

視覚に障害をもつ医療系学生に適した情報セキュリティ技術

岡本 健¹⁾, 山口通智²⁾, 三宅輝久¹⁾, 石塚和重³⁾, 野口栄太郎⁴⁾, 大越教夫⁴⁾

筑波技術大学 保健科学部 情報システム学科¹⁾

筑波技術大学大学院 技術科学研究科 保健科学専攻²⁾

筑波技術大学 保健科学部 保健学科 理学療法学専攻³⁾

筑波技術大学 保健科学部 保健学科 鍼灸学専攻⁴⁾

要旨: 情報セキュリティ対策は、安全な ICT 社会を構築する上で不可欠である。現在、実社会で利用されている情報セキュリティの多くは、画像に対する認識能力をもたなければ、安全なサービスを提供できない。このことは利用者が視覚に障害をもつ場合、セキュリティの観点で深刻な問題となる。また、一般的にサービス利用者は、十分な情報リテラシを持つときに限り、ICT の安全性が担保されるといふ現状がある。本稿では、視覚障害者を取りまく ICT および情報セキュリティの現状と問題点についてまとめると共に、これらの問題を解決するために、我々がこれまでに取り組んできた研究活動について紹介する。研究テーマの一つは、「CAPTCHA (キャプチャ)」と呼ばれるものであり、提案方式では、従来、認証を行う際に視力などの特定知覚に依存していたものを排除し、バリアフリー化を実現している。これらの特徴は、本学の医療系学生に適していると考えられ、種々の実証実験を行うことにより、提案方式の有効性を検証した。

キーワード: 情報セキュリティ, アクセシビリティ, 情報バリアフリー, CAPTCHA

1. はじめに

1.1 本学における ICT 利用の背景

ICT (Information and Communication Technology) の著しい発展は、日常生活において、我々に多くの恩恵をもたらしてきた。その一方で、視覚障害者は、身体的条件により、インターネットで提供されている各種サービスの多くが利用できないことが報告されている [5]。情報セキュリティに関しても、同様のことが言え、現在の視覚障害者の ICT 促進に対する大きな障壁となっている。

本学保健科学学部では、これまで視覚に障害を持つ学生に対し、eラーニングの利用 [9] やパソコンを使用した ICT の促進 [8] を試みてきた。本学の eラーニング導入にあたっては、情報のアクセシビリティを常に考え、利用する学生がどの学科・専攻に所属しているかを考慮し、専門性に合わせた取り組みが求められる。例えば、本学保健学科 (鍼灸学専攻, 理学療法学専攻の2専攻から構成) の学生の場合、国家試験に合格しなければ、卒業後、鍼灸士や理学療法士として就労できない。このため本学の医療系学生に適した国家試験対策の eラーニング教材を構築することは、学生の学力や就業力を向上させるという点で大変有益と考えられる。

一方で本学の場合、2つの大きな課題がある。1つめは、医療系学生共通の課題である。医療系の教育機関は、当然ながら医療に関する講義や実習の割合が高く、それらの多くは必修科目となっている。相対的に情報リテラシに関する授業時間数やパソコンに触れる機会は限られる。

2つめは本学の障害者支援の特殊性に関連した課題である。アンケートによれば、本学の医療系学生は、中途の視覚障害者の割合、および平均年齢が他学科と比較して高い。また、パソコンの習熟度についても、普段の生活において、パソコンをほとんど使用しないと回答する学生が多くいる。eラーニングの導入や情報セキュリティ対策については、これらの課題を十分考慮した上で、構築する必要がある。

1.2 ウェブアクセシビリティの問題

eラーニングは web を用いる場合が多いが、現状ではアクセシビリティに関して様々な問題がある。報告書 [12] によれば、視覚障害者がインターネット利用時に困る点として、「障害に配慮したホームページが少ない」、「画面が煩雑で見にくい」、「欲しい情報がない、また見つけるのが難しい」が上位にあげられている。

前者の2つは、提供されるウェブページに起因する。具

体的には、フレームなどのレイアウトや画像情報の説明不足が問題になる。さらに近年では、画像の認識・解釈能力を使用した情報セキュリティ技術が導入され、その利用に関する問題が発生している。

後者は、利用者の情報リテラシに起因する。これは視覚障害者に限った話ではなく、ICTに馴染みがなければ、その恩恵を十分に受けることはできない。

以上のように、ウェブアクセシビリティの問題は、情報の提供者と受益者の双方に原因がある。しかし、一般に受益者側の情報リテラシにはある程度の格差があり、容易にそれを改善することはできない。このため、サービスの提供者側は、情報リテラシの格差が影響しにくい形でデータを提供する必要がある。

1.3 情報セキュリティ技術のバリアフリー対応

パスワードを用いたログイン認証を例にあげる。テキストベースのパスワード方式は、キーボードを用いれば実装でき、特殊な機器を必要としないため、現在でも広く利用されている。しかしながら、パスワード管理の困難さや盗み見に弱いことから、最近では、新しい方式がいくつか提案されている。

例えば、画像を利用し、その特定のオブジェクトの組み合わせをパスワードとして認証する方式がある。盗み見への耐性をもつものとしては、ランダムに並べられた数字パネルを使い、パスワードを選択させる方式がある。また、数字パネルの代わりに、複数の画像を用いるタイプもあり、近年では、人間の画像解釈能力を組み合わせ、より高度な方式 [6] が提案されている。

新しい認証方式の傾向として、画像の解釈を用いた認証方法が利用されている場合が多い。これは、現在の計算機が、画像の意味論的な解釈を未だ困難としており、悪意あるプログラムからのオンラインを用いた総当たり攻撃に耐性があるためである。

一方、視覚障害者が、これら画像を用いた認証方式を利用することは本質的に不可能である。視覚障害者が安全性確保のため、これらを利用せざるを得ない場合は、支援者など人に頼らざるを得ない。また人から支援を得られたとしても、視覚障害者のプライバシーの問題は残る。



図1 画像型 CAPTCHA の例

この対応策として、カクテルパーティ効果やエピソード記憶など、画像の利用を必須としない解釈能力を用いた方式も提案されているが、現状では提案数自体が少ない。このことから、今日の情報セキュリティ技術の研究動向において、バリアフリー問題に注意を払う意識は低いと考えられる。

2. CAPTCHA

2.1 現状と課題

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart)は、人間と人工知能による自動プログラム(ロボット)を判別するチューリングテストの一種として知られており、ネット上での認証手続きに利用されている。2000年に発明されたこの

テストは、現在、ウェブを用いた投票やアンケート、無料のメールアドレス取得時などで使用され、今では重要な情報セキュリティ技術の一つとして認識されている。それにもかかわらず、現在でもこのテストは、視覚障害者がウェブを利用する際の大きな障壁になっている。

まず、歪んだ文字画像を読み取らせる画像型 CAPTCHA (図1)は、CAPTCHAのもっとも代表的なタイプであるが、視覚障害者は利用できない。代替案として、変形した音声を利用する音声型 CAPTCHA があるが、これは自動プログラムのみならず、人間にも判別が難しいと指摘 [1],[2] されている。さらに、我々の調査 [14] でも、Google のアカウント認証で利用可能な音声型 CAPTCHA に対する視覚障害者らの成功率が、0% であるとの結果が出ている。これらの事実は、音声型 CAPTCHA が、画像型の代替を果たしていないことを示している。社会情勢としても、2013年には、NFB (National Federation of the Blind) やオンライン請求サイト「Change.org」において、CAPTCHA のアクセシビリティ問題が相次いで指摘されており、世界的な注目を集めている。

2.2 必要要件の付加

従来、CAPTCHA に対して求められてきた必要要件を以下にまとめる。

- 操作性要件：利用の際に、複雑な操作を必要とせず、情報リテラシが低い利用者にも対応していること。
- 知識非依存性要件：テストの難易度が特定の知識の有無に強く依存しないこと。
- 識別性要件：人間には容易に解けるが、現状のロボットには解答が難しい問題を生成できること。
- 問題新規性要件：未使用で新規な問題を無数かつ自動で作れること。

これまでの研究では、CAPTCHA の安全性のみが強調され、利用者の取り巻く環境について、あまり考慮されなかつ

た。一方で我々の研究では、利用者の利便性、情報アクセシビリティの強化という観点から、以下の要件を明示的に加える。

- バリアフリー要件：特定の知覚のみの使用に限定されないこと。

既存の CAPTCHA の多くは、特定知覚による解釈能力の使用を容認して作られおり、これらはバリアフリー要件を満たしていない。このとき、その知覚に障害がある人は、その方式に伴う各種サービスを楽しむことができないことになる。一方で我々の研究では、人間が本来もつ言語解釈能力を用いて認証を行うことにより、バリアフリー要件を満たす CAPTCHA の構築を目指す。

3.3 バリアフリーな CAPTCHA に関する既存研究

先行研究として、人間のもつ知識に依存したクイズを用いる方式 [7] が提案されている。しかし、IBM の Watson や Apple の Siri といった自然言語で質問を受け付け、正しい解答をする人工知能の登場により、この方式では識別性要件を満たさない [3]。さらに、作問内容が一般常識の範囲を超えると、知識非依存性要件も問題となる。

特定の知覚や知識に依存しない方式には、文章の文意や文脈を解釈する問題（文意文脈解釈問題と称す）を用いた研究 [11], [16] がある。これらは、人間の作った自然な文章と、機械合成文や機械翻訳文のような不自然な文章を提示し、利用者に解答させる方式である。文法的に正しい文の違和感を問うので、現在のロボットには解答が難しいと期待できる。しかし、これらの方式では、問題生成に必要な文章に秘匿文章を利用している。これは、問題として提示された文章から、解答するためのヒントを検索で取得されないようにするための処置ではあるが、秘匿文章の分量は作問要求に比べて極めて少量であるため、問題文新規性を満たすことは難しい。

以上のように、バリアフリーな CAPTCHA を構成するためには、識別性要件と問題新規性要件が重要となる。問題新規性要件を満たさない方式は、攻撃者が CAPTCHA に何度も挑戦することで、問題文と解答のペアを収集できてしまう。この場合、自動プログラムがペアを保持すれば、容易に問題が解けてしまうことから識別性要件も満たしていない。

3. バリアフリー性をもつ提案方式

3.1 研究のアプローチ

我々の研究では、バリアフリー要件を満たすため、文意文脈解釈問題に着目した。文意文脈解釈問題は、文字情報として提示されれば十分なので、視覚ディスプレイ、点字ディスプレイ、音声読み上げのいずれの手段にも対応できるた

め、バリアフリー性が満たされる。さらに、スクリーンリーダや点字ディスプレイのサポートがあれば、問題の提示方法は、HTML に文章を記載することで共通化できる。これは、図 1 のような利用する知覚ごとに問題の提示方法を切り替える方式に比べ、利用者が解答に要する操作を簡潔にすることができる。

作問に必要な文章については、インターネット上の公開文章を利用し、数に制限のない作問をおこなう。しかしながら、攻撃者が公開文章を検索し、問題に対するヒントを得ることが可能となるため、このままでは識別性要件を満たさない。対策として、提案方式では、問題文として表示する文字列の子音を改変する。これは、方言などに見られる単語の子音の違いを指す。例えば、ザ行からダ行への子音交替では、「ざぶとん」を「だぶとん」と改変する。漢字に対しては、仮名に開いてから処理を適用する。子音交替は、文の誤植や音声での聞き間違えに似た処理である。改変率が一定値以下ならば、人間は文意文脈を解釈して、改変前の内容を推測できることを期待している。

利用する文意文脈解釈問題の種類については、共通話題識別テストと機械合成文識別テストを用いる。

本稿では、ページ数の制限のため、提案方式に関する主な特徴や概要のみを示す。詳細が必要な場合は、参考文献 [10], [14], [15] を参照されたい。

3.2 共通話題識別テスト

共通話題識別テストとは、共通する話題の文脈に現れる文を複数個、回答者に提示し、共通話題が何であるかを選択肢から回答させる問題である。作問者は、知識非依存性要件を満たすように、一般的な話題を選択する必要がある。

共通話題の収集は、回答として提示する選択肢に関連した文章を、検索によって取得する。ただし、選択肢の語句を単純に検索語として問い合わせると、問題文に直接回答が表示されてしまう。このため提案方式では、類似語を用いて検索をおこなう。またその際に、選択肢となる語句を Not 条件に加える事で、前述の問題を防止する。

また、提案方式では、KF ダイバージェンス [4] を用いた攻撃にも対策を講じた。これは、子音交替による形態素解析の妨害と、仮名展開し、既存データベースの利用を防ぐことで対応した。

さらに、問題に偽の話題を混入する。回答者には偽の話題が小数混入していることを告知して、多数を占める共通話題を答えさせる方式にすることで、ロボットによる回答をより困難にした。

3.3 機械合成文識別テスト

機械合成文識別テストとは、人間が作成した自然な文と、機械合成文とを並べて回答者に提示し、機械合成文を選ばせるテスト方式である。機械合成文には、式 (1) に示される N 階マルコフ連鎖モデルを用いて生成されるワードサラダを利用する。

$$\begin{aligned} \Pr[X_{n+1} = x | X_n = x_n, \dots, X_0 = x_0] = \\ \Pr[X_{n+1} = x | X_n = x_n, \dots, X_{n-N+1} = x_{n-N+1}] \end{aligned} \quad (1)$$

この方式は、文中のある形態素は、直前 N 個の形態素により決定される連鎖型共起表現であるという仮定に基づく。具体的には、問題の素材となる文章を形態素解析したものをコーパスとし、式 (1) に従い組み合わせることで、コンピュータにより効率的に生成できる。ワードサラダは、「てにをは」といった文法構造は正しいが、登場する単語はランダムに選ばれているため、内容が不自然な文章である。ワードサラダは、文法は正しい文章であるため、文法チェックなどのプログラムでは自然文との見分けがつかない。内容の自然さの識別には、常識が必要となるためロボットには難しいと期待できる。

ワードサラダの検出攻撃としては、離散的共起表現を用いた方式 [13] がある。提案方式では、子音交替による形態素解析の妨害と、離散的共起表現を取り込んだワードサラダの生成により、その攻撃に対抗する。提案方式における離散的共起表現の抽出は、構造解析による係り受け関係の情報を利用している。

これらのアプローチを用いて、我々は本学の医療系学生を対象として、各種の実証実験 [10] を行った。結論として、本学の医療系学生と、同じく本学の情報系学生の間で実験結果に統計的優位性は認められなかった。このことから、提案方式は情報リテラシの習得が必ずしも十分でない人達に対しても、安全なサービスを楽しむことができる方式であると考えられる。

4. まとめ

本稿では、ウェブアクセシビリティを阻害する要因として、情報セキュリティ技術のバリアフリー対応が不十分である点を指摘し、その代表例である CAPTCHA の現状を示した。さらに、バリアフリー要件を満たし、情報リテラシが必ずしも十分でない利用者にも対応した新しい CAPTCHA を提案し、その概要について説明した。提案方式は、文意文脈解釈問題を利用することで知覚依存を解消し、時々刻々作り出されるネット上の文章データを作問の種にすることで問題の新規性を保ち強度を与えるものである。また、ネット上の文章データを安全に使用するための問題点をあげ、対処法を示した。

参考文献

- [1] Jeffrey P. Bigham and Anna C. Cavender. "Evaluating Existing Audio CAPTCHAs and an Interface Optimized for Non-visual Use", CHI '09, pp. 1829–1838. ACM, 2009.
- [2] Elie Bursztein, Steven Bethard, Celine Fabry, John C. Mitchell, and Dan Jurafsky. "How Good are Humans at Solving CAPGCHAs? a Large Scale Evaluation". SP '10, pp. 399–413. IEEE Computer Society, 2010.
- [3] Robert M. French. "Moving beyond the Turing Test", Communications, 55(12):74–77, ACM, 2012.
- [4] B. Fuglede and F. Topsoe. "Jensen-Shannon divergence and Hilbert space embedding", IEEE International Symposium on Information Theory, pp. 31–31, 2004.
- [5] Elaine Gerber and Corinne Kirchner. "Who's Surfing? Internet Access and Computer Use by Visually Impaired Youths and Adults", Journal of Visual Impairment and Blindness, Vol. 95(3):176-181, 2001.
- [6] Eiji Hayashi, Jason Hong and Nicolas Christin. "Security Through a Different Kind of Obscurity: Evaluating Distortion in Graphical Authentication Schemes", CHI '11, pp. 2055–2064, ACM, 2011.
- [7] Pablo Ximenes, Andre Santos, Marcial Fernandez, and Jr. Celestino, Joaquim. "A CAPTCHA in the Text Domain", MMIS '06, Vol. 4277 of LNCS, pp.605–615, Springer-Verlag, 2006.
- [8] 池部実, 三宅輝久, 大越教夫, 小野東, "視覚障害者のアクセシビリティを考慮した Moodle システムの改良", 筑波技術大学テクレポ Vol.19 (2), 2012.
- [9] 岡本健, 坂尻正次, 三宅輝久, 石塚和重, 野口栄太郎, 大越教夫. "視覚に障害をもつ医療系学生のための e ラーニング支援", 情報科学技術フォーラム (FIT2013), 第 3 分冊, pp.655-656, K-043, 情報処理学会, 2013.
- [10] 岡本健, 山口通智, 三宅輝久, 石塚和重, 野口栄太郎, 大越教夫. "バリアフリーな CAPTCHA の基盤構築: 視覚に障害をもつ医療系学生を事例として", 暗号と情報セキュリティシンポジウム (SCIS2014), 4B2-1, 電子情報通信学会, 2014.
- [11] 鴨志田芳典, 菊池浩明. "文章合成の不自然さの評価と応用", ファジィシステムシンポジウム講演論文集, 26:1069–1074, 2010.

- [12] 総務省 IICP 情報通信制作研究所. “障がいのある方々のインターネット等の利用に関する調査報告書”, 調査研究報告書, 2003, 2012.
- [13] 森本浩介, 片瀬弘晶, 山名早人. “N-gram と離散型共起表現を用いたワードサラダ型スパム検出手法の提案”, 情報処理学会研究報告, データベース・システム研究会報告, 148:1-8, 2009.
- [14] 山口通智. “人間ロボット判別テストのバリアフリー化のためのネット上文章の採取加工技法”, 論文誌, , pp.337-352, Vol.15, No.4, ヒューマンインタフェース学会, 2013.
- [15] 山口通智, 岡本健. “人間ロボット判別テストのバリアフリー化のための言語的作問とその自然文生成技法”, コンピュータセキュリティシンポジウム (CSS2013), 3D3-3, 情報処理学会, 2014.
- [16] 山本匠, J.D. Tygar, 西垣正勝. “機械翻訳の違和感を用いた CAPTCHA の提案”, 情報処理学会研究報告. CSEC, [コンピュータセキュリティ], 2009(37):1-8, 2009.

Information Security for Medical Student who are Blind and Visually Impaired

OKAMOTO Takeshi¹⁾, YAMAGUCHI Michitomo²⁾, MIYAKE Teruhisa¹⁾,
ISHIZUKA Kazushige³⁾, NOGUCHI Eitaro⁴⁾, OHKOSHI Norio⁴⁾

Department of Computer Science, Faculty of Health Science,
Tsukuba University of Technology¹⁾
Course of Information Science, Division of Health Sciences, Graduate School of Technology and Science,
Tsukuba University of Technology²⁾
Course of Physical Therapy, Department of Health, Faculty of Health Science,
Tsukuba University of Technology³⁾
Course of Acupuncture and Moxibustion, Department of Health, Faculty of Health Science,
Tsukuba University of Technology⁴⁾

Abstract: In the case of CAPTCHA-based identification on a network, most of the existing methods make use of the difference in perceptual recognition between humans and computers. Researchers have pointed out that it is difficult for visually impaired people to leverage this difference. In this paper, we give our research results which can solve this problem. Since our proposed CAPTCHA generates linguistic questions to avoid relying on specific perceptual abilities, our scheme satisfies “Barrier-free” property. We also give experimental results which contain a case study of medical student who are blind and visually impaired. The results show that the user does not need to have the advantage of information literacy during the identification phase. Therefore we consider that our proposed scheme is suitable for use in barrier-free society.

Keywords: Information Security, Accessibility, Information Barrier-free, CAPTCHA