

「いじめ語」検出による学校裏サイト監視支援システム

(知能情報システム学) 竹田 麻友子

1. 緒言

近年、インターネットの普及により電子掲示板等、インターネット上での交流が盛んになった。それと共に「学校裏サイト」は 2008 年の時点で全国で 3 万 8260 件に及び、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「学校裏サイト」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毎日検索・監視を行っている学校も少なくないと言われている。

本研究では、そのような先生の負担を軽減するため、形態素解析ツール Mecab^[2]を利用して文章解析を行うことでインターネット上に書き込まれている「いじめ語」を検出するシステムを開発した。

2. いじめ語辞書登録

検出対象の「いじめ語」は、書籍やインターネット上から実際使用されている誹謗中傷語を収集し、Mecab 辞書に「いじめ語」として登録した。現在、「いじめ語」は 187 語登録している。(例:ウザイ、キモイ、死ね、等)

3. 開発環境

OS: Windows XP ・ 使用プログラム言語: Microsoft Visual C# 2005

PC: DELL OPTIPLEX760(CPU : Intel Core 2 Duo E7400 2.8 GHz , メモリ : 2 GB)

4. 処理概要

4.1 いじめ語検出機能

- (1)「いじめ語」検出を行いたい電子掲示板等の URL をあらかじめテキストファイルにリストアップしておく。
- (2)ファイルからリストを読み込み、「いじめ語検出」ボタンを押すとテキストファイルにリストアップされている Web ページの HTML ソースを Mecab にかかけ、形態素解析をおこなう。
- (3)解析後、いじめ語数の多い順に URL が表示される。また、個々の URL をクリックすると「いじめ語」(図1 Textbox1)、「いじめ語」数(図1 Textbox2)が表示されると同時に、ブラウザウィンドウ上で「いじめ語」にバックカラーを付け表示する。

4.2 クローラ機能

既報^[3]のクローラ機能を使用することで、ブラウザウィンドウ上に表示されているページからリンクされているページの「いじめ語」検出を行う。リンクを辿っていく深さの設定も可能である。

- (1)クローラ開始ボタンを押し、リンクを辿ったページの URL を Textbox3 にリストアップする。
- (2)「検出ボタン」を押すと、「いじめ語」の多い順に URL の表示、「いじめ語」や「いじめ語」数の表示、バックカラーを付ける作業が行われる。

4. 3. メール送信機能

学校裏サイトから「いじめ語」が抽出された場合、裏サイトに該当する学校にメールで通知する機能である。あらかじめアドレスをファイルに保存している学校は、Mail 作成ボタンを押すことで宛先が自動的に記載される。本文には、「いじめ語」が書き込まれているサイトの URL、いじめ語を記載する(図 2 参照)。



図1:実装フォーム

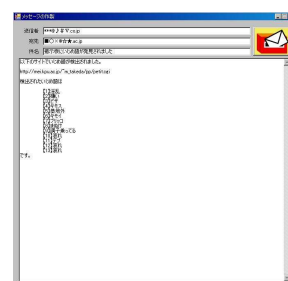


図 2:メール送信フォーム

5. 結言

登録してあるサイトの HTML ソースを解析し、「いじめ語」を抽出するシステムを開発した。本システムの課題として、日々増え続けるネット語に対応していくことが挙げられる。今後は、実用化の課題を明らかにし、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]京都大学情報学研究科-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/>
- [3]磯部 博行,「電子透かしパリティを用いた著作権侵害防止システム」, 京都府立大学院人間環境科学研究所 環境情報学専攻 修士論文(2008)
- [4]子どもたちのインターネット利用について考える研究会(子どもネット研), <http://www.child-safenet.jp/>
- [5]下田博次,「子供のケータイ利用と学校の危機管理」, 少年写真新聞社(2009)
- [6]下田博次,「学校裏サイト」, 東洋経済新聞社(2008)