

自然言語情報の開示制御技術 DCNL の実現に向けて ープライバシー情報検知手法ー

片岡春乃* 渡辺夏樹† 水谷桂子† 吉浦裕†

* 電気通信大学大学院 電気通信学研究科 † 電気通信大学 電気通信学部

*† 〒182-8585 東京都調布市調布ヶ丘 1-5-1

概要 SNS, Blog 等を介した Web 上のコミュニケーションにおいて, プライバシー情報など個人のセンシティブな情報が漏洩し問題となっている. 我々は, 自然言語テキストの意味と面白さを維持しながら不適切な部分を変更するアクセス制御技術 DCNL を提案し, 研究を進めている. 本論文では DCNL の実現に向け, 特定ユーザの 2 年分の日記を分析し, 自然言語処理を用いたテキスト文章からのプライバシー情報の検知手法を検討する.

キーワード 情報漏洩, プライバシー, アクセス制御, 人工知能, 認知科学

Implementing Disclosure Control of Natural Language Information

Kataoka Haruno* Watanabe Natsuki† Mizutani Keiko† Yoshiura Hiroshi†

* Graduate school of Electro Communications, University of Electro-Communications

† Faculty school of Electro Communications, University of Electro-Communications

*† 1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585 Japan

Abstract In Web-based communication media, such as SNSs and blogs, sentences posted by users cause security problems, such as privacy leakage and abuse. To solve these problems, our early paper proposed the concept of Disclosure Control of Natural Language Information (DCNL) that modifies the users' sentences so that they are no longer harmful while maintaining their meaning and liveliness. This paper discusses the implementation aspects of DCNL. Diaries that a user has written on SNS for two years are analyzed to clarify technical requirements for DCNL, from which algorithms and data structures of DCNL are derived.

Keyword Privacy protection, Web security, Disclosure control, Access control, Natural language analysis

1. はじめに

近年, Web 上のコミュニケーションメディアとして Blog や SNS(Social Networking Services) が注目されている. これらのメディアは, ヒューマンコミュニケーションを活性化する一方, プライバシー情報の漏洩や誹謗中傷などの不適切な表現が問題になっている.

情報開示を制御する従来技術としてアクセス制御[3]がある. アクセス制御では, 対象となる個人情報や機密情報を洗い出して分類する. その上で, 事前にアクセス主体のクラスに応じたアクセス制御ルールを定義して利用する.

一方, ヒューマンコミュニケーションでは, 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が難しく, 加えてどの語句, 文章からどのような情報が洩れるのかは, 状況や相手によってそれぞれ異なる. そのため従来

のアクセス制御を用いた場合, 事前に定義されていない言葉が頻出すると予想される. それらに対して一律にアクセス許可すれば制御の意味が無くなり, 不許可にすると言葉が失われ, 共有したい話題を相手に伝えることが出来なくなる. 結果, コミュニケーションが断絶され, メディアの面白さが失われてしまう恐れがある.

我々は, Web 上のコミュニケーションに適した自然言語情報の開示制御 DCNL(Disclosure Control of Natural Language information)を提案[1][2]し研究している. これは, ユーザが十分配慮せずにメディアへ投稿した文章を自動的にチェックし, 必要な開示制御を行う. これにより, ユーザに手間をかけず複数の相手と 1 つの話題を共有可能となり, メディア特有の面白さを維持することができる.

また, 近年相次ぐ SNS, Blog を介した略取誘

拐等の事件を背景に、未成年携帯電話利用者を対象にした、SNSを含む有害サイトアクセス制限サービス（フィルタリングサービス）の義務化に向けた動きがある[10]。しかし、未成年者が従わない可能性、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ある。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。未成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。

本稿では DCNL の実現に向け、自然言語テキストからのプライバシー情報の検知手法を検討する。

2. DCNL(Disclosure Control of Natural Language information)

2.1. DCNL の要件

- (1) コミュニケーションにおける文章をチェックし、プライバシー情報を洩らす恐れのある全ての語句について洩れないように言い換えるか、削除する。
- (2) 間接的な意味、言葉の組み合わせから生じる意味を推測し、Web 検索を通じてプライバシー情報に到達する可能性も考慮する。

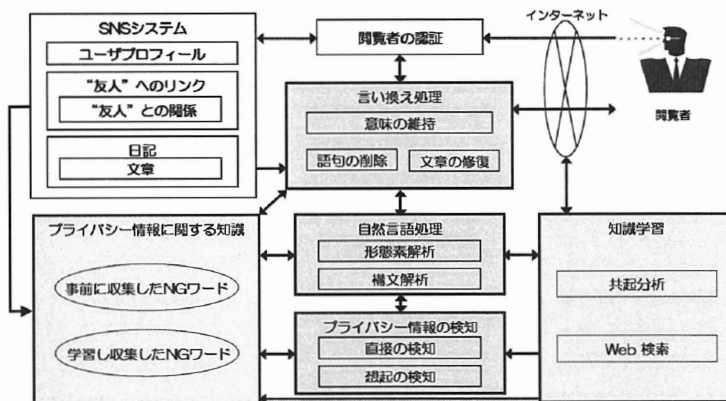


図 1 DCNL システム構成図

表 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な事柄	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

(3) 言い換えにあたって、文章の意味と面白さを維持する。

(4) ユーザへの負担を最小限に抑える。例えば、それぞれのクラスに対して言い換え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。そのため、言い換えルールは出来るだけ自動的に学習する。

2.2. DCNL のシステム構成

図 1 に示す。SNS にアクセスするとき、閲覧者が認証され、ユーザとの関係クラスが特定される。DCNL はクラスを基に原文を安全に言い換える。始めに、「自然言語処理」が、原文の語句を認識する。次に、「プライバシー情報の検知」が「プライバシー情報に関する知識」を用いて、語句がプライバシー情報を直接表すか、想起させるかを判断する。最後に「言い換え処理」が、文章の意味や面白さを維持しながら問題語句を言い換えるか削除し、最終的な文章を生成する。

本稿では、DCNL を実現するために、プライバシー情報の検知手法を検討する。

3. 実例文章の調査・分析

3.1. プライバシー情報

NEC 総研『現代人のプライバシー』[9]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。本論ではこの分類に準拠し、事例日記の調査を行った。

3.2. 特定ユーザに対する追跡調査

表1は、50代男性のmixi[4]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、分類した結果である。「直接」は、プライバシー情報が直接記述されている場合、「想起」は記述された情報をヒントとして、Web検索や年代早見表の参照によりプライバシー情報に到達する場合を指す。

調査対象149件中、76%(113件)の日記に509件のプライバシー情報が記述されていた。

3.3. 人間の思考過程の分析

3.2節において、プライバシー情報の記述がある7件と、他に20代女性ユーザ2名が投稿し、プライバシー情報の記述がある8件を加えた計15件の日記を対象に、人間が文章からプライバシー情報を読み解く思考過程を分析した。特に、図2に示す2件の日記について述べる。

日記1

人間は、「会社」を勤務先に関わる語として、「初台」を渋谷区の地名として認識する。「近く」が場所の位置関係を表す言葉であることから、「初台」を勤務先の所在地と理解する(図2-A)。「マンション」が住居に、「暮らす」が住所に関わる言葉であることを認識する。「初台」と「マンション」の文法的な関係性から、「初台」に住む(B)、「マンション」に住む(C)を理解する。たとえ、「多摩市」が未知の言葉であっても、「〇〇市」が地名に関係する特徴的な表現であ

日記1

月曜から土曜まで会社の近くの初台のマンションで暮らす
単身赴任者は、実は多摩市の住民です。

検知されるプライバシー情報	情報の種類
A 会社は初台の近くにある	通勤先
B 初台に住んでいる	住所
C マンションに住んでいる	住居
D 多摩市民である	住所

日記2

個人写真撮影の追加撮影会が決定
平成17年11月14日、15日 10時から19時 リサーチ1階
撮影対象は卒業研究着手者。
アルバム購入は卒業、修了が必須条件ではありません。
ちなみに、H科は昼夜合わせて52人が撮影済みらしいです。

検知されるプライバシー情報	情報の種類
E リサーチ→電気通信大学にある	学校
F H科→人間コミュニケーション学科	学校、専攻

図2 使用した事例日記とプライバシー情報

ることを認識できる。さらに、「住民」が住所に関わる言葉であることから、「多摩市」に住むことを理解する(D)。

日記2

「リサーチ」が未知の言葉であっても、「1階」が建物に関わる言葉であるため、名称と予測できる。そこで、「リサーチ」をキーワードにWeb検索(Google[9]など)する。検索結果から、「図形」との関係を推測する。しかし、先に建物の名称と予測していることから、「リサーチ1階」をキーワードに再度Web検索する。結果から、「リサーチ」が電気通信大学の建物であることを理解する(E)。

さらに、「H科」が学科に関係する略記であると認識する。そこで、「H科」をキーワードにWeb検索し、「人間コミュニケーション学科」との関係を認識する(F)。

人間はこのように、言葉の意味を認識し、意味に基づく言葉の関係性や、間接的に表す意味も理解する。また、未知の言葉でも、表現の特徴や周囲の語の意味から推測する。文法的な言葉の関係性を認識できる。さらに、Web検索結果から必要な情報を取得することもできる。

3.4. システム化に向けた分析

計算機の場合、言葉の意味を認識するためには、意味解析技術が必要となるが、技術は未だ確立されていない。そのため、代替となる方法が必要である。そこで、検知すべき情報は何か、プライバシー情報の種類を示す【種類語】と、ユーザそれぞれの具体的なプライバシー情報である【NGワード】を準備し、日記文章中の語と照合する方法を考える。例えば、【種類語】として「勤務先」を準備する。日記文章中に「会社」が現れた場合、「勤務先」との共起度を求める。共起度とは、2つの単語が同時に出現する頻度を統計的に測定した、関連度である[7]。「会社」と「勤務先」の共起度の高さから、関係性を認識する。なお、語の特定には形態素解析ツール[6]を利用する。

表現の特徴からの意味の推測については、プライバシー情報の種類を表す特徴的な【接尾辞】を準備することで計算機にも認識可能となる。例えば、「市」や「区」といった語である。

文法的な言葉の関係性は、構文解析ツール(JUMAN+knp)を用いて、係り

受け関係として認識する。

周囲の語からの意味の推測は、【種類語】、【接尾辞】を準備し、文法的な言葉の関係性を認識することで代替する。例えば、【種類語】に「住所」が準備され、日記文章が「多摩市の住民です」の場合、まず、「住民」と「住所」の共起関係が認識される。さらに、「住民」と「多摩市」の係り受け関係から、「多摩市」が【種類語】「住所」に関わる【NGワード】として推測できる。

4. プライバシー情報に関する知識

4.1. 知識の準備

以上の分析から、プライバシー情報に関する知識は、ユーザそれぞれの具体的なプライバシー情報である【NGワード】、プライバシー情報の種類を示す【種類語】と種類語と関係の強い特徴的な【接尾辞】の3種類があると分かる。

【種類語】は、人間であってもプライバシー情報が何か、情報の種類を知らなければ検知することができないため、必須の知識である。また、ある程度ユーザ同士に共通することが予想されるため、比較的容易に準備することが出来る。【接尾辞】は、他ユーザの知識を利用するなど、ある程度の準備は可能であるが、全てを網羅することは難しい。一方、【NGワード】はユ

ーザ個人の具体的な情報であるため、最も準備が難しい。そこで、ユーザが準備した【NGワード】の代表的な表現を基に、それに関連する語を事前に自動収集する仕組みが必要となる。

4.2. 事後学習

しかしそれでもなお、知識の網羅的な事前準備は難しい。それは、本人であっても予測できない語句がプライバシーに関わる、時間の経過と共に変更される、複数の表現・言い回しがある、などの場合が考えられるためである。

そこで、検知を通じて学習する。日記文章中に現れた語と、準備された知識との共起関係や、Web検索結果から知識へ到達する場合を「想起」として検知し、学習する。このように知識を強化し、検知精度と処理速度の向上を目指す。

4.3. 知識の形式

プライバシー情報に関する知識は、以下の組み合わせで準備されていると考えられる。

- ① 【種類語】と対応する【NGワード】
- ② 【種類語】と対応する【接尾辞】
- ③ 【種類語】のみ

5. 実現方法

5.1. 処理レベルの分類

縦軸を検知に必要な知識の準備の多さ、横軸を自然言語処理のレベルとして、プライバシー

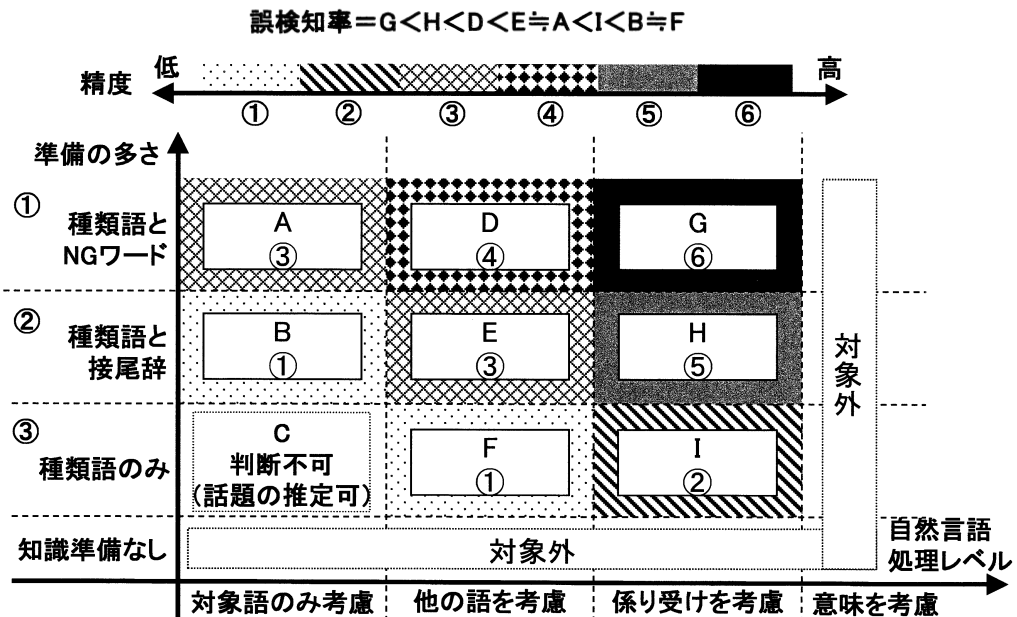


図 3 処理レベルの分類

情報検知システムの処理レベルを分類した（図3）。なお、このうち最も準備が少ない、知識準備なしは、4.1 節で分析した理由から本論文では対象外とする。また、最も高度な自然言語処理の意味解析については、技術的に確立されていないため、今後の課題とする。

ここでは、実現性が高く最も高度な処理である構文解析を用いて、係り受け関係を考慮する場合（図3 ケース I, H, G）について述べる。

ケース I（知識：【種類語】のみ）

文中の注目している語（対象語）が準備された【種類語】である場合、他の語のうち係り受け関係にある語を全て【NG ワード】として検知する。

例えば、【種類語】に「住所」が準備され、日記文が「単身赴任者は多摩市の住民です」の場合を考える。「住民」が【種類語】であるため、「住民」と係り受け関係にある「単身赴任者」、「多摩市」が【NG ワード】として検知される。しかし、「単身赴任者」は「住所」の【NG ワード】として不適であり、誤検知になる。

なお、【種類語】、【NG ワード】は準備されている知識の直接記述だけではなく、想起される

記述の場合も検知する。以下、他のケースでも同様に想起についても考慮する。

ケース H（知識：【種類語】と【接尾辞】）

ケース I と同様に、対象語が【種類語】である場合、係り受け関係にある語を選び、このうち、準備された【接尾辞】を持つ語を【NG ワード】として検知する。また、対象語の接尾辞が【接尾辞】で、係り受け関係にある語が【種類語】の場合にも検知する。

ケース I の例において、【接尾辞】に「市」の準備がある場合、「多摩市」と係り受け関係にある「住民」が【種類語】であるため、「多摩市」が【NG ワード】として検知される。

ケース H は、文中の【種類語】と【接尾辞】のペアが係り受け関係にある場合に検知するため、ケース I に比べて誤検知が少ない。一方で、ケース I は、特徴的な接尾辞が付属するとは限らない情報を検知可能である。例えば、「職業」を表す「プログラマ」「エンジニア」が相当する。

ケース G（知識：【種類語】と【NG ワード】）

対象語が【種類語】であり、係り受け関係にある語が【NG ワード】である場合に検知する。対象語が【NG ワード】で、係り受け関係にあ

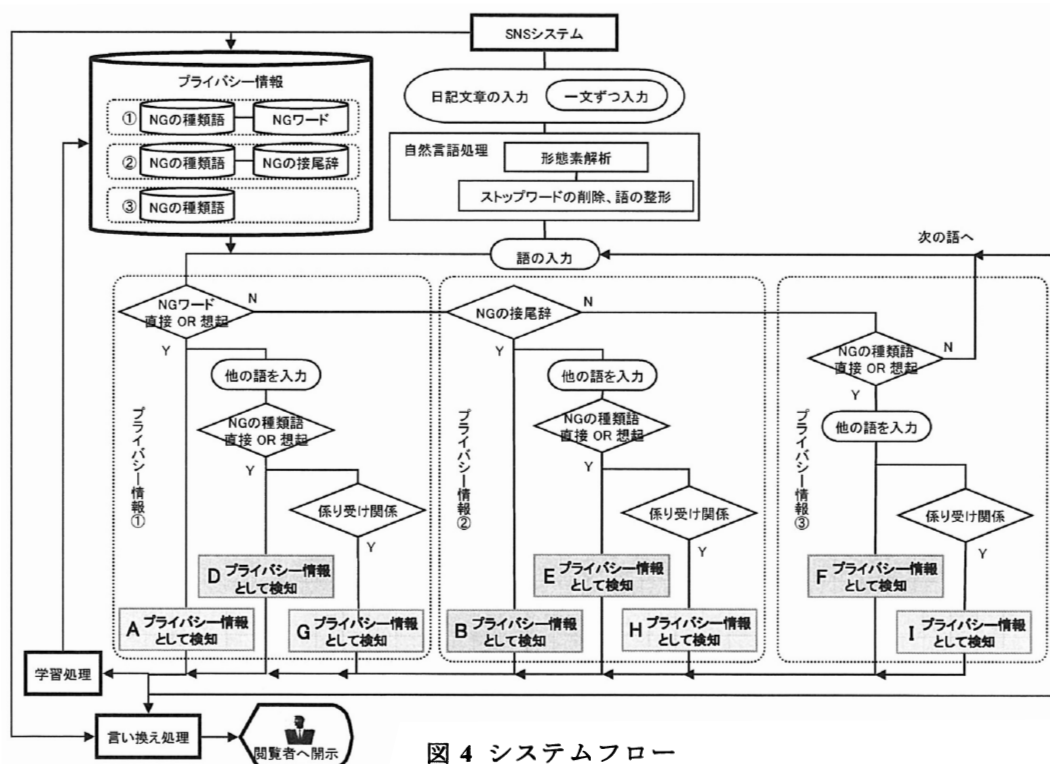


図 4 システムフロー

る語が【種類語】である場合にも検知する。

ケース I の例において、【NG ワード】に「多摩市」が準備されている場合、文中の「多摩市」が【NG ワード】の直接記述であり、係り受け関係にある「住民」が【種類語】であるため【NG ワード】として「多摩市」が検知される。

ケース G は、【種類語】の他、【NG ワード】とも一致する必要がある、【接尾辞】に比べ限定されている。このため、ケース H よりも誤検知が少ない。しかし、4.1 節に示した通り【NG ワード】の準備は難しい。

このように、準備される知識の形式は混在し、それぞれ、準備の難しさや検知を得意とする情報が異なる。そのため、全ての知識形式を考慮したシステムが必要となる。

5.2. システムフロー

以上の分析から、DCNL の実現に向けたプライバシー情報の検知アルゴリズムを提案する。システムフローを図 4 に示す。

閲覧者が日記を閲覧するとき、SNS システムから日記文章が入力される。検知システムでは、一文ずつ入力される。はじめに自然言語処理が形態素解析を行い、ストップワードの削除、語の整形を行う。ストップワードとは、本論では名詞、動詞を用いるためそれ以外の品詞を指す。語の整形は、動詞の活用形を原形に戻す処理や、形態素解析を指す。

自然言語処理を終えると、一語ずつ入力される。プライバシー情報は、準備の多い順番に検知処理を行う。プライバシー情報①の場合、はじめに対象語が【NG ワード】の直接または、想起させる表現か判断し、ケース A として検知する。ここでの想起は、共起度からの検知を指す。次に他の語を入力し、他の語が【種類語】を直接または、想起させる場合、ケース D として検知する。最後に、対象語とその語が係り受け関係にあるかを判断し、ケース G として検知する。ケース A として検知されなかった語に対して、同様に、プライバシー情報②のケース B,E,H の判断をする。さらに、ケース B として検知されなかった語に対して、同様にプライバシー情報③のケース F,I を判断する。ケース F として検知されなかった語は、プライバシー情報に関与しない語と判断し、次の語にうつる。

各々のケースでプライバシー情報と検知された後、新規に判明したプライバシー情報であ

る場合は、学習処理によってプライバシー情報の知識を強化する。最後に、言い換え処理を経て、安全な文章として閲覧者に開示される。

6. まとめと今後の課題

自然言語情報の開示制御 DCNL の実現に向け、特定ユーザの 2 年分の日記を調査した。調査対象の 76% にプライバシー情報が含まれていた。さらに、個別の日記文章の分析を通して、DCNL の最も重要な機能であるプライバシー情報の検知手法について検討し、事前に保持すべき知識の種類と形式を明らかにした。また、準備形式と使用する自然言語処理に応じて、処理レベルを分類し、システムフローを提案した。今後の課題は以下の通りである。

- (1) 実装、評価
- (2) 事後学習処理の確立
- (3) 言い換え処理の確立
- (4) 他のアプリケーションへの適用

参考文献

- [1] 片岡春乃, 他; 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—, 情報処理学会 第 38 回コンピュータセキュリティ研究会 (2007)
- [2] H. Kataoka, et al.; Disclosure Control of Natural Language Information to Enable Secure and Enjoyable Communication over the Internet, 15th International Workshop on Security Protocols (2007)
- [3] Ross Anderson, “Security Engineering”, John Wiley & Sons, 2001.
- [4] ソーシャルネットワークワーキングサービス “mixi,” <http://mixi.jp/>
- [5] 長尾真(編),『自然言語処理』,岩波書店, 1996
- [6] 松本裕治, 高岡一馬, 浅原正幸, 工藤拓, “茶釜と南瓜による日本語解析—構文情報を用いた文の役割分類”, 人工知能学会誌, Vol.19, No.3, pp.334-339, 2004
- [7] Julie Weeds and David Weir, “Co-occurrence retrieval: A Flexible Framework for Lexical Distributional Similarity,” Computational Linguistics, Vol.31, No.4, pp.439-475, 2005
- [8] P. Baldi, et al.; Modeling the Internet and the Web: Probabilistic Methods and Algorithms, John Wiley & Sons (2003)
- [9] NEC 総研,『現代人のプライバシー』,NEC 総研,2005
- [10] 総務省 調査研究会, “インターネット上の違法・有害情報への対応に関する検討会”, http://www.soumu.go.jp/joho_tsusin/policyreports/chousa/internet_illegal/index.html