルメディア上の多数の発言とを照合し，多数の発言の中から当該個人の発言を特定あるいは絞り込むシステムを提案する．その際，履歴書の内容と発言内容だけを用いて照合を行い，ソーシャルメディアのユーザプロフィールといった付加情報は用いない.

4.1 節で述べたように，このような照合が可能であることは，企業などの組織が履歴書情報を用いて，就職希望者や社員のつぶやきを特定できる可能性を意味する．つぶやきの特定ができれば，履歴書の人物について，様々な情報を入手することが可能となる．ソーシャルメディアのユーザプロフィールを用いないので，プロフィールをニックネームにするといった匿名化を行っても，なお特定の可能性がある.

このことは，現実的なプライバシー問題にもつながりうる．たとえば，欧米では，Background Checking とよばれる就職希望者を対象とした雇用前の身辺調査が広く行われている．近年行われている Background Checking では，就職希望者が提出した履歴書

などの情報を手がかりに，人事担当者が当該就職希望者のソーシャルメディアのアカウントを検索し，そこでの発言から，就職希望者の素行や友人関係，思想などを調査する．その結果次第では，選考対象から除外する企業もあるとされている[55]．ソーシャルメディアを用いた Background Checking は社会問題になっており，2012 年に米国カリフォルニア州にて，ソーシャルメディアのアカウントの開示を就職希望者に強要することを禁止する法案が提出された[56]．提案システムは，このような Background Checking をより広範囲かつ効率的に実行可能であることを示唆する．

履歴書に相当する情報は，実名指向のソーシャルメディアでユーザプロフィールとして多数公開されているだけでなく，大学の教員プロフィールや企業の役員プロフィール等の形で公開されることもある．そのため，提案システムは，これらの人物のソーシャルメディア上の発言が特定される可能性を示唆する．

なお，提案システムを個人の側が利用すれば，履歴書から自分の発言が特定されるかをチェックし，発言に注意を払う，早めに当該ソーシャルメディアから退会するといった対策につなげることも可能となる．

### 4.3.2　発言特定のモデル

本システムでは，当該個人のものを含むソーシャルメディアのユーザアカウント候補が与えられていると仮定した時に，その中から当該個人のユーザアカウントを特定する．特定手法は，当該個人の履歴書と各ソーシャルメディアユーザの発言内容との類似度を算出し，最も類似度が高いユーザを当該個人とみなす．

ソーシャルメディアの発言として Twitter のつぶやきを取り上げる．つぶやきは最大 140 文字であるため，複数のつぶやきの集合（M 件のつぶやきの集合）をまとめ，1 件の文書として扱う．ユーザ y の$i$番目のつぶやきを$T_{yi}$とすると，ユーザ y の$l$番目つぶやき集合（$M(l-1)+1$番目から$Ml$番目までの M 件のつぶやきの集合）は，以下の式 4.1 の$D_{yl}$で表される．例えば，$D_{Adam1}$は Adam の 1 番から M 番のつぶやきの集合である．本システムの発言特定処理の概要を図 4.1 に示す．

$$D_{yl} = \bigcup_{i=M(l-1)+1}^{Ml} T_{yi} \qquad \cdots \text{ 式4.1}$$

また，JIS Z8303:2008 解説の様式例に基づいた履歴書を調査に使用する．JIS 規格に基づいて，「氏名」「住所」「電話番号」「学歴」「職歴」「免許・資格」「志望動機」「アピールポイント」「通勤時間」などを記載している．なお，今回の調査では，中学卒業以降のデータを利用した．

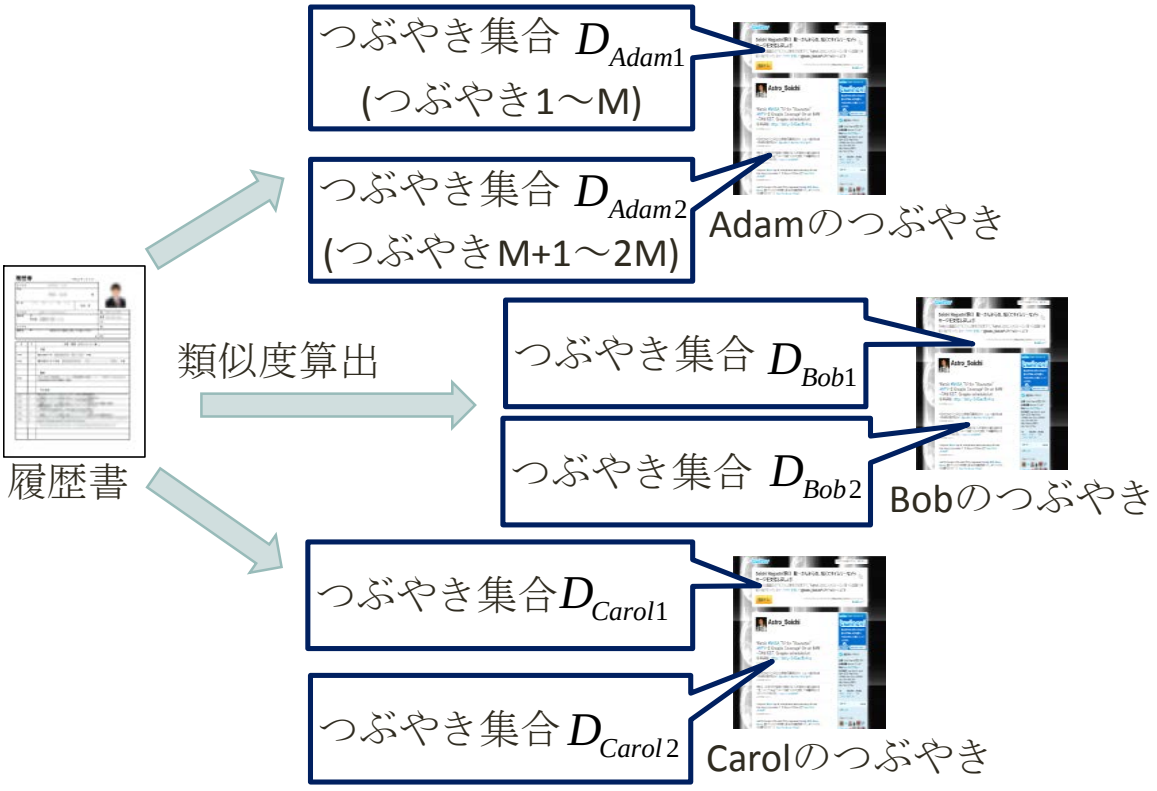


図 4.1　発言特定手法の概要

### 4.3.3　類似度算出の課題

4.2 節で挙げた通り，先行研究では，単一のソーシャルメディアを解析し個人情報を推定するか，または，複数のソーシャルメディア同士を照合していた．本研究では，異種メディアとしてソーシャルメディアと履歴書とを照合する．

ソーシャルメディアのコンテンツのうち，日記などは比較的文章が長い．つぶやきなどは短文であるが投稿頻度の高さから総文章数が多い．そのため，writing-style などコンテンツの特徴に着目した解析照合が可能となる．また，コンテンツの投稿時間や投稿場所，友人関係を表すソーシャルグラフなど，コンテンツ以外の付加情報が利用可能となる．

一方，履歴書の場合，学歴や職歴，資格などは学校名，会社名，資格名などキーワードの羅列であり，文章として記載する項目は志望動機などに限定され，その文章量も紙面に限りがあるため限定されている．そのため，writing-style などコンテンツの特徴に着目することはできない．また，履歴書には友人情報は記載されておらず，ソーシャルメディアにおける投稿時間や投稿場所に相当する情報も含まない．そのため，ソーシャルメディアと履歴書とを組み合わせて分析する場合，単語同士のマッチングのみ可能となる．しかし，履歴書は就職希望者が企業等に提出する書類であるため，記載される学校名や会社名など

は一般に公式名称である．一方，ソーシャルメディアはコミュニケーションが目的であるため，より口語的な表現となり，正式名称よりも略称や間接的な表現が使われると考えられる．そのため，同じ内容であっても直接照合することが出来ない可能性がある．

類似度算出アルゴリズムを検討するに当たり，これらの課題に対応する必要がある．

### 4.3.4　類似度算出アルゴリズム

履歴書とつぶやき集合の類似度を算出するアルゴリズムを類似度算出アルゴリズムと呼ぶことにする．本節では，類似度算出アルゴリズムについて述べる．

履歴書の情報はデータベースのレコードに近いものであるが，それを文書とみなすことで，履歴書とつぶやきの文書としての類似度を計算する．類似度の算出には，情報検索に用いられる代表的な手法の 1 つであるベクトル空間モデル[57]を用いる．これは文書を多次元ベクトルによって表現し，ベクトル間の類似度を算出する．類似度計算には余弦を用いる．個人 x の履歴書を$R_x$，ユーザ y の Twitter のl番目のつぶやき集合を$D_{yl}$とすると，余弦は以下の式により計算できる．

$$\mathrm{sim}(R_x, D_{yl}) = \cos(R_x, D_{yl}) = \frac{R_x \cdot D_{yl}}{\| R_x \| \| D_{yl} \|} \quad \cdots \text{式4.2}$$

ここで，・はベクトルの内積演算，‖ A ‖はベクトル A の大きさを表す．履歴書とつぶやき集合の類似度が高いほど，履歴書ベクトルとつぶやき集合ベクトルのなす角度が小さくなり，式 4.2 の値が大きくなる．各文書のベクトル表現法については，次章以降で説明する．

## 4.4 発言特定方式

### 4.4.1　基本的な方式

履歴書のベクトル表現として，式 4.3 のように履歴書中の名詞の重みを用いる．ここで，$W_{R_x,j}$は個人 x の履歴書$R_x$における名詞 j の重み，S は履歴書における単語数である．

$$\begin{pmatrix} \text{履歴書における名詞1の重み} w_{R_x,1} \\ \text{履歴書における名詞2の重み} w_{R_x,2} \\ \vdots \\ \text{履歴書における名詞Sの重み} w_{R_x,S} \end{pmatrix} \quad \cdots \text{式4.3}$$

Twitter のつぶやき集合についても，同様に，つぶやき集合中の名詞の重みを用いてベクトル表現とする．ここで，Twitter ユーザ y の$l$番目のつぶやき集合$D_{yl}$における名詞 j の重みを$w_{D_{yl},j}$で表すことする．名詞の重みを用いたベクトル表現であるため，word-style など文章の特徴は使用しておらず，履歴書との照合に用いることができる．

名詞の重み$W_{R_x,j}$，$w_{D_{yl},j}$を求めるために，形態素解析器である Mecab[39]と専門用語ツール[40]を用いて履歴書およびつぶやきの形態素解析を行い，名詞を抽出する．Mecab の解析用辞書には IPA（独立行政法人 情報処理推進機構）が提供する辞書があるが，それ以外に，独自の辞書として Wikipedia のカテゴリ名[41]とはてなキーワード[42]をそれぞれ名詞として登録し，現代語に対応した．以降の形態素解析にも同様の環境を用いることとする．

抽出した名詞に対して TF-IDF[58]を用いた重みを算出した．TF-IDF とは，文章中の特徴的な単語を抽出するために主に情報検索などで利用される指標であり，TF(TermFrequency : 単語の出現頻度)と IDF(Inverse Document Frequency : 文書の逆出現頻度)の 2 種類の指標の乗算により計算される．履歴書$R_x$，つぶやき集合$D_{yl}$における単語 j の TF-IDF を用いた重みは以下の式 4.4，4.5 により表される．

$$W_{R_x,j} = tf_{R_x,j} \times idf_j = \frac{n_{R_x,j}}{\sum_k n_{R_x,k}} \times log\frac{N}{N_j} \quad \cdots \text{式4.4}$$

$$w_{D_{yl},j} = tf_{D_{yl},j} \times idf_j = \frac{n_{D_{yl},j}}{\sum_k n_{D_{yl},k}} \times log\frac{N}{N_j} \quad \cdots \text{式4.5}$$

$n_{R_x,j}\left(n_{D_{yl},j}\right)$は，履歴書$R_x$（つぶやき集合$D_{yl}$）における単語 j の出現回数，N は参照文書の総数，$N_j$は参照文書のうち単語 j を含む文書の数である．ここでは，参照文書としてインターネット上の全 Web ページを用いることとし，N には Web ページの総数の予想値[59]を用いた．$N_j$には単語 j を含む Web ページの数を用いることとし，具体的には，単語 j をキーワードとして Yahoo! JAPAN が提供する検索 API[43]により Web 検索を行った時の Web ページの検索件数を用いた．

この重みを用いることで，式 4.2 は以下の式 4.6 のように変形できる．

$$sim(R_x, D_{yl}) = \frac{R_x \cdot D_{yl}}{\| R_x \| \| D_{yl} \|} = \frac{\sum_{j=1}^{S} W_{R_x,j} w_{D_{yl},j}}{\sqrt{\sum_{j=1}^{S} W_{R_x,j}^2} \sqrt{\sum_{j=1}^{S} w_{D_{yl},j}^2}} \quad \cdots \text{式4.6}$$

サンプルデータとして，社会人 Adam（個人情報保護のため仮名）の履歴書とつぶやきを用い，基本方式の性質を分析した．なお，Adam は，都内の理工系大学修士課程を卒業後，通信関係の会社に勤務している 20 代後半の男性である．2011 年 1 月を起点に Adam の 500 件のつぶやきを用意し，上記の手法により名詞の重みを算出した，図 4.2 にその結果を示す．図 4.2 の横軸は，Adam の履歴書に含まれる名詞である．縦軸は Adam のつぶやき集合におけるこれらの名詞の重み，すなわち式 4.5 の TF-IDF 値$W_{D_{yl},j}$を表している．

Adam のつぶやきにおいて「コンピュータ」「通信」などの語句の TF-IDF 値は 0 ではないが，Adam の出身大学名や所属学会名などのより個人特定につながる語句の TF-IDF 値は 0 である．即ち，つぶやきの中に一度も登場しないことが分かる．このように履歴書の本人のつぶやき集合であるにも関わらず，履歴書に含まれる 80 の名詞のうち 70 の名詞について，つぶやきにおける重み（TF-IDF 値）が 0 となる．この状態では，同一人物の履歴書とつぶやきの類似度が低くなる．

その原因として，履歴書に含まれる名詞がつぶやきの中で直接使われる機会が多くないことが挙げられる．たとえば「電気通信大学」の学生は，電気通信大学という正式名称を用いるかわりに，電通大や UEC といった略語を利用することが多い．また，履歴書中の名詞は個人の特定につながるものが多いため，プライバシーを意識する人は，「調布駅付近の理工系大学」などのように，特定のコミュニティに属する人の間でのみ通じる表現や単語の組み合わせによる間接的な表現に言い換えることもある．このように，つぶやきでは，履歴書の単語が直接表現されるとは限らず，単語の組合せや略称によって表現されることが多い．また，調布駅付近の大学は電気通信大学に限られるといった，外部知識との組み合わせによって履歴書の単語に到達できる表現も用いられる．このように直接記載されない情報については，つぶやきで示唆されていたとしても，TF-IDF のように同一語句の出現頻度に基づく手法を用いた場合，類似度に反映させることができない．

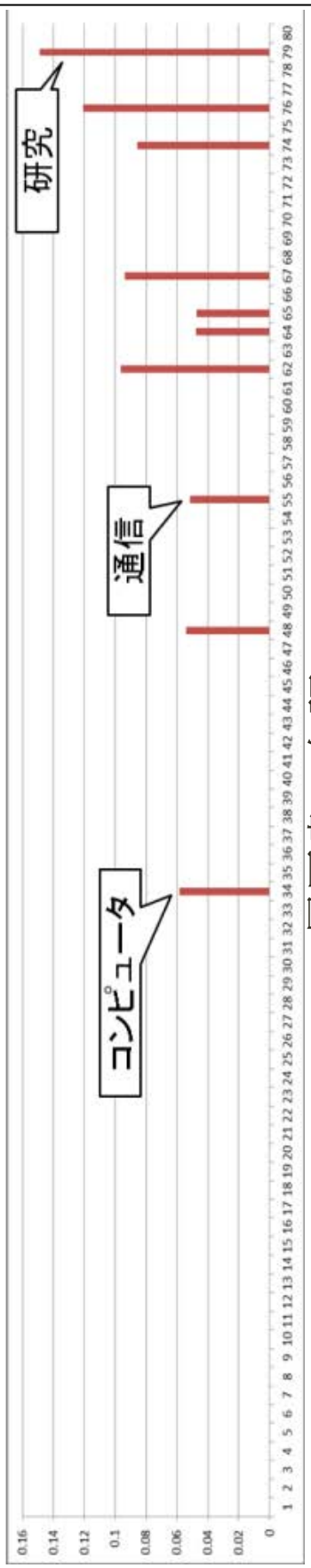


図4.2　Adamのつぶやきにおける名詞の重み（基本方式）

### 4.4.2　Web 検索の利用

上記の基本的な方式の問題点をまとめると下記のようになる．

(1) 履歴書において，プライバシー情報は，正式名称などの名詞によって表現されている．

(2) つぶやきにおいて，プライバシー情報は，正式名称だけでなく，略称，単語の組合せ，間接的示唆など多様な表現形態を有するので，語句の共通性だけでは，履歴書とつぶやきの類似度を正しく算出することができない．

そこで，システムは，つぶやきの多様な表現と履歴書の名詞との結びつきを検知し，類似度として定量化する必要がある．たとえば，つぶやきと履歴書の単語に全く関連がない場合には 0，つぶやき内に履歴書の単語が直接記載されている場合には 1，間接的な示唆の場合は 0.5 といった定量化を行う必要がある．

そのために，3 章で提案した DCNL の想起検知手法を拡張して用いることを考える．履歴書の名詞を想起検知の NG ワードと考えると，つぶやきの表現と履歴書の名詞とのつながりを Web 検索によって定量化することができると考える．つぶやきの表現と履歴書の名詞との繋がりをこのように定量化した値を Hidden Term Frequency（HTF，隠れ出現頻度）と呼ぶことにする．図 4.3 に HTF の考え方を示す．

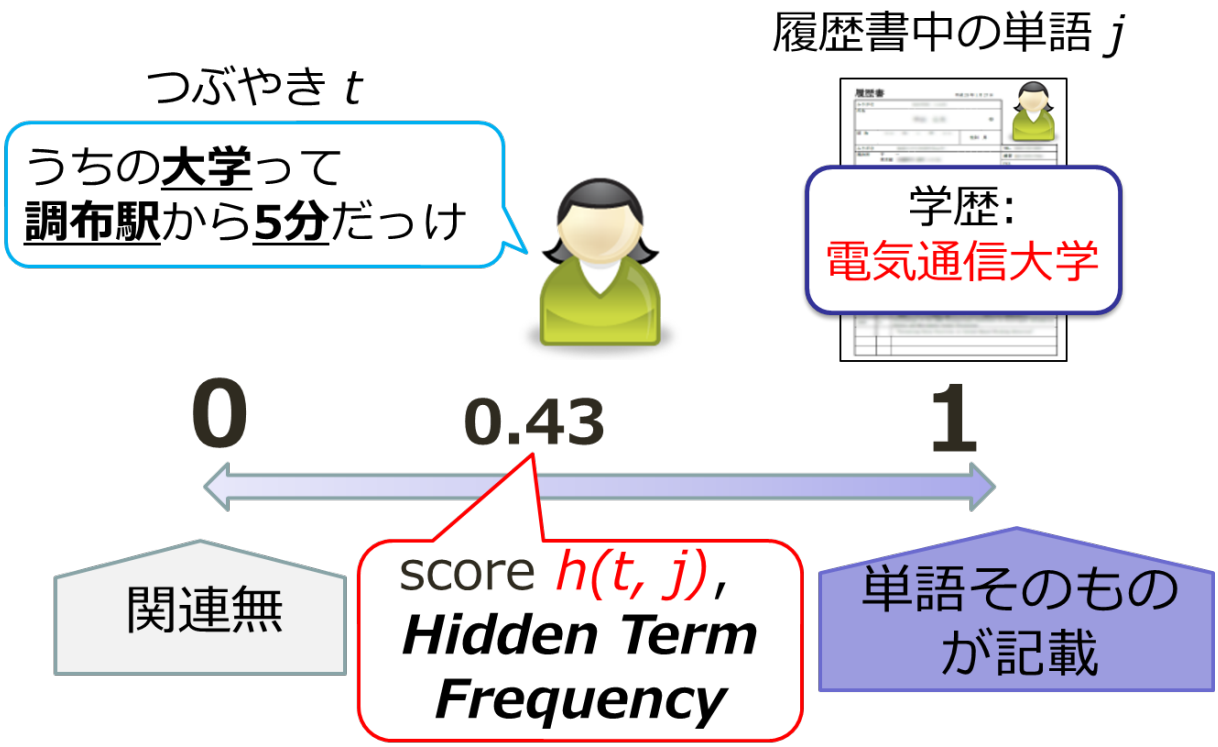


図 4.3　HTF の考え方

### 4.4.3　改良方式

つぶやき集合$D_{yl}$が履歴書中の名詞 j を示唆する強さを$HTF(D_{yl}, j)$とし，それを自動的に 0 から 1 の値に定量化する手法を提案する．以下に，その詳細を示す.

I. つぶやき集合$D_{yl}$からb$(1 \leq b \leq M)$番目のつぶやきを抽出する($D_{yl}$は M 件のつぶやきの集合).

II. 抽出したつぶやきから名詞とその複合語を形態素解析により抽出する．抽出した語句の個数を n とする.

III. その中から最大 m 件の語句の組合せをクエリとして抽出し, クエリリストを作成する．クエリリストの総数 Q は以下の式で表される．今回は実装の都合上 m=3 とした.

$$Q = \sum_{a=1}^{m} \binom{n}{a} \quad \cdots \text{ 式4.7}$$

IV. クエリリストに含まれるクエリ$q_i(1 \leq i \leq Q)$について以下の処理を繰り返し行う.

A) クエリ$q_i$で Web 検索し，検索結果の上位 k 件のタイトルとテキストサマリーを取得する.

B) c$(1 \leq c \leq k)$番目の検索結果のタイトルとテキストサマリーについて，履歴書の名詞 j が含まれているか探索し，以下の式により$f_c$の値を返す．今回は実装の都合上 k= 20 とした.

$$f_c = 1(if\ noun\ j\ is\ included),$$
$$0(if\ not\ included) \quad \cdots \text{ 式4.8}$$

C) $f_c$を用いて$q_i$のスコア$sqore(q_i)$を以下のように計算する．この値は，式 4.10 で表される D を用いて 0 から 1 の値に正規化されている.

$$score(q_i) = \frac{1}{D}\sum_{c=1}^{k}(k - c + 1)f_c \quad \cdots \text{ 式4.9}$$

$$D = \sum_{c=i}^{k}(k - c + 1) \quad \cdots \text{ 式4.10}$$

V. 全てのクエリ$q_i$のスコア$sqore(q_i)$から，その最大値を選定し，$h(b, j)$とする．これは b 番目のつぶやきが単語 j を示唆する強さを表す.

$$h(b,j) = \max_{1 \le i \le Q} (score(q_i)) \quad \cdots \text{ 式4.11}$$

VI. IからVまでの手順をM件のつぶやきについて繰り返し，$h(b,j)$の平均値を求める．その値を$HTF(D_{yl},j)$とする．

$$HTF(D_{yl},j) = \frac{1}{M}\sum_{b=1}^{M} h(b,j) \quad \cdots \text{ 式4.12}$$

手順 V で得られた値は，つぶやきが履歴書の名詞をどの程度強く示唆するかを算出して 0 から 1 の値に定量化したものである．例えば，検索結果 20 件全てに履歴書の名詞が含まれていたら，この値は 1 となり，19 件目と 20 件目に含まれていたら，値は$(2+1)/210 \cong 0.014$となる．この手法を用いることで，例えば，「大学院入試説明会が今週 5 月 22 日土曜日に開催されます. 同時に研究室公開も実施します」というつぶやきが「電気通信大学」を示唆する強さを定量化することが可能となる．なお，この例で得られる $h(b,\ j)$は 0.12 となる．

本手法によると式 4.9 の$(k-c+1)$の働きにより，履歴書の名詞が検索結果上位に来るほど示唆の強さが大きくなる．また，式 4.9 において$c=1$から$k$ までの和をとっていることにより，履歴書の名詞が多くの検索結果に含まれるほど示唆の度合いが大きくなる．

最後に，得られた HTF の値を式 4.5 の$tf_{D_{yl},j}$ の代わりに用いる．類似度計算は前回と同様に式 4.6 を用いる．

以上の I から VI の手順で求めた$HTF(D_{xl},j)$を$tf_{D_{yl},j}$の代わりとし，式 4.5，式 4.6 により類似度を算出する方式を改良方式と定義する．この改良方式をスクリプト言語 Python と 4.4.1 項で述べた形態素解析の環境を用いて実装した．

改良方式を利用した場合の名詞の重みを図 4.4 に示す．図 4.4 は，図 4.2 における縦軸を TF と IDF の乗算ではなく HTF と IDF の乗算の値に変えたものである．

基本方式では，履歴書内のほとんどの名詞がつぶやきに直接現れないため，類似度の算出に寄与していなかった．しかし，改善方式を用いることで，80 の名詞のうち，66 の名詞の重みが正の値になった．実際に「(出身大学名)」「セキュリティ」「(所属学会名)」などの，つぶやき内に直接記載されていなかった名詞について，示唆の強さが定量化され，類似度の算出に寄与していることが分かる．

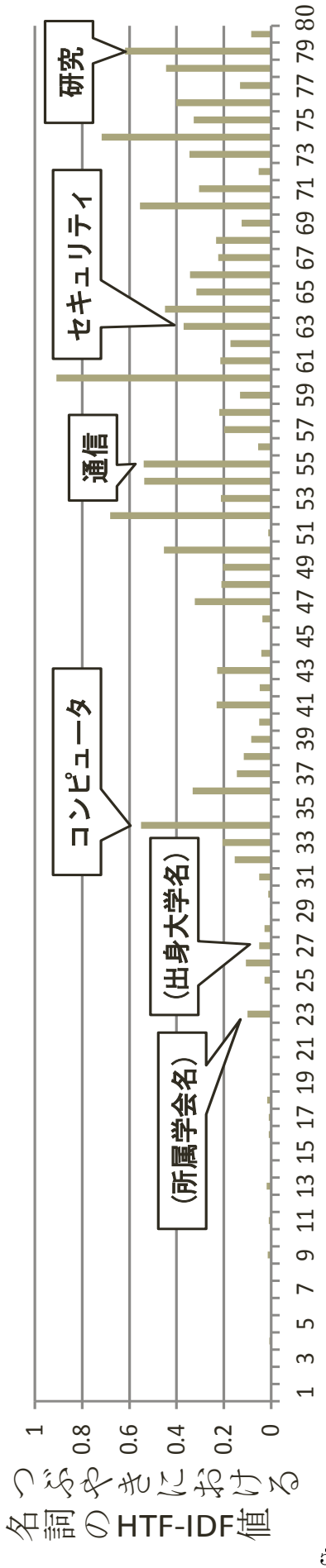


図4.4　Adamのつぶやきにおける名詞の重み（改良方式）

## 4.5 予備評価

### 4.5.1　概要

4.4 節で述べた改良手法を用いて, 多数のアカウントからの当該個人のアカウントの特定を行う．今回は，Twitter ユーザ 100 人と当該個人の合計 101 人のユーザのつぶやきの中から当該個人のつぶやきを特定できるか評価する．

### 4.5.2　実験 1

#### 4.5.2.1　サンプルデータ

予備実験で用いた社会人 Adam に加えて，社会人 Bob の履歴書を用いた．Bob は都内理工系大学の 30 代男性准教授である．これら 2 人のつぶやきの他に，2012 年 9 月 5 日，8 日，10 日の 3 日に渡ってつぶやいた，日本語を主言語とする Twitter ユーザを合計 100 人ランダムに選び，各 1000 件分のつぶやきを取得した．なお，Bob のつぶやきの総数は 700 件であったので，100 人についても 700 件を評価に用いた．つぶやき集合のサイズ M を 100 とし，つぶやき 100 件毎に履歴書との類似度を算出した．

#### 4.5.2.2　評価

図 4.5 に結果を示す．図 4.5 のグラフの横軸の数字は各ユーザのつぶやき集合の番号を表している．例えば，横軸の 1 は$D_{y1}$，すなわちユーザ y の 1 番目のつぶやき集合を表す．ここで，ユーザ y は，Adam を含む 101 人である．縦軸は，Adam の履歴書と各ユーザのつぶやき集合との類似度を表している．

図 4.5 中の■の点は Adam の履歴書と Adam のつぶやき集合との類似度を表している．箱ヒゲ図は，Adam の履歴書と他の 100 人のユーザのつぶやき集合との類似度の分布を表している. 黒い箱の中に第 1 四分位点から第 3 四分位点までの 50%のユーザのつぶやき集合が含まれる．箱の上下の直線は，上位および下位 25%のユーザのつぶやき集合の分布を示している．横軸 1 のグラフに着目すると，Adam の履歴書と Adam のつぶやき集合 1 との類似度が，Adam の履歴書と他の 100 人のつぶやき集合 1 の類似度の分布よりも上にあることが見て取れる. 図 4.5 より全てのつぶやき集合において, ■点が箱ヒゲ図よりも上に位置すること，すなわち履歴書と本人のつぶやきとの類似度が 101 人中最大となることが分かる．図 4.6 は図 4.5 のつぶやき集合 1（横軸 1）における他人 100 人の類似度の度数分布を示したものである．

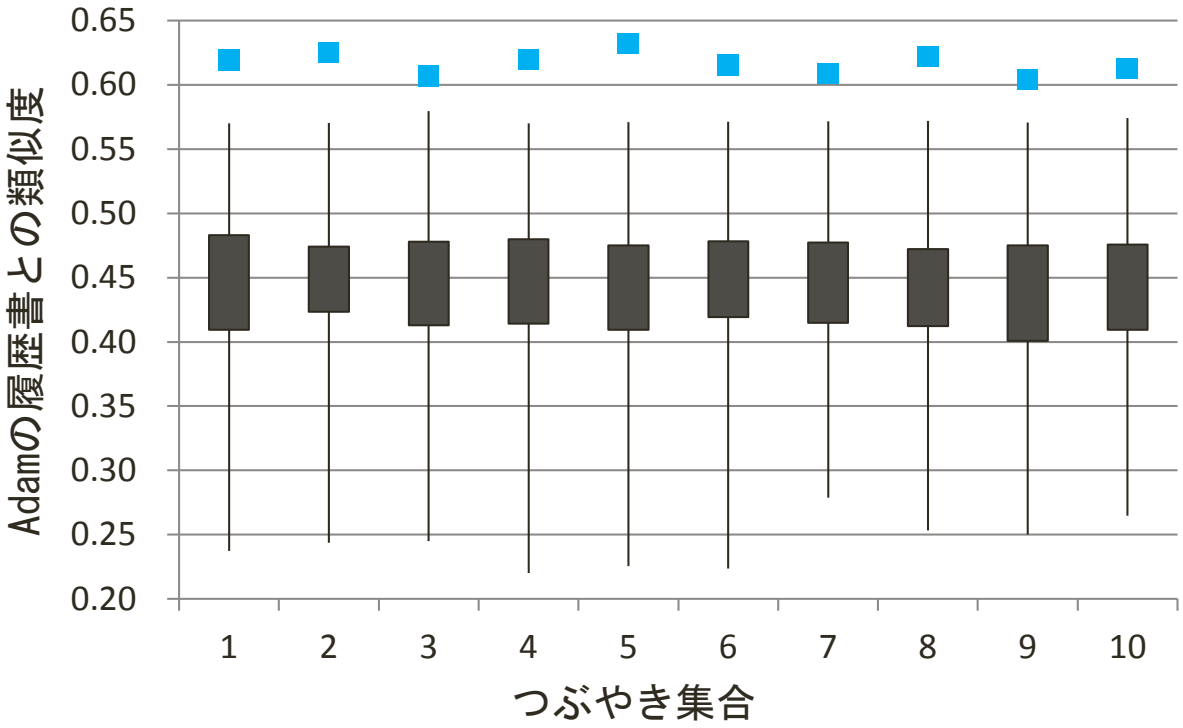


図 4.5　Adam の履歴書と各ユーザのつぶやきとの類似度

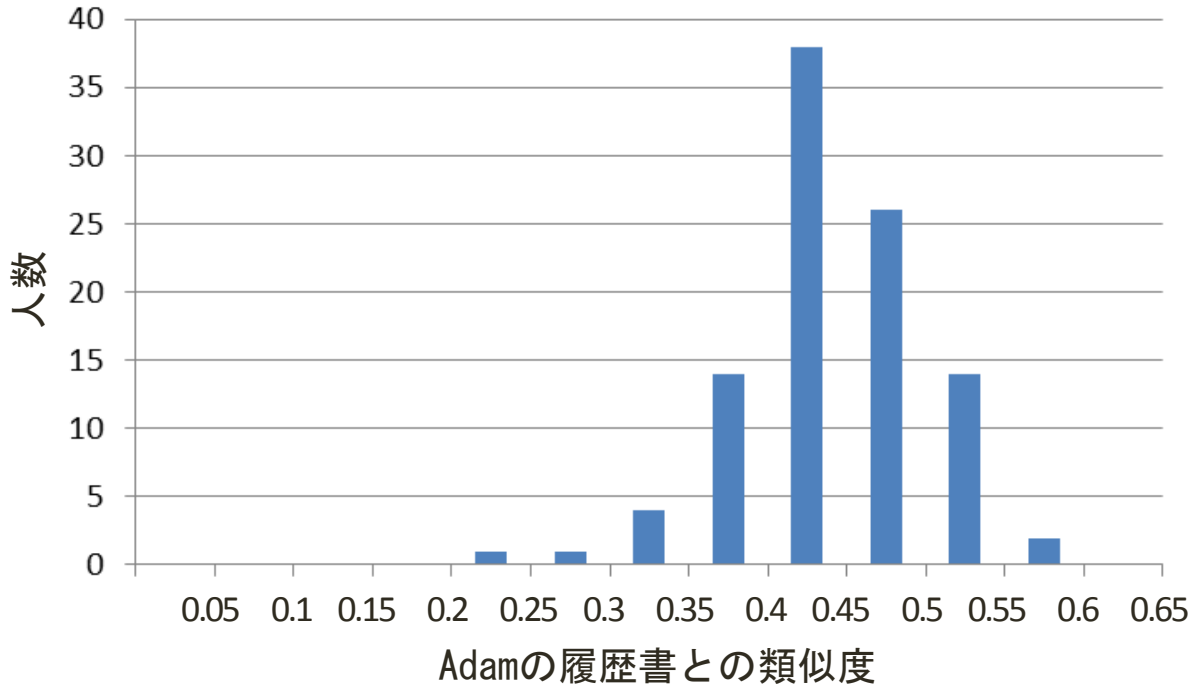


図 4.6　つぶやき集合 1 における 100 名のユーザの類似度の度数分布(Adam)

表 4.1　つぶやき特定の精度（Adam）

| 判定＼事実 | Adamのつぶやき | Adam以外のつぶやき |
|---|---|---|
| Adamのつぶやき | 10 | 0 |
| Adam以外のつぶやき | 0 | 1000 |

類似度が最大のつぶやき集合を Adam 本人のつぶやき集合と判定した場合の判定精度を表 4.1 に示す．Adam の場合，本人のつぶやき集合の類似度が常に最大になるため TPR(True Positive Rate)=1.0，他人のつぶやき集合の類似度は一度も最大とならないため TNR(True Negative Rate) = 1.0 である．

図 4.7，図 4.8，表 4.2 は Bob について同様の結果を表している．つぶやき集合 1 のグラフを見ると，Bob の履歴書と$D_{Bob1}$との類似度が最大でなく，2 番目に高い類似度となっている．このように，7 つのつぶやき集合のうち 1 つで本人よりも他人が高くなるため TPR=0.857，TNR = 0.998 となる．

2 人の評価結果の平均をみると，TPR=0.941，TNR=0.999 となり，高い精度で本人のつぶやきと他人のつぶやきを識別できることが分かる．

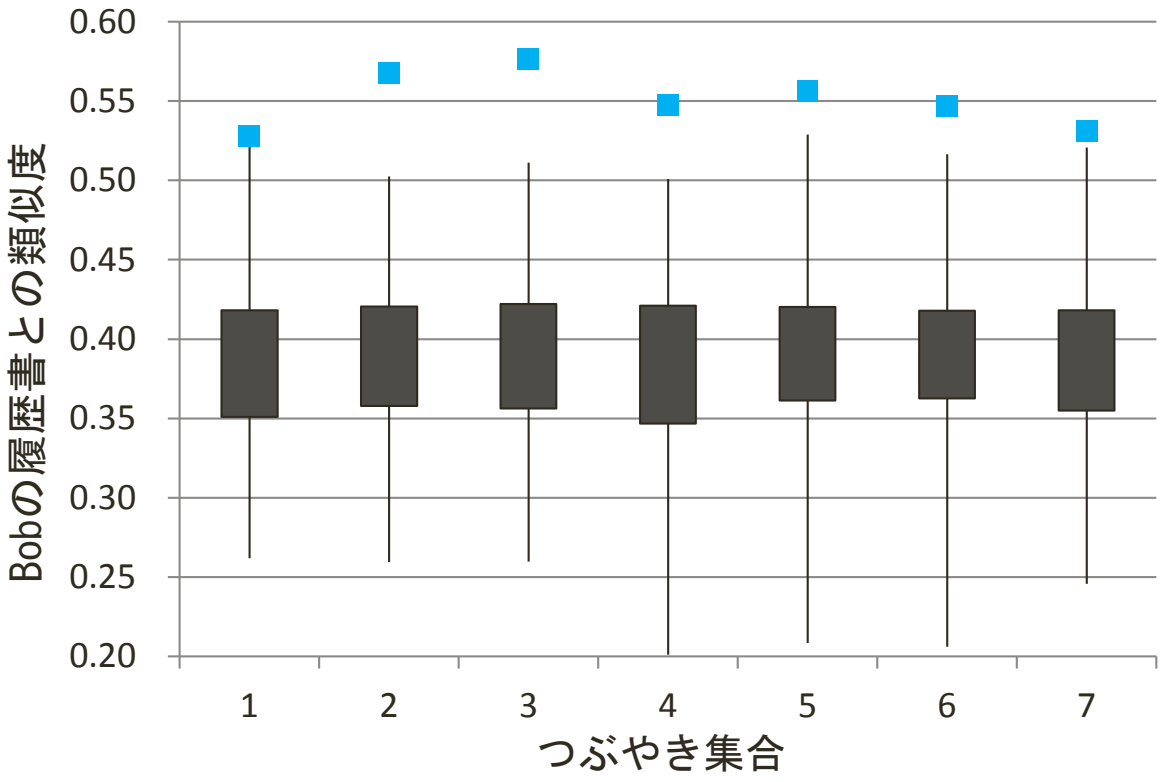


図 4.7　Bob の履歴書と各ユーザのつぶやきとの類似度

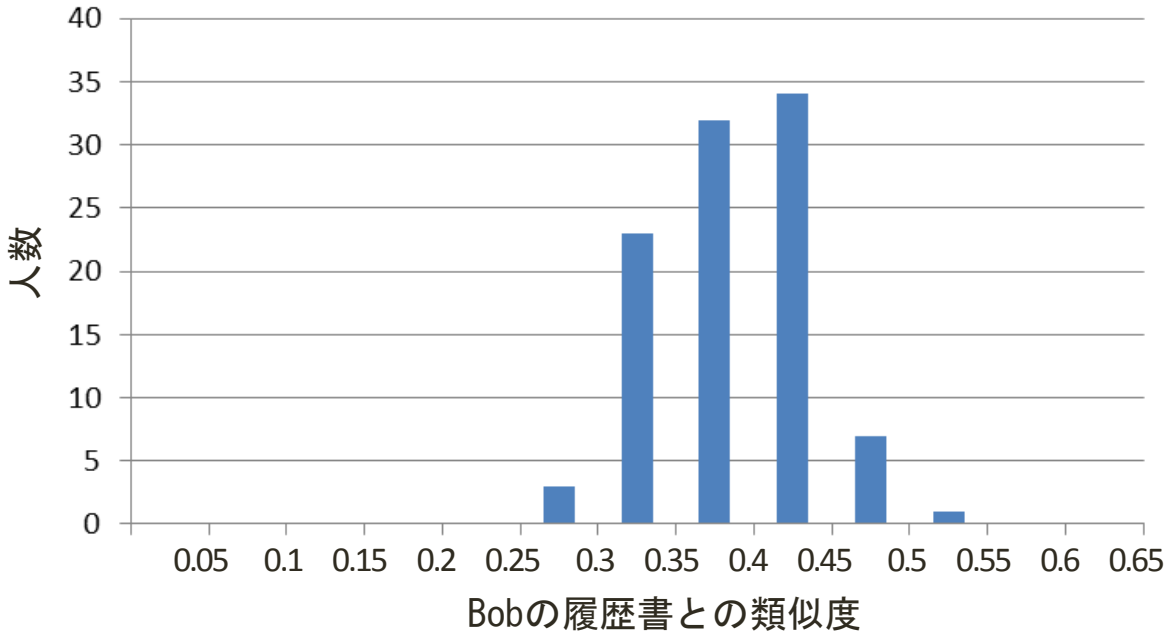


図 4.8　つぶやき集合 1 における 100 名のユーザの類似度の度数分布(Bob)

表 4.2　つぶやき特定の精度（Bob）

| 事実 / 判定 | Bobのつぶやき | Bob以外のつぶやき |
|---|---|---|
| Bobのつぶやき | 6 | 1 |
| Bob以外のつぶやき | 1 | 699 |

### 4.5.3　実験 2

#### 4.5.3.1　サンプルデータ

3.7.2 節と同様，電気通信大学情報理工学部総合情報学科 9 名，情報・通信工学科 1 名，知能機械工学科 1 名の 3 年男子学生より，本人の了解を得て，履歴書と Twitter アカウントを収集した．いずれの学生も Twitter で本名を公開していない．以降，10 人の学生をそれぞれ{学生 1，学生 2，学生 3…学生 10}と表記することにする．収集した履歴書情報の項目は，(1) 氏名，(2) 生年月日，(3) 性別，(4) 現住所（市・区・郡まで），(5) 帰省先住所（市・区・郡まで），(6) 学歴，(7) 職歴，(8) 交通手段，(9) 電車区間，(10) 得意科目，(11) 長所・特徴（自己 PR），(12) クラブ活動・サークル・趣味である．

この中で，実際に評価に用いたデータは(4)現住所，(5)帰省先住所，(6)学歴，(7)職歴，(8)交通手段，(9)電車区間，(10)得意科目，(11)長所・特徴，(12)クラブ活動・サークル・趣味の計 9 種類である．

今回のサンプルデータの特徴として，学生であるためか日常会話に近いつぶやきが多く見られることが挙げられる．また，予備実験の時に用いていた社会人の履歴書情報は 80 項目以上であったのに対し，どの学生においても履歴書情報が 30 項目前後であることがあげられる．したがって，類似度計算で作成されるベクトルの次元が 80 次元から 30 次元に減少する．このことから，社会人のデータを利用した実験 1 よりも特定精度が低くなることが予想される．

#### 4.5.3.2　手法

実験 1 の時と同様に，各ユーザの 1000 件のつぶやきを 100 件毎に 10 個のつぶやき集合に分割し，履歴書とつぶやき集合の類似度を算出した．手法は実験 1 と同じである．

#### 4.5.3.3　評価

学生 1 の履歴書と Twitter ユーザのつぶやき集合の類似度を実験 1 と同様に図 4.9 に示す．また，学生 2 から学生 10 までの結果を同様に図 4.10 から図 4.18 に示す．

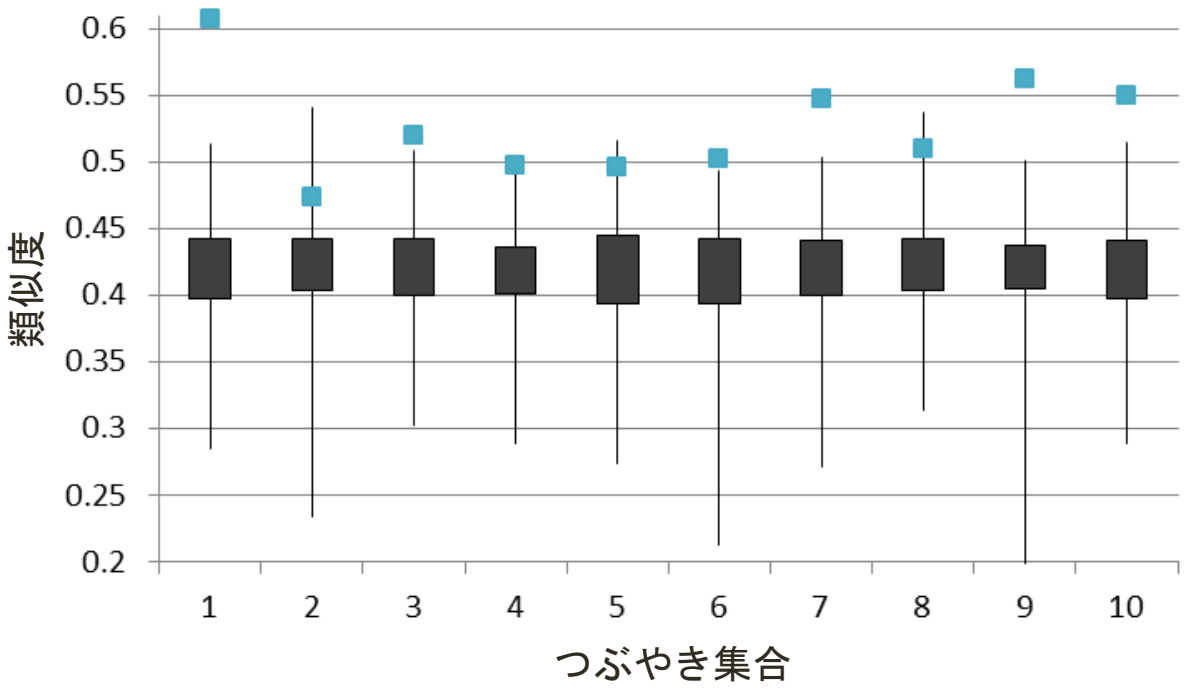


図 4.9　学生 1 の履歴書と各ユーザのつぶやきとの類似度

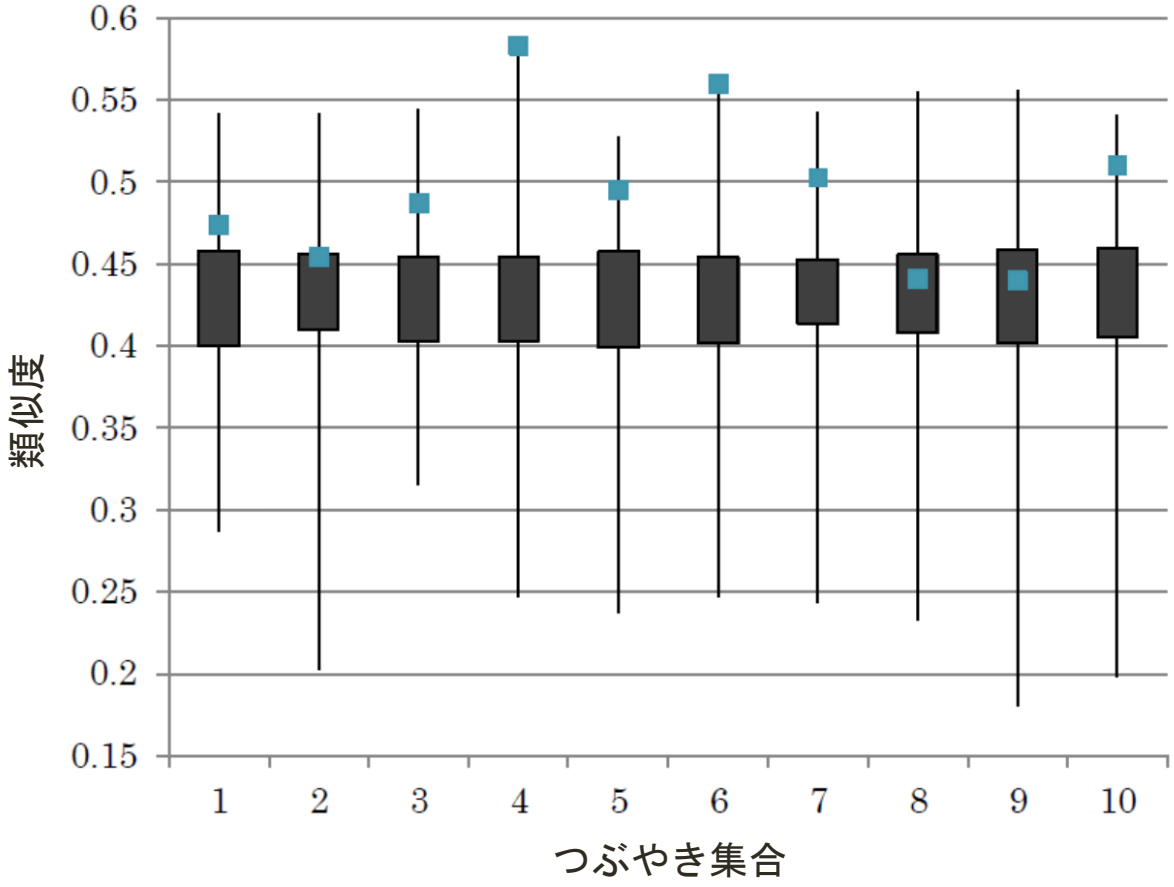


図 4.10　学生 2 の履歴書と各ユーザのつぶやきとの類似度

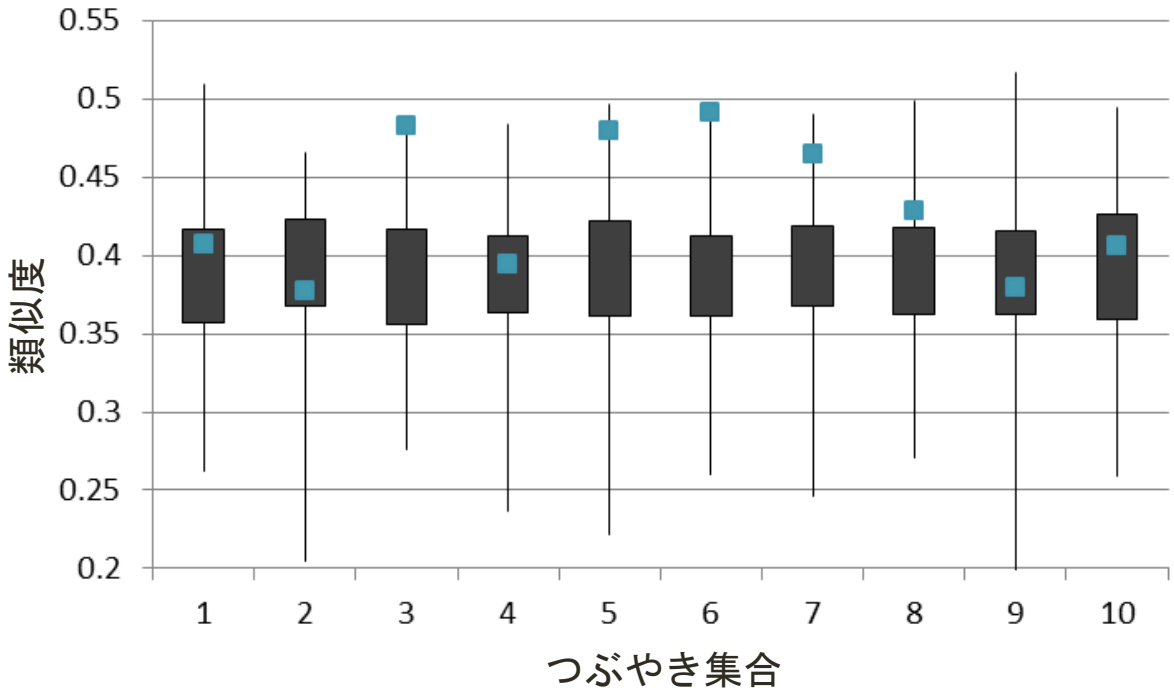


図 4.11　学生 3 の履歴書と各ユーザのつぶやきとの類似度

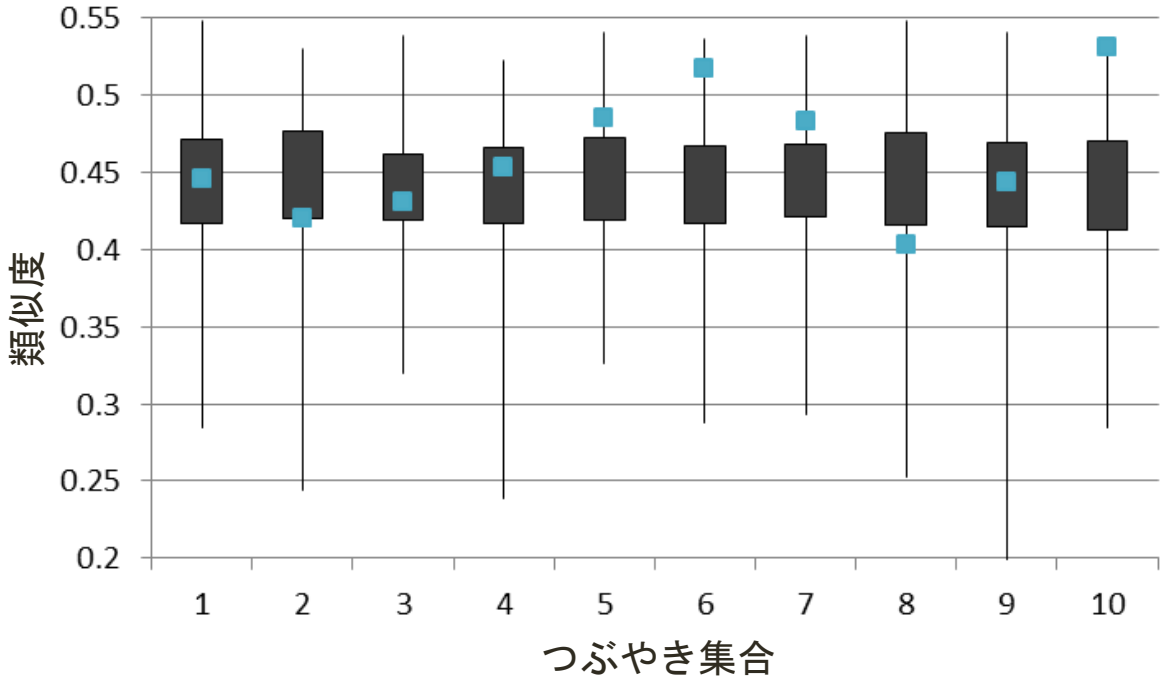


図 4.12　学生 4 の履歴書と各ユーザのつぶやきとの類似度

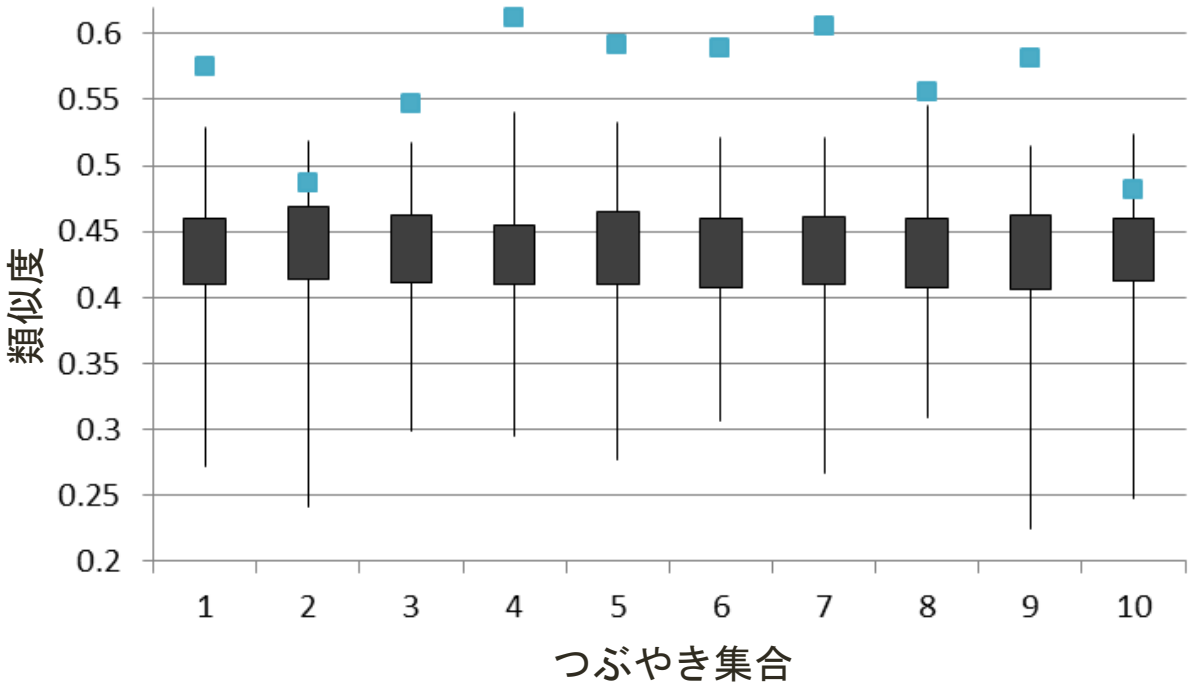


図 4.13　学生 5 の履歴書と各ユーザのつぶやきとの類似度

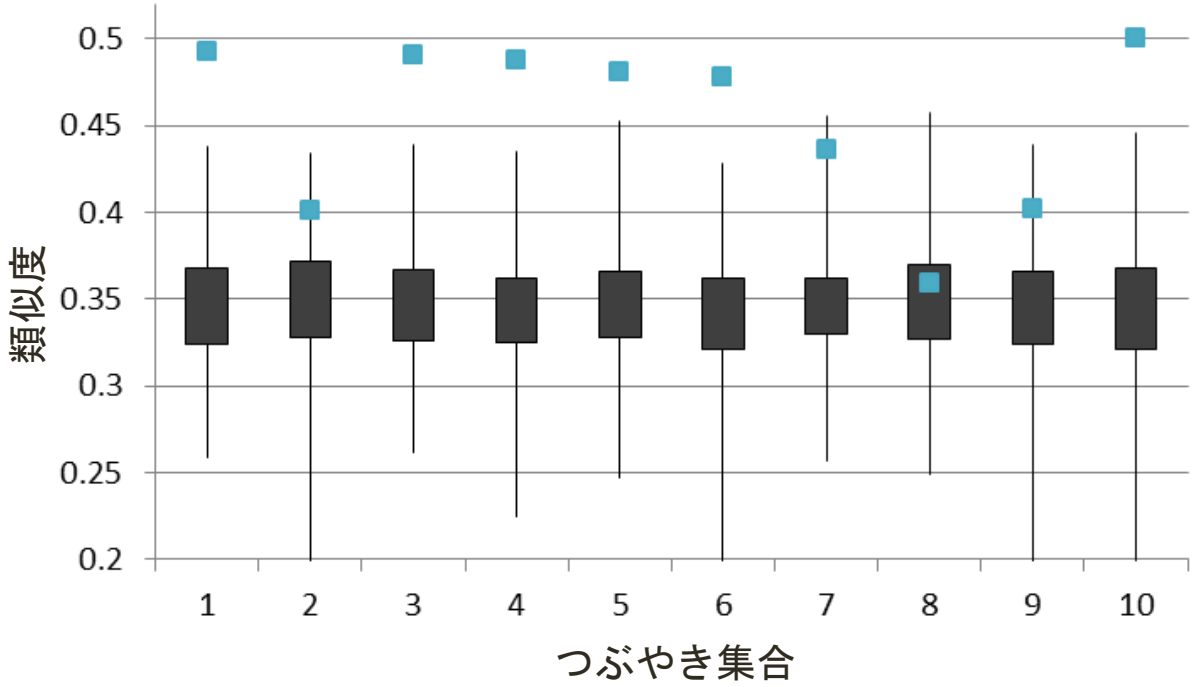


図 4.14　学生 6 の履歴書と各ユーザのつぶやきとの類似度

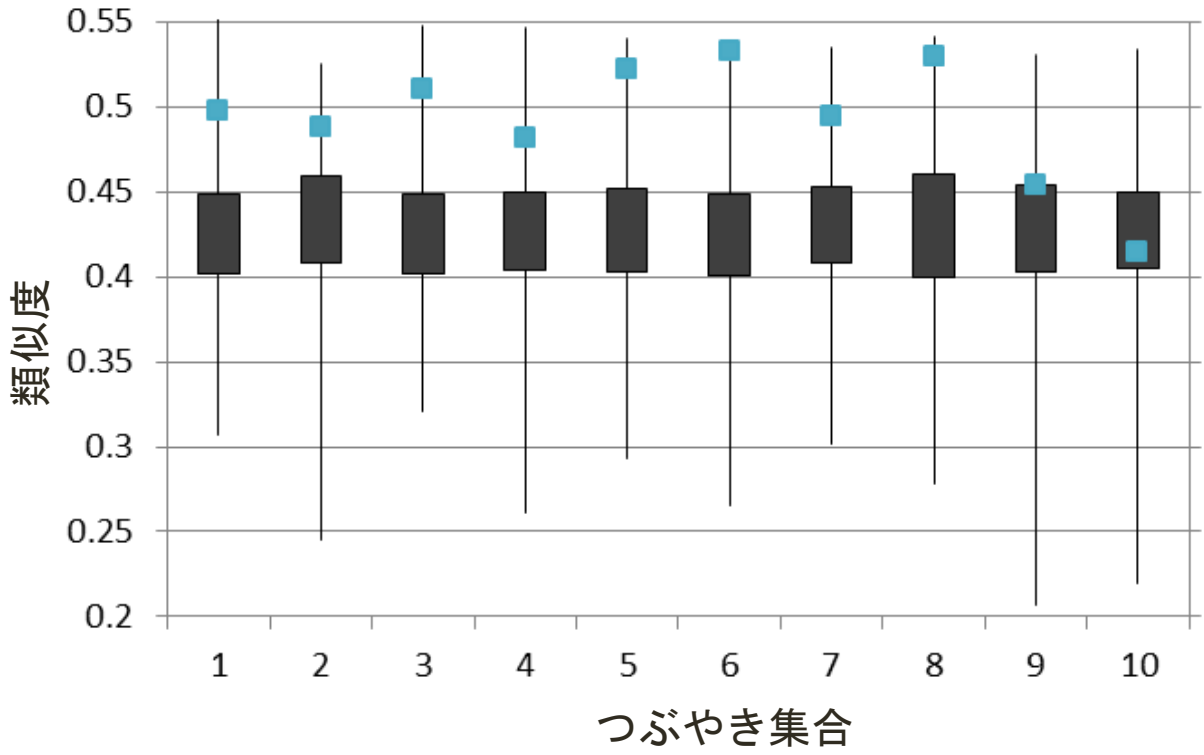


図 4.15　学生 7 の履歴書と各ユーザのつぶやきとの類似度

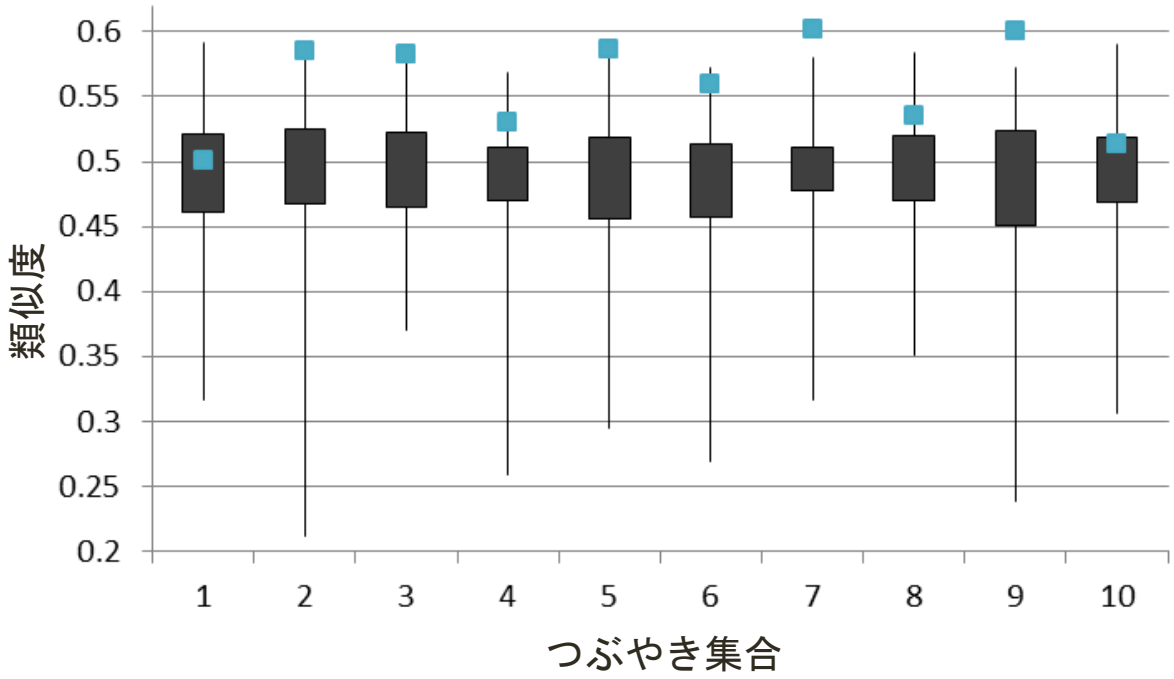


図 4.16　学生 8 の履歴書と各ユーザのつぶやきとの類似度

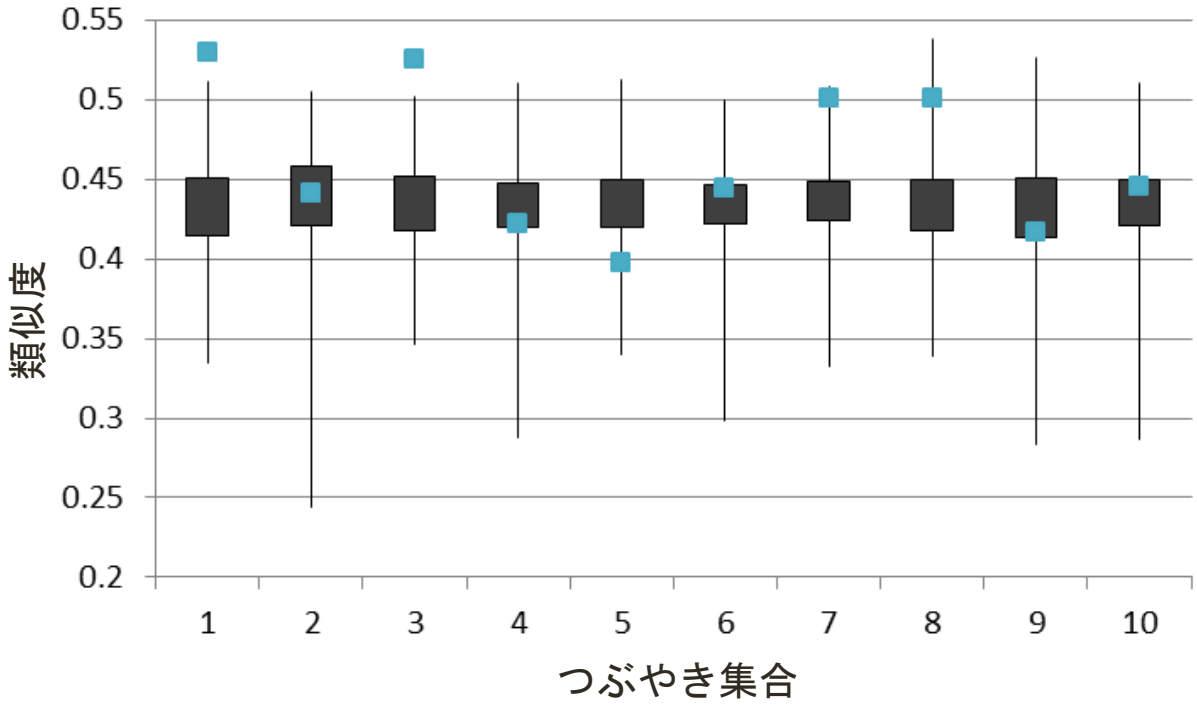


図 4.17　学生 9 の履歴書と各ユーザのつぶやきとの類似度

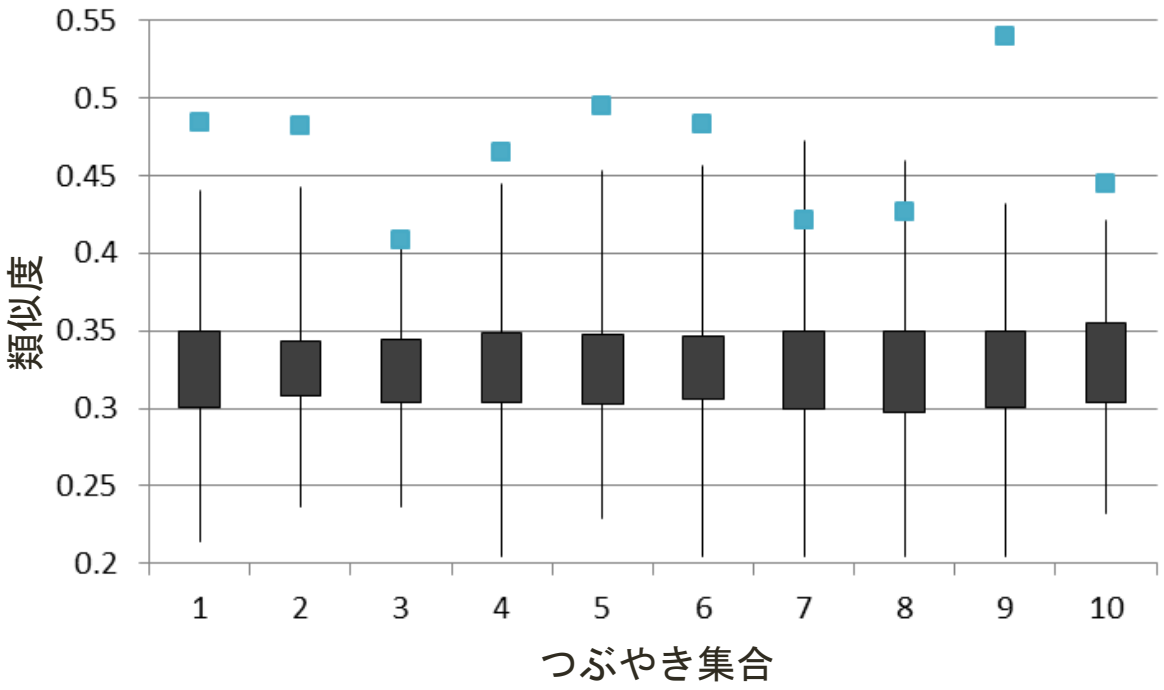


図 4.18　学生 10 の履歴書と各ユーザのつぶやきとの類似度

特定・絞り込みの精度を確認するために，それぞれの学生について，履歴書と本人のつぶやき集合との類似度が，101 人中上位何人に入るかを表 3 にまとめた．たとえば，学生 1 の行は，10 個のつぶやき集合のうち 7 個の類似度が 101 人中最大であり，10 個全てが上位 5 人に含まれることを示している．表 3 より，学生本人のつぶやき集合の類似度が最大となるのは，のべ 100 個のつぶやき集合のうち 37 個である．また，本人のつぶやき集合の類似度が 101 人中上位 5 人以内となったのは 59 個，上位 10 人以内となったのは 67 個であった．このように，実験 1 と比べると精度は全体的に低くなっている．

しかし，10 個のつぶやき集合の類似度の平均値（つまり，履歴書と 1000 件のつぶやき集合との類似度）に着目することで，本人を特定することが可能になる．図 4.19 は，各学生の履歴書と 101 人のユーザの 1000 件のつぶやき集合との類似度の分布を示したものである．■点は本人のつぶやきとの類似度を表している．箱ヒゲ図は他 100 人のユーザのつぶやきとの類似度を表している．図 4.19 から学生 10 人のうち学生 1・学生 5・学生 6・学生 7・学生 8・学生 10 の 6 人の類似度が箱ヒゲ図よりも高くなるため，本人のつぶやき集合を特定できることが分かる．また，学生 2・学生 3・学生 9 の 3 人においては，101 人中上位 4 人に入る．このことから，サンプルデータの学生 10 人のうち 6 人を特定でき，9 人は 101 人中上位 4 人までに絞り込める．

表 4.3　各学生の発言の特定・絞り込み精度

| 履歴書の人物 | 上位１人 | 上位５人 | 上位１０人 |
|---|---|---|---|
| 学生１ | 7 | １０ | １０ |
| 学生２ | 2 | 5 | 6 |
| 学生３ | 1 | 4 | 4 |
| 学生４ | 0 | 1 | 2 |
| 学生５ | 8 | 8 | 8 |
| 学生６ | 6 | 7 | 9 |
| 学生７ | 0 | 4 | 8 |
| 学生８ | 3 | 6 | 6 |
| 学生９ | 2 | 4 | 4 |
| 学生１０ | 8 | １０ | １０ |
| （総計） | ３７／１００ | ５９／１００ | ６７／１００ |

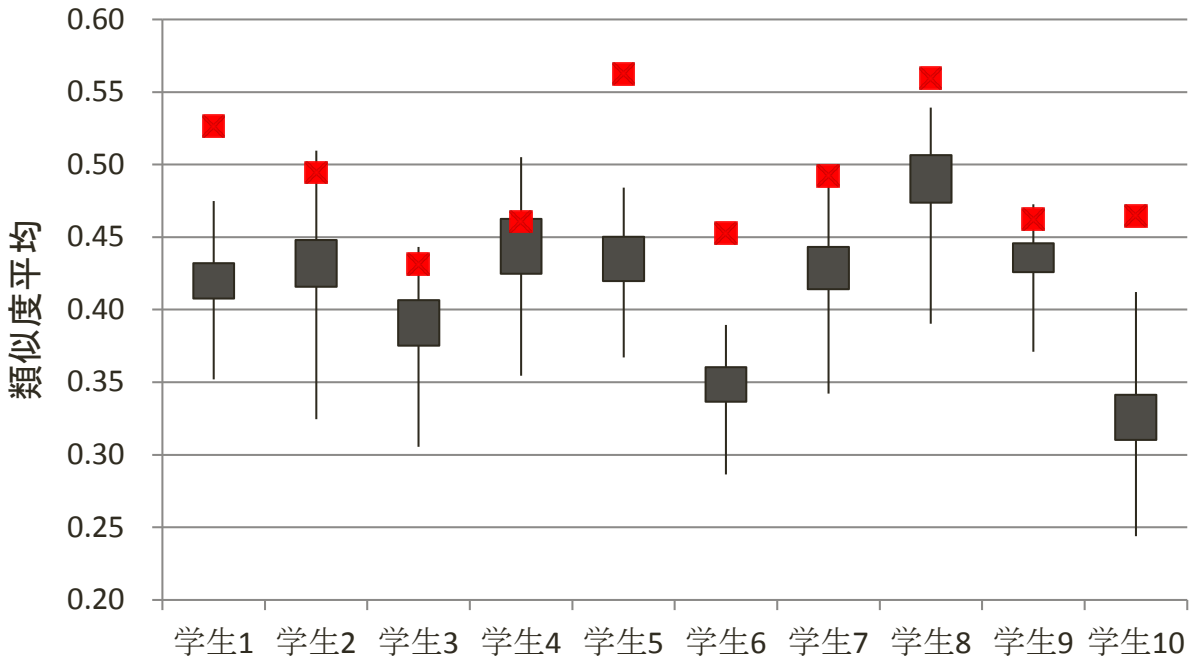


図 4.19　各学生の履歴書と 101 人のつぶやきとの類似度

### 4.5.4　考察

実験 1 の結果と実験 2 の結果を比べると，実験 2 では本人特定精度が低いことが分かる．この原因として，5.3.1 項で述べたように，履歴書の項目数が少ないことが考えられる．また，一番精度が低かった学生 4 のつぶやきの内容を手作業で確認したところ，履歴書情報に関するつぶやきは 10 件程度であった．その他は「でかける」「おはようございます」「起きた」など日常的な発言が非常に多く見られた．提案手法は履歴書情報とつぶやきとの類似度に基づくため，履歴書と関連性のないつぶやきが多い人物の場合は，つぶやきの特定が難しいことが明らかになった．

また，学生のつぶやきには，口語や現代語，ネットスラングが多いことが考えられる．これらの語句は，形態素解析用の辞書に登録されていない場合がある．現在用いている形態素解析器 Mecab は辞書にない単語（未知語）を全て「名詞」と判定するため，これらの単語が類似度算出のノイズとなり，精度の低下につながっていることが考えられる．

以上述べた精度低下の原因については，今後より詳細な分析を行い，方式の改良につなげる必要がある．

### 4.5.5　評価のまとめ

社会人 2 人，学生 10 人の合計 12 人の履歴書情報と，12 人および他の 100 人の各 1000 件のつぶやきを用い，4.4.2 節の改良方式が 101 人分のつぶやきの中から履歴書の人物のつぶやきを特定できるか実験を行った．ただし，社会人 Bob のみは，収集できたつぶやきが 700 件であったため，700 件を用いた．つまり，実験に用いたつぶやきは，社会人 1000 件および 700 件，学生 1000 件×10 人，他人 1000 件×100 人の合計 111700 件である．

まず，同一人物の 1000 件のつぶやきを 100 件毎に分割して 10 個のつぶやきセットとし，セット単位で特定を試みた．社会人 Adam のつぶやきセット 1 と他人 100 人のつぶやきセット 1（合計 101 セット）について，Adam の履歴書との類似度を算出したところ，Adam 本人のつぶやきセット 1 を類似度最大として正しく特定することができた．同様に，Adam のつぶやきセット 2 と他人 100 人のつぶやきセット 2 の中から，Adam 本人のつぶやきセット 2 を正しく特定することができた．Adam のつぶやきセット 3 から 10 についても，各々101 人の中から，類似度最大として正しく特定することができた．社会人 Bob については，7 個のつぶやきセットのうち 2 から 7 の 6 セットについて正しく特定することができた．以上から，社会人 2 人についての特定精度は，True Positive Rate (TPR) = 0.941, True Negative Rate (TNR) = 0.999 となった．なお，TPR および TNR の定義は以下のとおりである．

$$TPR = \frac{\text{真に本人であった件数}}{\text{本人とみなした件数}} = \frac{16\text{セット}}{17\text{セット}} \quad \cdots \text{ 式4.13}$$

$$TNR = \frac{\text{真に他人であった件数}}{\text{他人とみなした件数}} = \frac{1699\text{セット}}{1700\text{セット}} \quad \cdots \text{ 式4.14}$$

学生 10 人について同様の評価をおこなったところ，のべ 100 個のつぶやきセット（10 人×10 セット）のうち 37 個においてのみ，101 セットの中から類似度最大として特定できた．

そこで，つぶやきセットのサイズを 100 件ではなく 1000 件として，同様の評価を行った．なお，Bob についてのみ，つぶやきセットのサイズを 700 件とした．この場合，各人が 1 つのつぶやきセットで評価される．評価の結果，学生 10 人のうち 6 人については，101 人分のつぶやきセットの中から本人のつぶやきセットが類似度最大として特定できた．また，残り 4 人のうち 3 人については，本人のつぶやきセットが類似度の上位 4 セットの一つになった．社会人 2 人については， 2 人とも本人のつぶやきセットを特定できた．この評価結果をまとめると，1000 件のつぶやきを評価の単位をする場合には，66.7%の被験者（8 人／12 人）について 101 人のユーザの中からつぶやきを特定することができ，また 91.7%の被験者（11 人／12 人）について 101 人のユーザの中から 4 人のつぶやきにまで絞り込むことができた．

## 4.6 社会的リスクに関する考察

4.5 節の評価により，履歴書との照合を通じて，履歴書本人のつぶやきを特定できる可能性が明らかになった．ほとんどの組織は構成員の履歴書を保有している．また，構成員でなくとも就職希望者等の履歴書を入手する機会は多い．さらに，大学等では教員の履歴書相当の情報を Web 上で公開している場合もある．したがって，ソーシャルメディアに公開されたつぶやきと異種情報である履歴書との照合は，現実的な社会的リスクをもたらすと言える．

4.4 節で述べたように，つぶやきの中で，履歴書の単語をほとんど開示しなくても，Web 検索を通じて照合が可能となる．これは，図 4.20 に示すように，つぶやき中の単語と履歴書中の単語の両方を含む Web ページが存在し，検索エンジンによって利用できるからである．これらの Web ページの情報の介在により，つぶやきの投稿者が履歴書の単語を使わないといった通常の注意を払ったとしても，履歴書と照合され，個人が特定される．そのためプライバシー侵害のリスクは大きいと考えられる．

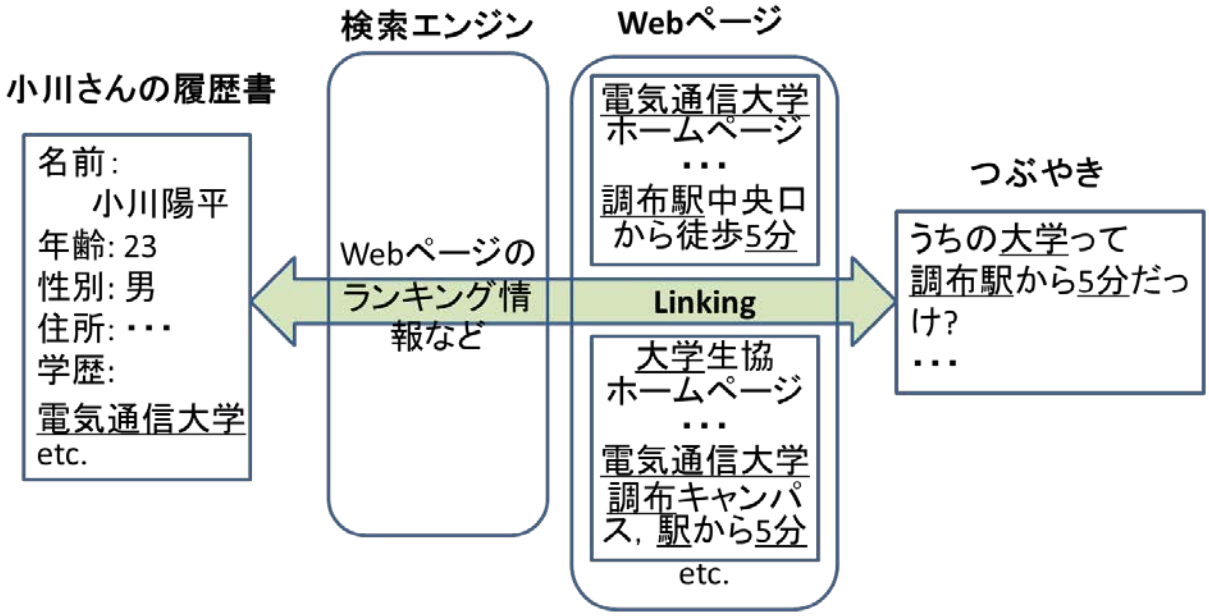


図 4.20　履歴書とつぶやきの照合における Web 情報の介在

## 4.7　まとめと今後の課題

同一人物に関する複数の情報の照合を通じて，プライバシー侵害につながる情報の推定が可能になることを明らかにするために，個人の履歴書情報とソーシャルネットワーク上に開示された発言内容との照合システムを提案した．提案システムは，ソーシャルネットワーク上の多数の発言について，履歴書との間で文書としての類似度を算出し，類似度が最大の発言を履歴書の人物の発言として特定する．

発言と履歴書の類似度を算出する際に，発言内で履歴書の単語が直接記載されるとは限らず，単語の組合せや略称，間接的な表現が用いられることが問題となる．そこで，3 章で述べた DCNL の想起検知手法を拡張し，履歴書を NG ワード集として，発言内の多様な表現と履歴書の名詞とのつながりを Web 検索によって検知する手法を提案した. 提案したシステムおよび手法を実装し，12 人の履歴書の各々について，履歴書の人物を含む 101 人の Twitter ユーザのつぶやきの中から，履歴書の人物のつぶやきを特定できるか評価した．同一人物の 1000 件のつぶやきを 1 セットとし，履歴書の人物のつぶやき 1 セットおよび他の 100 人のつぶやき各 1 セット，合計 101 セットのつぶやきから，履歴書の人物のつぶやきを特定したところ，8 人の履歴書の人物については，当該人物のつぶやきセットを特定できた．3 人の人物については 101 人のつぶやきセットから 4 人のつぶやきセットに絞り込むことができた．

今後の課題としては，以下が挙げられる．

(1)より大規模な評価実験を通じて，提案手法の有効性を確認する．本研究では，電気通信大学の 3 年生を主たる被験者として評価した. そのため, つぶやきに講義名や実験など，履歴書との関連の高いキーワードが多く含まれ，つぶやきと履歴書との照合が比較的容易

であった可能性がある．そこで，被験者の属性を変えて評価実験を行う．

(2)類似度算出アルゴリズムの中間結果を用いて，ソーシャルメディアの発言のなかの本人特定につながる表現を検知し，ソーシャルメディアユーザに警告する技術を開発する．

# 第5章

# 結論

## 5.1　まとめ

ソーシャルメディアは，世界中で 10 億人以上の人々が多様な目的のために利用し，それに合わせたメディア機能の拡張も継続的に行われており，現代社会のコミュニケーション基盤としての地位を確立しつつある．しかしながら，その一方で，ソーシャルメディアを通じたプライバシー情報や機密情報の漏洩，それに伴う犯罪などの反社会的な行為や言論の自由への懸念といった問題が生じている．本論文では，ソーシャルメディアにおけるプライバシー情報の漏洩問題および従来の対策を分析し，プライバシーを保護する新しい技術の提案と実装評価を行った．また，ソーシャルメディアにおけるプライバシーリスクとして，従来明らかにされていなかった異種情報の照合によるプライバシー情報の推定を取り上げ，そのような照合が技術的に可能であることを明らかにした．

本論文の 1 章では，ソーシャルメディアの発展とプライバシー問題を概観して，研究背景を明らかにし，コミュニケーションとの両立が可能な新しいプライバシー保護技術の確立および異種情報の照合によるプライバシーリスクの明確化という 2 つの研究目的を設定した．

2 章では，ソーシャルメディアにおける従来のプライバシー保護技術として，伝統的なセキュリティ技術であるアクセス制御と，ソーシャルネットワークサービスで利用されている公開範囲設定の 2 つを取り上げて分析し，以下を明らかにした．

- プライバシー情報およびその表現の多様性のためにアクセス制御ルールおよび公開範囲の適切な設定が困難であり，プライバシー情報の漏洩を充分に防止できない．
- プライバシー情報は個々のユーザに依存するため，アクセス制御ルールや公開範囲の設定をユーザが行う必要があり，ユーザへの負担となる．
- 公開範囲の設定においてプライバシー保護を優先すると，コンテンツ（日記やつぶやき）の公開毎に範囲を厳密に考え設定する必要がある．
- 公開範囲の設定はコンテンツ単位であるため，コンテンツ内に 1 か所でもプライバシー情報の漏洩個所があると，コンテンツ全体を非公開にする必要がある．

以上を総括し，従来のプライバシー保護技術は，コミュニケーションとプライバシーの二者択一を迫る技術であり，ソーシャルメディアの本来の目的である「コミュニケーションの楽しみ」を損なう可能性があることを明らかにした．これらの分析に基づき，コミュニケーションとの両立が可能な新しい技術として，自然言語情報の開示制御技術 DCNL（Disclosure Control of Natural Language Information）を提案した．DCNL は，ユーザの投稿文を検査して，プライバシー情報の漏洩をもたらす個所を単語単位で検知し，ユーザへの警告あるいは該当箇所の言い換えを行う技術である．DCNL のシステム要件として，以下を示した．

- プライバシー情報の検知ルールを定義するといったユーザへの負担を最小化する．
- プライバシー情報を表す単語の直接的な記載だけでなく，単語の組合せや間接的な意味によるプライバシー情報の漏洩も検知する．
- 単語単位の検知および言い換えにより，コンテンツの大部分は公開可能とし，コミュニケーションかプライバシーかの二者択一を避ける．

以上の要件を満たすシステム構成案として，ソーシャルメディアとユーザの間に介在し，プライバシー情報検知処理と漏洩防止処理の 2 つのメイン処理から成り，自然言語処理，プライバシー情報定義，外部情報利用の 3 つの共通モジュールを用いる形態を示した．

3 章では，DCNL のうちプライバシー情報検知処理について検討した．50 代の社会人ユーザの mixi における日記 7047 文を分析し，漏洩しているプライバシー情報を分類すると共に，その表現形態として直接的な記載と間接的な示唆の 2 種類の存在を明らかにした．間接的な示唆には，単語の組合せによる表現，外部知識との組み合わせによりプライバシー情報に到達する表現があった．この分析に基づいて，プライバシー情報をシステム内に定義し，投稿文の語句と照合することで情報漏洩を検知することとし，以下の技術課題と基本方針を設けた．

- ユーザに手間をかけないプライバシー情報の定義を課題とし，NG ワードと呼ばれる一つの単語によってプライバシー情報を定義する．
- 間接的な示唆によるプライバシー情報の漏洩を検知するために，投稿文中の単語と NG ワードの直接的な照合だけでなく，投稿文中の単語およびその組み合わせをキーワードとして Web 検索を行い，検索結果に NG ワードが含まれる頻度によって情報漏洩を検知する．

以上の方針に基づいて，プライバシー情報の直接表現を検知する直接検知と，単語の組合せや間接的意味を検知する想起検知から成るアルゴリズムを設計，実装した．11 人の学生のつぶやき各 1000 件（合計 11,000 件）および上述した 50 代の社会人ユーザの mixi の日記をサンプルとし，通学・通勤先および職業情報の検知精度の評価を行った結果，再

現率（漏洩した件数のうちシステムが検知した率）は平均して 90%程度であり，情報漏洩を高い確率で検知できた．残り 10%の False Negative については，キーワードの抽出に失敗したため検知できなかった．一方，適合率（システムが検知した件数のうち真に漏洩している率）が平均して 30%程度と低く，ユーザに不必要なアラートを発することが明らかになった．しかしながら，文書ソフトの校正機能のように，不必要なアラートであってもユーザに煩わしさを感じさせることの少ないインターフェースの提供は可能と考えられる．一方，システムがプライバシー情報の漏洩を検知した文章を人間が読み返した結果，これまで人間が気付かなかった情報漏洩に気づいたケースが，情報漏洩した文章数の 10%程度に達した．このことから，提案システムにより人間の知識や能力を超えた検知が可能となり，人間の注意力を補えることが明らかになった．

4 章では，異種情報の照合によるプライバシーリスクについて検討した．Twitter のつぶやきと個人の履歴書との照合を通じて，多数のつぶやきの中から特定個人のつぶやきを抽出する技術を検討した．つぶやきの実例を調査した結果，履歴書の名詞が直接表現されるとは限らず，単語の組合せや略称によって表現されることが多かった．また，調布駅付近の大学は電気通信大学に限られるといった，外部知識との組み合わせによって履歴書の単語に到達できる表現も用いられる．そこで，つぶやきの多様な表現と履歴書の名詞とのつながりを検知するために，履歴書を NG ワード集とみなし，3 章で提案した想起検知アルゴリズムを拡張して用いた．学生 10 名，社会人 2 名の履歴書とつぶやきをサンプルとして，履歴書の人物を含む 101 人の Twitter ユーザのつぶやきから，履歴書の人物のつぶやきを特定できるか評価した．同一人物の 1000 件のつぶやきを 1 セットとし，履歴書の人物のつぶやき 1 セットおよび他の 100 人のつぶやき各 1 セット, 合計 101 セット（101,000 件）のつぶやきから，履歴書の人物のつぶやきを特定したところ，8 人の履歴書の人物については，当該人物のつぶやきセットを特定できた．3 人の人物については 101 人のつぶやきセットから 4 人のつぶやきセットに絞り込むことができた．これにより，たとえば就職希望者や，社員などを注目人物とし，その履歴書を用いることで，当該人物のつぶやきを特定できることを示し，異種情報の照合による社会的リスクの存在を明らかにした．

以上の研究を通じて，プライバシー情報という曖昧性，個人性の大きい情報を検知するための計算モデルとして，個人毎のプライバシー情報を NG ワードによって表現し，Web 検索を通じてプライバシー情報の多様な表現と NG ワードを照合するモデルを提案し，その有効性を明らかにした．また，この計算モデルがプライバシーの保護と攻撃の両面に利用できることを明らかにした．

## 5.2　今後の課題

DCNL に関しては，以下の課題があげられる．

(1) 通学・通勤先および職業だけでなく，様々なプライバシー情報の検知精度を評価する．
(2) 異種情報の照合で行った履歴書の利用などを参考にして，NG ワードの適切な設定手法を検討する．
(3) 形態素解析を改良し，キーワード抽出の精度向上を行う．
(4) ユーザに煩わしさを感じさせないインターフェースを検討する．
(5) プライバシー情報の漏洩が検知された個所の言い換え処理を検討し，ユーザ間の関係に基づいて，適切な言い換え候補を提示できるようにする．
(6) 以上を通じて DCNL 全体を実装し，ユーザビリティ評価を行う．

異種情報の照合については，今回は電気通信大学の学生を主たる被験者としたが，被験者の属性を変えて評価実験を行う．また，今回は多数の投稿のなかから特定人物の履歴書に合致するものを選定したが，次は，多数の履歴書のなかから特定の投稿に合致するものを選定する技術を検討する．履歴書を選定することは人物を特定することと同義であることから，この技術は，履歴書を保有する機関が，特定のつぶやきの発言者を特定できることを意味し，プライバシーおよび言論の自由に関してより大きなリスクを明らかにすると考える．

本研究では，プライバシーの保護技術である DCNL と，攻撃技術である履歴書照合技術の両面から，プライバシー技術を検討した．DCNL における Web 検索を用いたプライバシー漏洩検知手法（想起検知）を発展させて，履歴書照合技術におけるつぶやきとの照合手法を開発した．今後は，履歴書照合技術を発展させて，DCNL の想起検知の向上に用いたい．このように，保護技術と攻撃技術を相まって発展させていきたい．

# 謝辞

本研究を遂行し，学位論文としてまとめるにあたり，主任指導教員として終始多大なるご指導とご教示を頂いた吉浦裕教授，および副指導教員としてご指導とご教示を頂いた兼子正勝教授，坂本真樹准教授に心より感謝の意を表します．また、博士論文の審査委員として，御指導いただいた市川晴久教授，高橋裕樹准教授に深く感謝申し上げます．様々なご指導ご助言を頂きました市野将嗣助教，そして，吉浦研の研究室の皆様に深く感謝申し上げます．

最後に，社会人として働きながらの学位取得にご理解いただき激励くださった NTT サービスエボリューション研究所の職場の皆様，そして，子育てをしながらの学位取得を温かく見守り，時に我慢し，辛抱強く支援してくれた娘 凜，夫，両親に深い感謝の意を表して謝辞といたします．

# 参考文献


[1] Facebook Reports First Quarter 2014 Results , http://investor.fb.com/releasedetail.cfm?ReleaseID=842071

[2] Twitter ANNUAL REPORT 2013, https://materials.proxyvote.com/Approved/90184L/20140328/AR_202076/pubData/source/Twitter,%20Inc%202013%20Annual%20Report.pdf

[3] LINE 株式会社 広告事業グループ 広告事業部, LINE 2014 年 4-9 月媒体資料 http://linecorp.com/ads/pdf/0E8613FE-9926-11E3-950E-80FF7B512F31

[4] R. Gross, and A. Acquisti, “Information Reve-lation and Privacy in Online Social Networks,” in Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, 2005,pp.71–80.

[5] Acquisti, A., Gross, R.: Imagined communities: awareness, information sharing, and privacy on the Facebook, In: Proc. Workshop on Privacy Enhancing Technologies (PET), LNCS 4258, Cambridge, pp.36-58 (2006).

[6] Baatarjav, E., Ram, D., Phithakkitnukoon, S.: Privacy management for Facebook, In: Proc. 4th International Conference on Information Systems Security, pp. 273-286, Hyderabad (2008)

[7] Viegas, F.: Bloggers' expectations of privacy and accountability: an initial survey, Journal of Computer-Mediated Communication, 10(3), article 12 (2005)

[8] Meeder, B., Tam, J., Kelly, P.G., Cranor L.F., RT@ IWantPrivacy: widespread violation of privacy settings in the Twitter social network, In: Proc. Web 2.0 Privacy and Security Workshop, Oakland (2010)

[9] Gurses, S., Rizk, R., Gunther, O.: Privacy design in online social networks: learning from privacy breaches and community feedback, In: Proc. 29th International Conference on Information Systems, pp.1-10, Paris (2008).

[10] Novak, J., Raghavan, P. Tomkins, A.: Anti-aliasing on the Web, In: Proc. 13th International World Wide Web Conference (WWW2004), pp.30-39, New York (2004)

[11] Lam, I., Chen, K., Chen, L.: Involuntary information leakage in social network services, In: Proc. the 3rd International Workshop on Security (IWSEC), LNCS 5312, pp.167--183, Takamatsu (2008)

[12] Narayanan, A., Shmatikov, V.: De-anonymizing social networks, In: Proc. 30th IEEE Security & Privacy, pp.173-187, Oakland (2009)

[13] O. Goga, et al., “On Exploiting Innocuous User Activity forCorrelating Accounts AcrossSocial Network Sites,” ICSI Technical Reports -University of Berkeley, 2012.

[14] TWITTER ROBBERY OF ARIZONA MAN COULD FREAK SOME FOLKS OUT http://ptgtravel.blogspot.jp/2009/06/twitter-robbery-of-arizona-man-could.html

[15] SNS の「友達」急変　サイバーストーカー増殖 http://www.nikkei.com/article/DGXNASDG02002_S3A400C1CR0000/

[16] WeKnowYourHouse.com, http://weknowyourhouse.com/

[17] PleaseRobMe, http://pleaserobme.com/

[18] What's your status? A survey carried out by UK home security experts Friedland has revealed social media is being put to use by today’s home burglars. https://www.friedland.co.uk/en-GB/News/Pages/Whats-your-status.aspx

[19]インプレスジャパン, ”インターネット白書 2011,”インプレスコミュニケーションズ, p196, 2011.

[20]Ross Anderson, “Security Engineering”, John Wesley & Sons, 2001.

[21]Y. Liu, K.P. Gummadi, B. Krishnamurthy, and A. Misolve, “Analyzing Facebook Privacy Set-tings: User Expectations vs. Reality,” in Pro-ceedings of the 2011 ACM SIGCOMM Con-ference on Internet Measurement Conference, 2011, pp.61–70.

[22]宮川大輔が間違えて携帯番号公開「死ぬほど電話かかってきた」
http://www.j-cast.com/tv/2011/10/18110361.html

[23]「インターネットサービス利用時の情報公開範囲の設定に注意！」
http://www.ipa.go.jp/security/txt/2013/10outline.html

[24]ガイアックス、『Facebook ユーザーの時間・シチュエーションによる利用動向調査』を実施
http://gaiax-socialmedialab.jp/press/003

[25]ソーシャルネットワーキングサービス “mixi,” ,
http://mixi.jp/

[26]Google,
http://www.google.co.jp/

[27]青少年インターネット環境整備法等について
http://www8.cao.go.jp/youth/youth-harm/seibi_law/

[28]フィルタリングの導入促進
http://www.it-anshin.go.jp/policy/filter.html

[29]Warren, S, Brandeis, L.: The Right to Privacy. Harvard Law Review, Cambridge, Vol.4, No.5 (1890)

[30]Altman, I.: The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding, Brooks/Cole Publishing Company, 1975.

[31]Schoemean, F.: Philosophical Dimensions of Privacy, Cambridge University Press, 1984.

[32]NEC 総研: 現代人のプライバシー, NEC 総研, 2005.

[33]Organization for Economic Co-Operation and Development: Recommendation of the Council Concerning Guide-Lines Governing the Protection of Privacy and Transborder Flows of Personal Data, 1980.

[34]横浜地判平成元・五・二三 判タ七〇〇号一四四頁 判時一三一九号六七頁

[35]iProspect；Search Engine User Behavior Study(2006)

[36]Barnard J.Jansen, Amanda Spink, and Tefko.Saracevic. Real users, and real needs:A studyand analysis of usr queries on the web. Information Processing and Management, Vol. 36, No. 2, pp. 207-227, 2000

[37]工藤拓,”MeCab: Yet Another Part-of-Speech and Mor-phological Analyzer.”, 2002

[38]Yahoo!デベロッパーネットワーク テキスト解析,
http://developer.yahoo.co.jp/webapi/jlp/

[39]MeCab - Browse /mecab-ipadic at SourceForge.net
http://sourceforge.net/projects/mecab/files/mecab-ipadic/

[40]中川裕志, 森辰則, “専門用語（キーワード）自動抽出システム,”
http://gensen.dl.itc.u-tokyo.ac.jp/

[41]Wikipedia, “データベースダウンロード,”
http://dumps.wikimedia.org/jawiki/

[42]はてなキーワード一覧ファイル,
http://developer.hatena.ne.jp/ja/documents/keyword/misc/catalog

[43]Yahoo!デベロッパーネットワーク - 検索 - ウェブ検索 Web 検索 API
http://developer.yahoo.co.jp/webapi/search/websearch/v2/websearch.html

[44]G. Kótyuk and L. Buttyan, “A Machine Learning Based Approach for Predicting Un-disclosed Attributes in Social Networks,” in Proceedings of IEEE 4th International Work-shop on Security and Social Networking, 2012, pp.361–366.

[45]L. Banks and S. Wu, “All Friends Are Not Created Equal: An Interaction Intensity Based Approach to Privacy in Online Social Net-works,” in Proceedings of 12th IEEE Interna-tional Conference on Computational Science and Engineering, Vol.4, 2009, pp.970-974.

[46]H. Mao, X. Shuai, and A. Kapadia, “Loose Tweets: An Analysis of Privacy Leaks on Twit-ter,” in Proceedings of the 10th ACM Workshop on Privacy in the Electronic Society, 2011,

[47]町田史門, 嶋田茂, 越前功, “SNS 上のプライバシーセンシティブ情報の漏洩検知に基づく公開範囲の設定方式,” 情報処理学会コンピュータセキュリティシンポジウム, 2013, pp566-573

[48]町田史門,梶山朋子,嶋田茂,”センシティブデータの漏洩検知による適応的な公開範囲設定システムのプロトタイプ実装,” 信学技報 113(480), 51-56, 2014-03-07

[49]Ngoc, T.H. ; Fac. of Inf. Technol., Univ. of Sci., Ho Chi Minh City, Vietnam ; Echizen, I. ; Komei, K. ; Yoshiura, H. “New Approach to Quantification of Privacy on Social Network Sites”, Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference, 2010

[50]曽根原愛理,白鷹靖子,小舘亮之,並河大地,南裕也,下村道夫,”ペルソナ情報を用いた画像共有サービスにおける開示制御技術の提案,” 信学技報 LOIS, ライフインテリジェンスとオフィス情報システム, 2012

[51]A. Narayanan,andV. Shmatikov, ”Robust De-anonymization of Large Sparse Datasets, In Proc. of the 29th IEEE Symposium on Se-curity and Privacy,” pp.111-125 (2008)

[52]A. Narayanan, et al., “On the Feasibility of Internet-Scale Author Identification,” in Pro-ceedings of the 33rd IEEE Symposium onSe-curity and Privacy, 2012, pp.300–314.

[53]Polakis, et al., “Using Social Networks to Harvest Email Addresses,” in Proceedings of the 9th ACM Workshop on Privacy in Electonic Society, 2010, pp.11–20.

[54]A. Acquisti, R. Gross, and F. Stutzman, “Faces of Facebook: Privacy in the Age of Augmented Reality,” in BlackHatUSA, 2011.

[55]CareerBuilder.com, “One-in-Five Employers Use Social Networking Sites to Reasearch Job Candidates, CareerBuilder.com Survey Finds,”
http://www.careerbuilder.com/share/aboutus/pressreleas-esdetail.aspx?id=pr459&sd=9%2F10%2F2008&ed=12%2F31%2F2008.

[56]Reuters, “California Schools, Employers Banned from Social Media Snooping,”
http://www.reuters.com/article/2012/09/27/us-usacalifonia-privacy-idUSBRE88Q1UI20120927.

[57]G. Salton,et al., “A Vector Space Model for Automatic Indexing,” Communications of the ACM, Vol.18, No.11, pp. 613-620(1975)

[58]K. Jones, “A Statistical Interpretation of Term Specificity and its Application in

Re-trieval," Journal of Documentation, Vol.28, No.1, pp.11-21(1972)
[59] The size of the World Wide Web (The Internet),
http://www.worldwidewebsize.com/

# 関連論文の印刷公表の方法および時期

- 学術論文

1. 全著者名：Haruno Kataoka, Natsuki Watanabe, Keiko Mizutani, and Hiroshi Yoshiura

   論文題目：DCNL: Disclosure Control of Natural Language Information to Enable Secure and Enjoyable E-Communications

   印刷公表の方法および時期：International Journal of u- and e- Service, Science and Technology, Vol. 3, No. 1, pp.1-10, Mar. 2010.

   （2 章，3 章に関連）

2. 全著者名：片岡春乃, 奥野智孝, 木村聡一, 内海彰, 吉浦裕

   論文題目：ソーシャルネットワークから注目者の発言を特定するシステムの提案と予備評価

   印刷公表の方法および時期：日本セキュリティ・マネジメント学会誌, 27 巻 3 号, pp. 13-28, Jan. 2014.

   （4 章に関連）

- 国際学会発表論文

1. 全著者名：Haruno Kataoka, Natsuki Watanabe, Keiko Mizutani, Hiroshi Yoshiura

   論文題目：DCNL: Disclosure Control of Natural Language Information to Enable Secure and Enjoyable E-Communications

   印刷公表の方法および時期：Proceedings of 2nd International Conference on U- and E-Service, Science and Technology, Springer CCIS 62, pp.131-140, Dec. 2009.

   （2 章，3 章に関連）

2. 全著者名：Haruno Kataoka, Yohei Ogawa, Isao Echizen, Tetsuji Kuboyama, Hiroshi Yoshiura

   論文題目：Effects of External Information on Anonymity and Role of Transparency with Example of Social Network De-Anonymisation

   印刷公表の方法および時期：Proceedings of 4th International Workshop on Resilience and IT-Risk in Social Infrastructures, pp.461-467, Sep. 2014.

   （4 章に関連）

- 国内口頭発表

1. 全著者名：片岡春乃, 内海彰, 広瀬友紀, 吉浦裕
   論文題目：意味と面白さを維持する自然言語情報の開示制御技術の提案―SNS のプライバシー保護への試適用―
   印刷公表の方法および時期：情報処理学会 第 36 回コンピュータセキュリティ研究会報告, pp.321-326, Mar. 2007.
   （2 章に関連）
2. 全著者名：片岡春乃, 渡辺夏樹, 水谷桂子, 吉浦裕
   論文題目：自然言語情報の開示制御技術 DCNL の実現に向けて ―プライバシー情報検知手法―
   印刷公表の方法および時期：情報処理学会 第 40 回コンピュータセキュリティ研究会報告, pp.237- 242, Mar. 2008.
   （3 章に関連）
3. 全著者名：吉浦裕, 片岡春乃, 中山心太
   論文題目：多様化するメディア環境に適応するヒューマンコミュニケーションセキュリティの構想
   印刷公表の方法および時期：第 38 回コンピュータセキュリティ研究会報告, pp.125-131, Jul. 2007.
   （2 章に関連）
4. 全著者名：片岡春乃, 小川陽平, 吉浦裕
   論文題目：ソーシャルネットワークのプライバシーと個人の特定, ―つぶやきと履歴書の照合を例として―
   印刷公表の方法及び時期:電気学会 電子・情報・システム部門大会論文集, pp.332-337, Sep. 2014..
   （4 章に関連）

# その他の研究業績

- 学術論文

1. 全著者名：大塚雅博，片岡春乃，末田欣子，下村道夫，淺谷耕一，水野修
   論文題目：共同体験型コミュニケーションサービスの受容性
   印刷公表の方法及び時期：電子情報通信学会論文誌 B Vol.J95-B No.2 pp.366-370, Feb.2012.

- 国際学会発表論文

1. 全著者名：Haruno Kataoka, Masashi Toyama, Yoshiko Sueda, Osamu Mizuno and Kenji Takahashi
   論文題目：Web Contents Collaborative Method in Call-to-Web Session Linkage System
   印刷公表の方法及び時期：7Th Annual IEEE Consumer Communications & Networking Conference, Jan. 2010.
2. 全著者名：Haruno Kataoka, Masashi Toyama, Yoshiko Sueda, Osamu Mizuno and Kenji Takahashi
   論文題目：Demonstration of Web Contents Collaborative System for Call Parties
   印刷公表の方法及び時期：7Th Annual IEEE Consumer Communications & Networking Conference, Jan. 2010.
3. 全著者名：Haruno Kataoka, Daichi Namikawa, Hiroya Minami, Michio Shimomura and Naoki Uchida,
   論文題目：SightFinder: Enhanced Videophone Service Utilizing Media Processing,
   印刷公表の方法及び時期：Intelligence in Next Generation Networks (ICIN 2012), Oct. 2012.
4. 全著者名：Hiroaki Nishihata, Masahiro Otsuka, Koichi Asatani, Osamu Mizuno, Haruno Kataoka, Michio Shimomura
   論文題目：Proposed presence system for safety confirmation
   印刷公表の方法及び時期：Intelligence in Next Generation Networks (ICIN 2012), Oct. 2012.

- 国内口頭発表

1. 全著者名：片岡春乃 ,末田欣子 ,外山将司, 村上幸司, 水野修
   論文題目：通話セッション情報に基づく Web 共有制御方式の検討

印刷公表の方法及び時期：電子情報通信学会 2009 総合大会, Mar. 2009.

2. 全著者名：片岡春乃，外山将司・末田欣子，水野修，高橋健司
   論文題目：SIP-Web 間セッション連携におけるコンテンツ共有制御方式の検討
   印刷公表の方法及び時期：電子情報通信学会 情報ネットワーク研究会, Sep. 2009.
3. 全著者名：片岡春乃, 車谷 駿介, 外山 将司, 末田 欣子, 千葉 一深, 西永 誠司, 下村 道夫
   論文題目：通話者間 Web 共有制御システムにおけるユーザビリティテスト
   印刷公表の方法及び時期：HIS2010, Sep.2010.
4. 全著者名：外山将司, 片岡 春乃 , 車谷 駿介, 末田 欣子 , 下村 道夫
   論文題目：SIP-Web 間セッション連携システムにおける既存 Web API との連携方式の検討
   印刷公表の方法及び時期：電子情報通信学会 情報ネットワーク研究会, May. 2010.
5. 全著者名：車谷 駿介 , 外山 将司 , 片岡 春乃 , 末田 欣子 , 下村 道夫
   論文題目：多様な Web ブラウザ搭載機器での双方向通信実現に関する一検討
   印刷公表の方法及び時期：電子情報通信学会 2010 総合大会, Mar. 2010
6. 全著者名：赤星拓未, 末田欣子, 片岡春乃, 下村道夫, 小松尚久
   論文題目：認証サーバ選択方式による OpenID フィッシング対策の検討
   印刷公表の方法及び時期：暗号とセキュリティシンポジウム SCIS 2011, Jan. 2011
7. 全著者名：大塚 雅博 , 片岡 春乃 , 末田 欣子, 下村 道夫 , 淺谷 耕一 , 水野 修
   論文題目：共同体験型コミュニケーションサービスの受容性評価手法の提案
   印刷公表の方法及び時期：電子情報通信学会 コミュニケーションクオリティ研究会, Apr. 2011
8. 全著者名：大塚 雅博 , 片岡 春乃 , 末田 欣子, 下村 道夫 , 淺谷 耕一 , 水野 修
   論文題目：共同体験型コミュニケーションのサービス条件 : コンテンツへのユーザ期待度に対する評価
   印刷公表の方法及び時期：電子情報通信学会 2012 総合大会, Mar. 2012
9. 全著者名：清水 裕貴 , 大塚 雅博 , 片岡 春乃 , 下村 道夫 , 淺谷 耕一 , 水野 修
   論文題目：非接触型デバイスを用いた安否情報登録システム
   印刷公表の方法及び時期：電子情報通信学会 2012 総合大会, Mar. 2012
10. 全著者名：西畑 拓晃 , 大塚 雅博 , 片岡 春乃 , 下村 道夫 , 淺谷 耕一 , 水野 修
    論文題目：サービス条件の変化に適応するプレゼンスサービス提供方式
    印刷公表の方法及び時期：電子情報通信学会 2012 総合大会, Mar. 2012
11. 全著者名：西畑拓晃, 片岡春乃, 下村道夫, 水野修
    論文題目：プレゼンスサービスのためのスマートフォンを用いた情報入出力方式

印刷公表の方法及び時期:電子情報通信学会 ネットワークシステム研究会, Jan. 2013

12. 全著者名：伊藤一喜， 西畑拓晃， 片岡春乃， 下村道夫， 水野修
論文題目：見守られる人物のプレゼンス変化に適応した見守りシステム
印刷公表の方法及び時期：電子情報通信学会 2013 総合大会, Mar. 2013

- 特許

1. 特願 2009-111388
Web サービス制御システム及び Web サービス制御方法
片岡春乃，外山将司，村上幸司，末田欣子
2009.4.30（登録査定）
2. 特願 2009-198874
Web コンテンツ共有システム及び Web コンテンツ共有方法
片岡春乃，外山将司，末田欣子
2009.8.28
3. 特願 2010-093348
Web コンテンツ共有システム及び Web コンテンツ共有方法
片岡春乃，外山将司，末田欣子
2010.4.14
4. 特願 2010-190688
Web コンテンツ共有システム及び Web コンテンツ共有方法
片岡春乃，外山将司，末田欣子
2010.8.27（登録査定）
5. 特願 2011-134710
サービス提供システム、認証サーバ、サービス提供装置およびプログラム
片岡春乃，末田欣子，下村道夫，田村健範，赤星拓未，小松尚久
2011.6.17
6. 特願 2011-264840
サービス提供方法、システムサーバ、及びプログラム
片岡春乃，並河大地，南裕也，下村道夫
2011.12.2
7. 特願 2012-154708
情報提示システム及び情報提示サーバ
片岡春乃，南裕也，並河大地，下村道夫
2012.7.10

8. 特願 2012-185228
   システムサーバ及びシステムサーバの制御方法
   片岡春乃，西永誠司，下村道夫，水野修
   2012.8.24
9. 特願 2012-185808
   システムサーバ及びシステムサーバの制御方法
   片岡春乃，西永誠司，下村道夫，水野修
   2012.8.24
10. 特願 2013-001610
   システムサーバの制御方法及びシステムサーバ
   片岡春乃，西永誠司，下村道夫，水野修
   2013.1.9
11. 特願 2013-020478
   遠隔監視方法及び遠隔監視システム
   片岡春乃，西永誠司，下村道夫，水野修
   2013.2.5
12. 特願 2009-123166
   サービス提供システムおよびサービス提供方法
   村上幸司，千葉一深，末田欣子，外山将司，片岡春乃
   2009.5.21（登録査定）
13. 特願 2009-153900
   サービス提供システムおよびサービス提供方法
   村上幸司，末田欣子，外山将司，片岡春乃
   2009.6.29（登録査定）
14. 特願 2009-154096
   サービス提供システムおよびサービス提供方法
   村上幸司，末田欣子，外山将司，片岡春乃
   2009.6.29
15. 特願 2009-152800
   通信制御システム、通信制御方法および通信制御プログラム
   末田欣子，村上幸司，外山将司，片岡春乃
   2009.6.26（登録査定）
16. 特願 2009-240809
   アクセス制御システムおよびアクセス制御方法

外山将司，片岡春乃，末田欣子

2009.10.19（登録査定）

17. 特願 2010-096390

サービス提供装置、サービス提供方法およびサービス提供プログラム

外山将司，片岡春乃，末田欣子

2009.4.19（登録査定）

18. 特願 2010-228938

アクセス制御システム及びアクセス制御方法

末田欣子，外山将司，片岡春乃

2010.10.8

19. 特願 2011-063292

共有制御システム、制御方法、制御プログラム

外山将司，片岡春乃．末田欣子

2011.3.22（登録査定）

20. 特願 2011-148309

通信管理装置及び通信管理方法

佐藤啓之，南裕也，片岡春乃，並河大地

2011.7.4

21. 特願 2011-148304

音像定位制御システム、コミュニケーション用サーバ、多地点接続装置、及び音像定位制御方法

佐藤啓之，南裕也，片岡春乃，下村道夫，並河大地

2011.7.4

22. 特願 2012-028861

有用情報提示システム及び有用情報提示システムの制御方法

南裕也，片岡春乃，並河大地，荒川則泰，下村道夫

2012.2.13

23. 特願 2012-028823

有用情報提示システム及び有用情報提示システムの制御方法

南裕也，片岡春乃，並河大地，荒川則泰，下村道夫

2012.2.13

24. 特願 2012-028489

有用情報提示システム及び有用情報提示システムの制御方法

南裕也，片岡春乃，並河大地，荒川則泰，下村道夫

2012.2.13

- 表彰

1. 第 16 回 電子情報通信学会 情報ネットワーク研究賞
   “SIP-Web 間セッション連携におけるコンテンツ共有制御方式の検討”
   片岡春乃，外山将司，末田欣子，水野 修，高橋健司
   Mar.2010.
2. ICIN2012 Best Paper Award
   "SightFinder: Enhanced Videophone Service Utilizing Media Processing,"
   Haruno Kataoka, Daichi Namikawa, Hiroya Minami, Michio Shimomura and Naoki Uchida,
   Oct. 2012.