

卒業研究要旨

「いじめ語指数」と個人名検出を併用した学校裏サイト監視支援システム

(知能情報システム学) 伏見 朋恵

1. 緒言

インターネットの利用者は年々増加し続け、若年層も利用する機会が多くなっている。中でも学生や生徒は、自身の HP やブログを開設したり、ソーシャルネットワークに登録したりと、インターネット上での交流が盛んになった。それと共に学校非公式サイト数は2008年3月には全国で3万8260件に達し、インターネット上で特定の人の誹謗中傷を行う「ネットいじめ」の件数も年々増えてきている^[1]。現在は学校非公式サイトを監視する企業も現れているが、費用等の問題で企業に依頼することが出来ず先生が検索・監視を行っている学校も少なくないと言われている。

そのような先生の負担を軽減するため開発された既報の「『いじめ語』検出による学校裏サイト監視支援システム」^[2]をベースとして、本研究では検出された「いじめ語」が人に与えるダメージの程度を表す指数(以下、「いじめ語指数」と表記)を用いて、通報対象となる裏サイト検出の自動化の検討を行った。

2. 開発環境

OS: Windows 7 Professional 使用プログラム言語: Microsoft Visual C# 2008

PC: DELL OPTIPLEX780(CPU: Intel Core 2 Duo E7400 2.8 GHz, メモリ: 2 GB)

3. いじめ語指数の決定

既報の研究^[2]で MeCab^[3]に辞書登録されている「いじめ語」232語を、意味により53グループに分類した。そして、各グループの代表的な「いじめ語」について、人に与えるダメージの程度を計測した。

各グループの代表的な「いじめ語」53語を、大学生10名(20代;男性5、女性5)、高校生7名(男性3、女性4)の計17名に一対比較してもらった。その結果を基に、「いじめ語指数」として各グループに150~450点が与えられた。また、330点以上の「いじめ語指数」の「いじめ語」を、特に注意が必要な「いじめ語」として、「レッドカード」表記対象とし、それ以外の「いじめ語」を「イエローカード」表記対象とした。

4. 個人名の登録

「松茸用人名テキスト」^[4]記載の「個人名」(苗字および名前)を計約80000語、MeCabに登録した。

5. 処理概要

監視対象のURLを、リスト(以下、「URLリスト」と表記)に登録し、以下の処理を行う。

5.1 「いじめ語指数」によるソート

(1) URLリストを読み込み、リストアップされているWebページのHTMLソースをMeCabにかけ、形態素解析を行う。

(2) 解析後、Webページの「いじめ語指数」の合計が高い順にURLが表示される。

各々のTextboxには、[いじめ語]、[いじめ語数]、[いじめ語指数]、[個人名]を表示し、ブラウザウィンドウ上では「レッドカード」表記対象の「いじめ語」に赤、個人名に青のバックカラーを付ける。

5.2 Web ページの分類

以下の、(1)、(2)、(3)、(4)の基準を順に適用して、Web ページを分類する。(1)、(2)、(3)、(4)は、深刻度の高い順になっており、いずれにも該当しない場合は、(5)ブルー(「いじめ語」なし)と分類する。

(1)通報候補

個人名と「レッドカード」表記の「いじめ語」が前後 10 形態素以内で存在していること。もしくは、「いじめ語」から後 10 行以内にある「いじめ語指数」の合計が 900 点以上であること。

(2)レッド

「レッドカード」表記の「いじめ語」があること。

(3)オレンジ

「イエローカード」表記の「いじめ語」があり、そこから 10 行以内にある「いじめ語指数」の合計が 760 点以上であること(この 760 点は、「イエローカード」表記の「いじめ語指数」の平均の 3 倍)。

(4)イエロー

「イエローカード」表記の「いじめ語」があること。

6. 適用例

京都市内の小学校 1 校、中学校 4 校、高等学校 4 校の計 9 校の裏サイト(プロフ、個人サイトも含む)で、パソコンで閲覧可能なサイト 341 件の URL を、URL リストに登録し、以下のテストを行った。

6.1 検出テスト

本研究のプログラムを用いて、過去に通報されたものと同じ文章を実験用掲示板に投稿し、検出を行ったところ、通報された 25 件のうち 21 件が「通報候補」、残り 4 件が「レッド」と判定された。

6.2 監視テスト

URL リストに登録された 341 件について 5 日間(1 日 1 回)監視したところ、「通報候補」と判定された 5 件のうち、実際に通報されたのは 1 件であった。また「通報候補」とならなかったが、通報されたものが 1 件あった。本研究のプログラムでは、この 1 件は、「レッド」と判定された。

7. 結言

登録してあるサイトの「いじめ語」を検出し、誹謗中傷の深刻度を判断するシステムを開発した。今後の課題として、日々増え続けるネット特有の言葉への対応、携帯端末でのみ閲覧可能なサイトへの対応、URL リストの作成の自動化、が挙げられる。さらに、今後も使用者の負担をより軽減するシステムの検討が必要である。

参考文献

- [1]文部科学省、「青少年が利用する学校非公式サイトに関する調査報告書」(2008)
- [2]浅田太郎,竹田麻友子,吉富康成,田伏正佳,『「いじめ語」検出による学校裏サイト監視支援システム」,第 9 回情報科学技術フォーラム講演論文集,3(2010),679-680.
- [3]京都大学情報学研究科ー日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクト, <http://mecab.sourceforge.net/>
- [4]松茸用テキスト, <http://www.vector.co.jp/soft/data/writing/se066955.html>