

多様化するメディア環境に適應する ヒューマンコミュニケーションセキュリティの構想

吉浦 裕[†] 片岡 春乃[‡] 中山 心太[‡]

[†]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

[‡]電気通信大学電気通信学研究科 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: tyoshiura@hc.uec.ac.jp

あらまし Web などのネットワークメディアの発展に伴い、人と人との多様なコミュニケーション形態が可能になる一方、プライバシー情報の漏洩、フィッシング詐欺、誹謗中傷などの問題が生じている。本論文では、ネットワークメディアを用いたコミュニケーションにおいて被害者および加害者にならないための技術を提案する。提案技術は、メディアを通じて発信・受信するコンテンツを監視し、リスクを検知する。リスクの検知に必要な知識はインターネットから取得する。検索と Web マイニングと認知処理によるシステム構成を提案し、応用を検討する。

キーワード 情報セキュリティ, メディア, プライバシー保護, フィッシング

Human communication security to adapt to ever-expanding media environment

Hiroshi YOSHIURA[†] Haruno KATAOKA[‡] and Shinta NAKAYA[‡]

[†]Faculty of Electro-Communications, The University of Electro-Communications,
1-5-1 Chofugaoka, Chofu 182-8585 Japan

[‡]Graduate School of Electro-Communications, University of Electro-Communications,
1-5-1 Chofugaoka, Chofu 182-8585 Japan

E-mail: tyoshiura@hc.uec.ac.jp

Abstract As network-based media such as Web systems are improved, various new forms of human communication become possible but, at the same time, various security problems, such as leaking privacy information, phishing, and slandering, increase. This paper proposes a new technology that protects users of network-based media from being victims and being unintentional attackers. The proposed technology detects the risk by checking content being sent and received through the media. The knowledge necessary for the risk checking is learned from the internet. The system structure, which consists of the internet retrieval, Web mining, and cognitive process, is proposed and applications are discussed.

Keyword Information security, Media, Privacy protection,

1. はじめに

Web などのネットワークメディアの発展に伴い、人と人との多様なコミュニケーション形態が可能になる一方、様々なセキュリティ上の問題が発生している。たとえば、blog や SNS を通じたプライバシー情報の漏洩、BtoC コミュニケーションにおけるなりすましやフィッシング詐欺、匿名掲示板における不適切発言などがあげられる。チャット、レビュー／コメントシステムにも不適切発言のリスクがある。電子メールにも以前からセキュリティ上の問題が存在していた。これらの問題は、多数多様な身元不明の人と接するインターネットの特性および、発展を続けるネットワークメデ

ィアへの適合性および注意力の限界という人間の特性に起因するので、今後も完全に解決されることはないと考えられる。

しかし、ネットワークメディアが多様な発展を続けたとしても、それらが特殊な専門家向けでなく一般人向けであるかぎり、一貫した特性が存在する。その特性とは、コンテンツが自然言語や図面、写真、映像などの human-readable content であり、ユーザは、human-readable content を読んで判断し、human-readable content を書き込んで、開示あるいは送信する。メニュー選択のようなクリック操作の場合でも、その入力ページ全体としては、human-readable

content になっており、例外は少ない。したがって、human-readable content の読み書きの段階で内容を監視すれば、ネットワークメディアの種類に関わらず、人と人とのコミュニケーションのリスクを低減することができると考える。

一方、近年、大量かつ多種類の知識がインターネット上に蓄積されており、情報セキュリティ技術への適用が期待されている[1]。これらの知識の中でも、自然言語テキストを中心とする human-readable knowledge は、Web 上の解説、レビュー、blog や SNS 上の日記、コメント、百科事典など、量と多様性において際立っている。しかしながら、インターネット上の human-readable knowledge を情報セキュリティに適用した例は殆どない。インターネット上の human-readable knowledge の多くは Web 上に集積されているので、上述した human-readable content の監視技術に適用しやすいと考えられる。すなわち、インターネット上に既に蓄積された human-readable knowledge を利用することで、新たに読み書きする human-readable content の監視精度を高めることが可能と考える。その利用を通じて、インターネット上の human-readable knowledge を情報セキュリティに適用する一つの方法論を確立することが期待できる。

また、近年、インターネットの検索技術が大幅な進歩を遂げているが[2]、情報セキュリティの向上に適用した例は殆どない。上述した human-readable content の監視技術は、インターネットの現状把握が必要となる。たとえば、ある会社の取引ページがユーザに表示されている場合に、もしその会社のホームページが存在しなかったら、不正ページであると推定できる。したがって、human-readable content の監視技術は、検索技術を自然に利用することになる。その利用を通じて、検索技術を情報セキュリティに適用する一つの方法論を確立することが期待できる。

以上の考察に基づき、本論文では、human-readable content に着目し、ネットワークメディアを用いたコミュニケーションにおいて被害者および加害者にならないための技術を提案する。以下では、この技術を「ヒューマンコミュニケーションセキュリティ」技術と呼ぶことにする。

2. ヒューマンコミュニケーションセキュリティの提案

2.1 ヒューマンコミュニケーションセキュリティ

下記の性質を持つヒューマンコミュニケーションセキュリティ技術を提案する。

(1) ネットワークメディアを用いた人間のコミュニ

ケーションを、human-readable content の読み書きの段階で監視する。監視の結果、必要に応じて、コンテンツの修正、ユーザへの警告、メディアシステムへの制御を実行する。

- (2) 想定する攻撃者は、悪意あるいは不注意により不適切なコンテンツを送信・開示するメディアユーザである。被害者は、不適切コンテンツを受信するメディアユーザおよび、プライバシー情報などを開示された者である。ユーザの不注意で自身のプライバシー情報を開示してしまう場合には、攻撃者と被害者が同一になる。監視システムは、ユーザの計算機などで動作する。
- (3) human-readable content からリスクの有無や種類を判断するにあたって、ネットワーク上に蓄積された同じ層の知識 (human-readable knowledge) を利用する。
- (4) human-readable content からリスクの有無や種類を判断するにあたって、検索技術に基づいて、情報空間の状態を把握し判断する。
- (5) 基本的な部分については、ネットワークメディアの種類に依存しない統一的な技術である。

2.2 応用例

ヒューマンコミュニケーションセキュリティの応用例として、SNS (social network service) のプライバシー情報漏えい監視[3, 4]と、フィッシング詐欺の検知[5]を検討し、システムを開発中である。SNS の文章の例を下記に示す。

「昨日、調布駅で友達に会った。やはり卒業研究が大変そうだ。」

この文章は、電気通信大学の学生が自分の SNS ページに掲載したものである。本人は、電気通信大学の学生という身元を公開していない。ところが、この文章のうち「調布」と「卒業研究」の組み合わせは電気通信大学を想起させるので、非公開の身元情報が漏洩してしまう。ここでのヒューマンコミュニケーションセキュリティの目標機能は、上記の文章から「調布」を除き、「昨日、駅で友達に会った。やはり卒業研究が大変そうだ。」のように修正することである。このような修正のためには、対象文の解析が必要となるが、それに加えて、「調布」と「卒業研究」が「電気通信大学」を想起させることの検知が必要となる。この検知の方法として、インターネット上の human-readable knowledge に基づく方法と、検索エンジンを用いる方法を検討している。前者の human-readable knowledge に基づく方法では、たとえば、インターネット上のテキ

ストをサンプルとする共起分析[6]¹によって、「調布」&「卒業研究」と「電気通信大学」との共起度を求めることを考えている。後者の検索に基づく方法では、「調布」と「卒業研究」をキーワードとする検索によって「電気通信大学」が検索できるか否か、検索ランクはどの程度かに基づいて、「電気通信大学」への到達度を求める。

後者のフィッシング検知では、フィッシングサイトは正規サイトの模倣であるとの仮定に基づいて、疑わしいWebページのコンテンツから類似検索を行って、正規サイトの候補を求める。疑わしいサイトと正規サイトとの照合によってフィッシング判定を行う。この類似検索のためには、フィッシングサイトのコンテンツを解析して、検索キーワードを抽出する必要がある。その方法として、インターネット上のhuman-readable knowledgeに基づく方法を考えている。たとえば、インターネット上のテキストを母集団とするTF-IDF[7]²によって語の特徴度を求める。

3. ヒューマンコミュニケーションセキュリティの課題

ヒューマンコミュニケーションセキュリティが実用的であるためには、下記の課題を解決する必要がある。なお、下記の一部は、2.1 節で述べた定義の繰り返しであるが、課題を明確化するために改めて記載する。

- (1) 情報空間の変化への適応
Web サイトの出現、消滅、ネットニュースの配信など、情報空間の最新状態および履歴に基づいて判断する。
- (2) 多様性への対応
対象とするメディアのあらゆる human-readable content (テキストや図面、写真など) が情報漏えいなどの可能性を有する。これらのコンテンツは極めて多様であり、また、リスクの有無はコンテンツ自体だけでなく、それが現れる文脈に依存する。
- (3) 事前予想の難しさへの対応
どのようなコンテンツが現れるか事前に予想することは困難である。
- (4) 危険度に応じた網羅性
多様性と事前予想困難性に完全に対処すること

¹ 共起分析とは、サンプルテキストにおいて2つの語が同時に用いられる頻度を測定することで、2つの語の近さを求める手法である。

² TF-IDFとは、対象テキストとサンプルテキストにおける語の出現頻度を測定することで、対象テキスト

は不可能と考えられる。危険度の高い事態ほど確実に対応し、全体として、本技術を用いない場合よりも安全となることが必要である。

(5) ユーザ負担の限定

判断のための知識の入力など、本技術を用いるためのユーザ負担が小さいこと、またメディアのユーザビリティの低下が小さいことが必要である。

4. ヒューマンコミュニケーションセキュリティへのアプローチ

4.1 実現方針

まず、実現のための基本処理について述べ、次に、これらの基本処理の連動について述べる。

4.1.1 基本処理

(1) 認知処理を用いたコンテンツの分析

2.2 節で例示したような判断を行うには、文字列マッチングのようなコンテンツの表層的な分析だけでは不十分であるから、構造や意味を抽出する認知処理を実施する。なお、ここでいう認知処理とは、単語、構文、意味を抽出する自然言語処理[8]や、線分、2次元構造、3次元構造を抽出する画像処理など、human-readable contentの要素、構造、意味を抽出する処理を指す。

(2) Web マイニングを通じたインターネット上の human-readable knowledge の利用

2.2 節で例示したような判断を行うには、判断のための知識が必要であり、それをインターネットから収集する手段としてWeb マイニングを用いる。現時点では、主にテキスト知識を考えており、共起分析などのテキストマイニング手法[6, 7]を用いる。

(3) インターネット検索を用いた情報空間の状態把握

1 章で例示したように、ある会社の取引ページがユーザに表示されている場合に、もしその会社のホームページが存在しなかったら、不正ページであると推定できる。このように情報空間の状態に基づいて判断するために検索を用いる。

4.1.2 基本処理の連動

ヒューマンコミュニケーションセキュリティの課題解決にあたって上記の基本処理をどのように組み合わせるかについて述べる。組み合わせ方の指針としては下記が挙げられる。

- (1) 汎用性、柔軟性が高く、多様なメディアに適用できること。

における語の特徴度を求める手法である。

- (2) システム開発者にとって自然で理解しやすいこと。
- (3) 個々の基本技術の弱点を相互に補うこと。

ここでは、2.2節で述べた具体例の検討に基づいて、下記の連動方式を提案する。

- 認知処理によってコンテンツから文字列キーワードを求め、文字列キーワードを用いてインターネット検索を実行し、検索結果に基づいて判断処理を実行し、判断結果に基づいてリスク回避処理を実行する。

2.2節の例のうちSNSのプライバシー漏洩防止の場合は、SNS上の文章を解析して、プライバシー漏洩につながる語句の候補（調布、卒業研究）を求め、その語句を用いて身元情報（電気通信大学）を検索し、検索結果に基づいてプライバシー漏洩のリスクを判断する。また、フィッシング検知の場合も、疑わしいサイトのコンテンツを解析して、類似検索のための情報を求め、それを用いて正規サイトを検索し、疑わしいサイトと照合してフィッシングのリスクを判断する。このように考えると、認知の結果を用いて検索し、その結果を用いて判断するという順序が、一つの自然な順序であると考えられる。

また、インターネット全体を対象とした高効率・高信頼の検索は、文字列キーワードを用いるものに限られ、画像や構造データを直接の検索語とするものは存在しない。そこで、認知処理が抽出する検索語は文字列キーワードとするのが妥当である。

- Webマイニングを用いて認知処理を補足、代替する。「調布」と「卒業研究」がプライバシー漏洩につながる語句であることは、一般的な意味辞書やオントロジーだけでは推定できず、地域や学校に関する応用依存の知識が必要となる。これらの知識を応用ごとに構築・更新するのはコストがかかるため、Webマイニング（たとえば共起分析）を通じてインターネット上の human-readable knowledge から抽出したい。

また、認知処理のうち低次の部分（自然言語の形態素解析、画像からの線分抽出など）については技術が確立し、利用可能なツール（形態素解析ソフトなど）が整備されている。しかし、意味理解などの高次部分については、技術が未確立であり、ツールも整備されていないため、その実装は容易ではない。そこで、認知の高次部分を直接実装せず、Webマイニングを通じて同等の結果を得ることが望ましい。たとえば、上記の例の場合、意味辞書やオントロジーを用いた意味解析をそもそも省略し、「調布」、

「卒業研究」の意味を扱わず、共起分析などの統計処理のみによって、「電気通信大学」へのつながり度合いを求めるのが望ましい。

以上のように、認知処理のうち高次部分の補足できれば代替としてWebマイニングを利用する。Webマイニングを起動するタイミングについては、認知処理の一部としての起動と、事前の知識学習としての起動が考えられる。認知処理の一部とする場合には、オンライン処理となるので性能への大きな影響が生じる。したがって、知識のうち必要性が予測できる部分についてはオフラインのマイニングとし、オンラインのマイニングをどこまで行うかは応用による。

4.2 システムモデル

前節の考察に基づいて、ヒューマンコミュニケーションセキュリティのシステムモデルを図1、図2のように提案する。個々の応用システムは、このモデルを具体化したものになる。図1は提案システムの位置づけ、図2はシステム構成を示す。

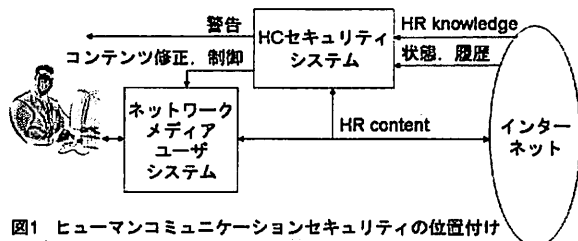


図1 ヒューマンコミュニケーションセキュリティの位置付け
(注) HC: Human Communication, HR: Human-Readable

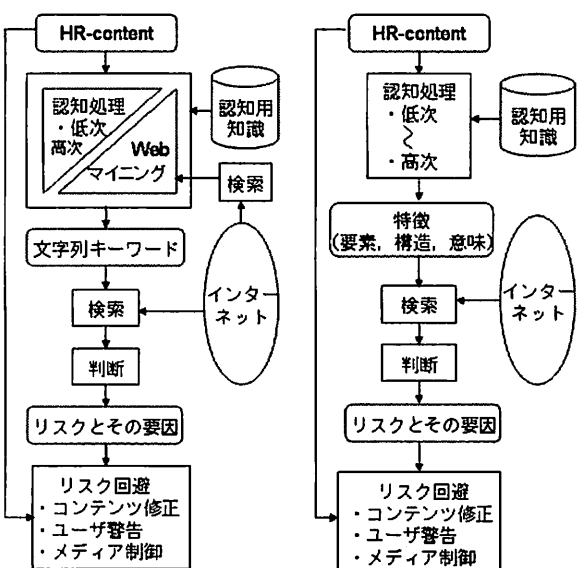


図2 システム構成

図3 直感的モデル

4.3 システム構築方法

具体的なシステムを図2のモデルから直接構築することは難しいと考えられる。まず第一に、図2はシステム開発者が直観的に理解しやすいモデルでない。また、どのようなWebマイニングを用いればよいか、認知の低次部分とWebマイニングをどのように接続するか、高次部分とWebマイニングをどのように相補的に連結するか、各々の処理の入出力は何か、などを直接考えることは困難である。さらに、最初から図2のシステムを開発すると、将来的に監視精度を向上するための方向性（たとえば将来意味処理を充実させると何が可能になるか）が不明確になる。そこで、図3のような直観的モデルに基づいてシステムを設計した後、それを図2の形態に変更するという2段階の構築方法を提案する。実際、2.2節で述べた応用例も、この2段階のモデルに沿って設計した。

(1) 直観的モデル (図3)

膨大なコストをかけて、低次から高次までの認知処理全体を構築すると仮定したモデルである。具体的には、自然言語理解や画像理解の高次部分を実用的な規模で構築するものと仮定する。また、認知に必要な知識も膨大な人手をかけてコーディングするものとする。検索は、画像や構造体データを直接の検索語として、高度なパターンマッチングなどを実行できると仮定する。すなわち、コストと時間が無限にあり、実験室レベルの技術が自在に利用できると仮定したモデルである。このモデルに沿って考えることで、認知処理の高次部分が何をするのか、検索が何をするのか、どのような知識を用いるのか、について、直観的に理解できる。

(2) 実現モデル (図2)

前節で提案したモデルである。認知処理の高次部分に関する理想的な機能を、認知処理とWebマイニングに分割し、どのようなWebマイニングを用いて、理想機能のどれだけを実現するかを設計する。また、画像や構造体データを用いた検索機能を、文字列キーワードを用いた検索で置き換えるために、認知処理が何をするかを設計し、理想的な検索機能のどこまでを実現するか設計する。

5. 具体例

2.2節の応用について、上述した直観的モデルと実現モデルの具体例を示す。

5.1 SNSのプライバシー保護

この応用における直観的モデルを図4、実現モデルを図5に示す。図4の直観的モデルでは、コストをかけて構築した自然言語処理と知識を仮定している。自然言語処理は、プライバシー漏洩の可能性がある要注意表現（語句や文章）のリストを出力する。検索処理では、それらの要注意表現およびその組み合わせを検索語として、インターネット検索を行う。NG表現推定処理は、各々の要注意表現およびその組み合わせについて、検索されたサイト情報にプライバシー情報が含まれる場合に、対応する要注意表現および組み合わせをNG表現と判定する。修正・警告処理は、NG表現を上位語に置換するなどにより、文章を修正し、ユーザーに警告を出す。

図5の実現モデルでは、自然言語処理の構文解析、意味解析を共起分析で置き換えている。また、解析知識のうち文法規則、意味辞書、分野知識を、語間の共起度で置き換えている。図5の直観的モデルを経たことで、図6の設計へ自然に導かれる。また、直観的モデルにより、将来の拡張として、下記が明らかになる。

- (1) 構文解析、意味解析の導入により形態素解析の誤り率を低減し、要注意語の抽出精度を向上させる。
- (2) 意味解析によって、未知語のカテゴリを推定し、たとえば、人名であれば要注意語と判断する。な

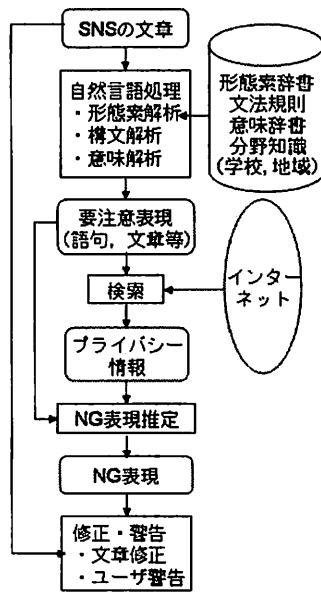


図4 SNSプライバシー保護の直観的モデル

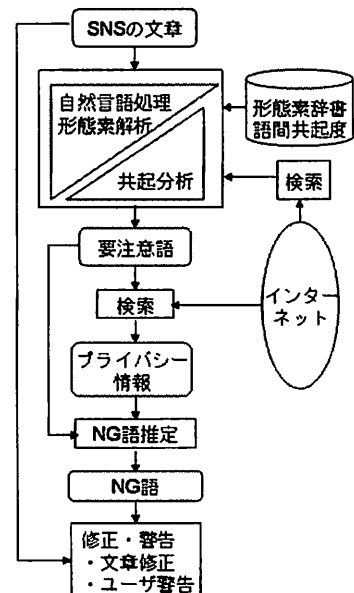


図5 SNSプライバシー保護の実現モデル

お、この程度の意味解析であれば、一般的な意味辞書により比較的容易に実装できる。

- (3) 共起分析以外のテキストマイニング手法を併用することで、要注意語の抽出精度を向上する。

5.2 フィッシング検知

直観的モデルを図6、実現可能モデルを図7に示す。疑わしいサイトのコンテンツはテキストとは限らず、図面や画像、動画、音声も含むので、直観的モデルの認知処理はマルチメディア解析処理となる。認知処理の出力はコンテンツの特徴を表す情報であり、テキスト、図面、画像などの特徴的な要素、構造、意味である。類似検索処理は、これらの特徴情報を用いて類似コンテンツのサイトを検索し、正規サイトの候補とする。照合処理は、疑わしいサイトと正規サイト候補を比較し、疑わしいサイトがフィッシングであるかを判定し、警戒レベルを決定する。警戒レベルに基づいてユーザへの警告を実施する。

実現モデルでは、マルチメディア解析処理を自然言語処理で置換し、形態素解析とTF-IDFで構成している。また、知識を形態素辞書と語の特徴度(TF-IDF値)で置換している。自然言語処理の出力は特徴語のリストである。これらの特徴語およびその組み合わせを検索語として、キーワード検索を実行する。図6の直観モデルを経たことで、図7の設計へ自然に導かれる。また、直観モデルにより、将来の拡張として、下記が明らかになる。

- (1) 上記 5.1 節(1)-(3)と同様に、構文解析、意味解析、別種類のテキストマイニング手法を導入することで、正規サイト候補の検索精度を向上する。
- (2) 画像処理、図面解析処理、音声処理などを導入することで、テキスト以外の部分から特徴情報を抽出し、それらの特徴情報を文字列(特徴語)に変換し、検索語に加えることで、検索精度を向上する。
- (3) 画像処理、図面解析処理、音声処理などを導入することで、照合処理の精度を向上する。

6. 関連研究

著者らの知る限り、2.1 節で提案したヒューマンコミュニケーションセキュリティと同一の技術は存在しない。しかし、部分的に類似した技術は存在する。以下に主な技術をあげる。

- (1) フィルタリング

SPAMメールのフィルタリングは、メ

ールテキストを解析して SPAM の程度を認識し、振り分けやコメント付け等を判断し、実行する[9]。テキストの解析は形態素解析を含む。また、協調型フィルタリングでは、多くのユーザからの SPAM 報告を利用する。この処理は、多くのユーザの受信したメールを解析して判断精度を向上するので、同じ層の human-readable content を判断に利用していると言える。

- (2) ビデオサーベイランス

撮影映像を解析することで、不審者の存在や事故の発生などを認識し、必要な対応を判断し、通報などを実行する[10]。

- (3) フィッシング検知

[11] は、メールや Web ページのテキストから、どの企業を名乗っているかを認識し、ホワイトリストとの照合によりフィッシングの可能性を判断し、ユーザへの警告を実行している。

7. まとめと今後の課題

ネットワークメディアを用いた人と人とのコミュニケーションにおいて被害者および加害者にならないためのヒューマンコミュニケーションセキュリティを提案した。提案技術は、メディアを通じて発信・受信する human-readable content を監視し、プライバシー情報の漏洩や不適正表現を検知する。Web マイニングを通じてインターネット上の human-readable knowledge を利用し、検索を通じて情報空間の現状を反映した判断を行う。提案技術の課題を明らかにし、Web マイニングと検索と認知処理を連動したシステム構成および構築方法を示した。また、SNSのプライバ

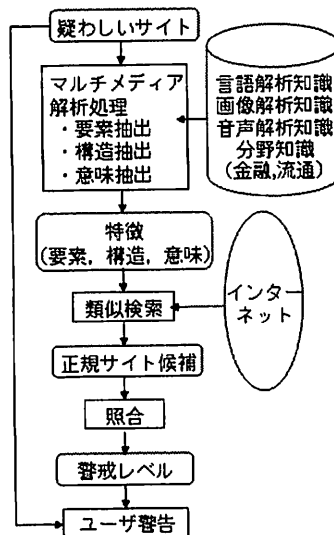


図6 フィッシング検知の直観的モデル

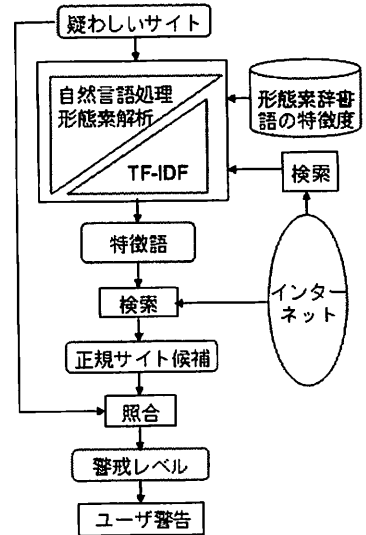


図7 フィッシング検知の実現モデル

シー保護とフィッシング検知における検討例を示した。

今後は、上述した2つの応用システムを完成、評価するとともに、第3の応用を検討して、本構想の有効性と汎用性を検証する。

謝辞

本研究は平成19-20年度科学研究費補助金（特定領域研究・課題番号19024032）の助成を受けたものである。また、被害者にも加害者にもならないという発想は東京電機大学佐々木良一教授の著書よりいただいた。ここに御礼申し上げる。

参考文献

- [1] 特集：情報セキュリティと AI, 人工知能学会誌, Vol.21, No.5, 2006年9月.
- [2] 特集：検索エンジン 2005, 情報処理, Vol.46, No.9, 2005年9月.
- [3] 片岡春乃, 内海彰, 広瀬友紀, 吉浦裕：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－, DPS/CSEC 合同研究会, 2007-CSEC-36, pp.321-326, 2007年3月.
- [4] H. Kataoka, A. Utsumi, Y. Hirose, H. Yoshiura : Disclosure Control of Natural Language Information to Enable Secure and Enjoyable Communication over the Internet, Proc. 15th International Workshop on Security Protocols, pp.136-147, Brno, April 2007.
- [5] 中山心太, 片岡春乃, 吉浦裕：模倣コンテンツの特性に基づくフィッシング検知方式, ISEC/SITE/CSEC 合同研究会, 2007-CSEC-38, 2007年7月.
- [6] J. Weeds, D. Weir : Co-occurrence retrieval: A Flexible Framework for Lexical Distributional Similarity, Computational Linguistics, Vol.31, No.4, pp.439-475, 2005.
- [7] P. Baldi, P. Frasconi, P. Smyth : Modeling the Internet and the Web: Probabilistic Methods and Algorithms, John Wiley & Sons, 2003.
- [8] 長尾真(編)：自然言語処理, 岩波書店, 1996年
- [9] 安藤一憲：フィルタリング, 情報処理, Vol.46, No.7, pp.758-761, 2005年7月.
- [10] 特集：安心と安全のための画像処理技術, 情報処理, Vol.48, No.1, 2007年1月.
- [11] 柴田賢介, 荒金陽助, 塩野入理, 金井敦：Web サイトからの企業名抽出によるフィッシング対策手法の提案, DPS/GN/EIP 合同研究会, 2006-DPS-128, pp.17-22, 2006年9月.