

Web サイトの動的な変化を考慮した コンテンツ探索手法に関する研究

山本雄平[†] 中村健二[‡] 田中成典^{‡*} 林翔太^{‡*}

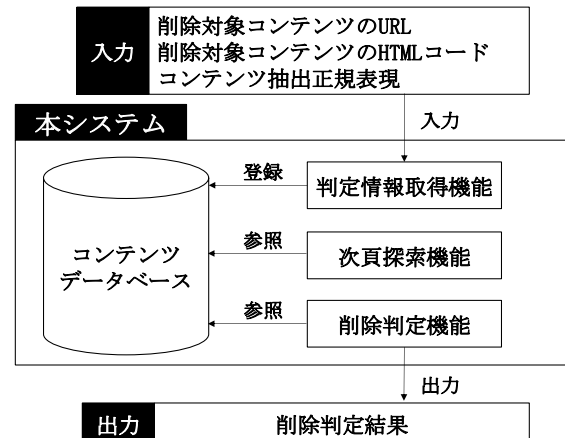
関西大学大学院総合情報学研究科[†] 大阪経済大学情報社会学部[‡] 関西大学総合情報学部^{‡*}

1. はじめに

過去の交際相手に関するコンテンツ（個人情報、写真など）を嫌がらせ目的で SNS（Social Networking Service）や画像投稿サイトなどの CGM（Consumer Generated Media）に公開するリベンジポルノ [1][2] が社会問題になっている。これらのコンテンツは、削除申請があった場合、プロバイダ責任制限法に従って削除されなければならない。2014 年以降、リベンジポルノに関する法整備 [3] が進み、申請から削除までの日数が短縮 [4] された。しかし、被害者の名誉やプライバシーを尊重する一方、投稿者の表現の自由が優先され、削除されない事例 [5] もある。そのため、被害者側が、様々なサイトで削除されたかどうかを確認せざるを得ず、その上、手間もかかる。さらに CGM は、動的な Web サイトであるため、コンテンツがユーザの投稿によりサイト外部へ流出している場合や、1 ページに表示するコンテンツの量に応じて同一 Web サイト内部でコンテンツが別のページへと移動している場合がある。そこで、本研究では、後者の場合を対象として、Web サイト内のコンテンツの動的な変化を考慮し、削除対象のコンテンツを探索する手法を提案する。実験では、動的な Web サイトを構築し、そのサイトにて特定のコンテンツを見つけ出すことができるかを評価する。

2. 研究の概要

本システムの概要を図 1 に示す。システムの入力は、削除対象のコンテンツが含まれるページの URL、該当箇所の HTML コードとコンテンツを抽出するための正規表現とし、出力は、削除判定結果とする。



2. 1 判定情報取得機能

Web サイト内では、削除対象コンテンツは動的に移動するため、削除された場合との区別がつかない。そこで、削除対象コンテンツの前後の HTML コードを削除有無の判定のための手がかりとして用いる。本機能では、まず、削除対象コンテンツの Web ページ内の位置を特定する。次に、削除対象コンテンツの前後の投稿内容の HTML コードを判定情報として取得する。最後に、その HTML コードと入力情報をコンテンツデータベースに登録する。

2. 2 次頁探索機能

動的な Web サイト内では、1 ページに表示するコンテンツの量が増加する場合、複数のページに分割することが一般的である。これらの分割されたコンテンツへは、「次のページへ」や「前のページへ」などのリンクを用いてアクセスできる。本機能では、ページを解析して、入力された URL の前後のページを関連ページとして取得する。

2. 3 削除判定機能

本機能では、判定情報取得機能にて取得した前後の投稿内容を含む削除対象コンテンツの HTML コードを用いて削除されたかを判定する。その削除判定では、次頁探索機能で抽出した関連ページを解析し、前後の投稿内容及び削除対象コンテンツの投稿内容の有無を確認する。前後の投稿内容のみが存在し、削除対象コンテンツの投稿内容が含まれない場合、削除済みと判定する。

Research of Method for Searching Contents Considering Dynamic Web Site

[†] Yuhei Yamamoto

Graduate School of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki City, Osaka 569-1095, Japan

[‡] Kenji Nakamura

Faculty of Information Technology and Social Science, Osaka University of Economics, 2-2-8 Osumi, Higashiyodogawa-ku, Osaka 533-8533, Japan

^{‡*} Shigenori Tanaka, Shota Hayashi

Faculty of Informatics, Kansai University

3. 実証実験と考察

本実験では、システムの有用性を確認するため、削除対象コンテンツの削除の有無を評価する。ここでは、実験のために構築した動的な Web サイトを対象に実施する。様々な Web サイトで本システムが利用できることを確認するため、図 2 に示す 4 種類の Web サイトを構築した。

3. 1 削除判定の評価実験

本実験では、対象の Web サイトにて、削除対象コンテンツの削除の有無を判定する。次に示す 3 つの移動パターンにて、削除対象コンテンツの削除の有無の判定が可能であるかを検証する。

- A) コンテンツが移動していない場合（以下、移動なし）
- B) コンテンツが Web サイト内の別ページへ移動しており、前後のコンテンツが同一ページ内に存在する場合（以下、移動あり）
- C) コンテンツが Web サイト内の別ページへ移動しており、前後のコンテンツが複数のページに跨って存在している場合（以下、ページ跨ぎ）

3. 2 結果と考察

3 つの移動パターンについて、コンテンツが削除された場合とそうでない場合に、削除判定を行った結果を表 1 に示す。実験結果を移動パターンごとに比較すると、移動パターン A) と B) に関しては、コンテンツ削除済みである場合と、コンテンツ未削除である場合の両方で適切に判定できることを確認した。しかし、移動パターン C) に関しては、図 1 の Web サイト(i)と(ii)は適切に判定できたが、(iii)と(iv)では判定できなかった。その原因の一例として、「次のページへ」や「前のページへ」などの前後関係を表すリンクがなく、「1 ページ目」や「2 ページ目」などのように、特定のページへ直接リンクする構造となっている。直接リンクする構造の Web サイトでは、ページ間を関連付けることができず、コンテンツの新旧を判断できないため、削除判定ができない状況であった。この問題に対しては、取得した Web ページ内の投稿日時から、コンテンツの新旧を推定することで、対処できると考えられる。

4. おわりに

本研究では、コンテンツが Web サイト内で動的に移動するページ同士を関連付けてコンテンツを探索する方法を提案した。本提案手法によりページ間が前後のリンクで繋がっている場合、ページを適切に関連付け、削除判定を正しく行



図 2 実験用 Web サイト

表 1 移動パターンごとの削除判定結果

	削除済み	未削除
A (移動なし)	4/4 件	4/4 件
B (移動あり)	4/4 件	4/4 件
C (ページ跨ぎ)	2/4 件	2/4 件

(正解判定数/Web サイト数)

えることを確認した。しかし、特定のページ番号へ直接リンクする Web ページは関連付けることができなかった。今後、このような特徴を持つ Web ページの解析手法を検討し精度の向上を目指すと共に、Web サイト外へとコンテンツが流出するケースについても検討する。

参考文献

- [1] 井樋三枝子：リベンジポルノ等の犯罪化に関する各州立法動向，
<http://dl.ndl.go.jp/view/download/digidepo_8677795_po_02600003.pdf?contentNo=1>, (入手 2015.1.9) .
- [2] 内閣府：青少年のインターネット利用環境に関する制度、法及び政策とその背景，
<http://www8.cao.go.jp/youth/youth-harm/chousa/h25/net-syogaikoku/2_16.html>, (入手 2015.1.9) .
- [3] 衆議院：第 187 回国会総務委員会第 3 号，
<http://www.shugiin.go.jp/internet/itdb_kaigiroku.nsf/html/kaigiroku/009418720141118003.htm>, (入手 2015.1.9) .
- [4] 日本経済新聞：リベンジポルノに懲役 3 年以下の罰則自民、法案提出へ，
<http://www.nikkei.com/article/DGXLASF511H03_S4A011C1PE8000/>, (入手 2015.1.9) .
- [5] 総務省：プロバイダ責任制限法の検証に関して考えられる個別の論点，
<http://www.soumu.go.jp/main_content/000086738.pdf>, (入手 2015.1.9) .