TOM の改良

03H094 藤川 慎也

1　はじめに

大学に入ると、一般的にレポート課題が増える。他の大学の友人たちも皆レポートのたびに四苦八苦しているが、世の中の学生達は実際に自分で考えた文を書いているのかと疑問に感じた。

大垣斉講師の講義では、講義中に出席確認をするのではなく、講義内容を要約したメイルを送信し、出席確認をする。TOM(Tabun Onagi Mail) はその出席確認メイルの送信時にメイルの文章に禁則文字を使用している場合や、他人のメイル内容をコピーしているなどの違反がある場合に、警告文を送信する機能である。現在の TOM には他人のメイル内容をコピーしている場合に警告文を送信する機能は存在するが、ウェブサイトの内容をコピーしている場合に警告を出す機能がない。今回の研究では、ウェブサイトの内容をコピーしてメイルを書いている場合に警告文を送信する機能を TOM に付加する。文章をコピーしていると、学生が自身で本当に考えた内容を書かなくなり、さらに個性や研究に対する意欲が薄れ、新たな意匠が生まれなくなる。ウェブサイトの文章をコピーし、そのままメイルを書いている場合に警告文を送信することができれば、学生がウェブサイトの文章をコピーすることが減少すると考えられる。

2　現在の TOM について

現在 TOM が警告を出す違反の種類は、5 種類ある。

- 禁則チェック
  行頭に捨て仮名や句読点、閉括弧の使用、行末に開け括弧など禁則文字が使用している場合に警告を出す。
- ワードラップチェック
  文章の行末の 1 つの単語を途中で改行し、2 行に分割した場合に警告を出す。
- 文体チェック
  文章の文体が常体か敬体に統一されていない可能性がある場合に警告を出す。
- 類似警告
  自分のメイル内容と他人のメイル内容と 60 パーセント以上類似している場合は、文章をコピーしてる可能性があるとして警告を出す。
- 変更差分
  前回自分が提出したメイルと今回自分が再提出したメイルの変更差分をパーセントで出し、75 パーセント以上酷似している場合に警告を出す。

3　システムの手順

1. メイルの文章を段落毎に分割し、それぞれの文章をわかち書きにする。
2. わかち書きされた文章から使用回数の多い名詞を取り出し検索語にする。
3. 抽出された検索語を Yahoo!検索 WEB サービスを用いてウェブ検索する。
4. 検索結果からソースを抽出し、文章の部分だけを取り出す。
5. メイルの文章とウェブサイトの文章をわかち書きにする。
6. わかち書きされた文章の区切りの空白に改行を入れ、語ごとに比較する。
7. メイルの文章とウェブサイトの文章の差分を出す。
8. 全体のメイルの行数から文章の差分を引き、それを全体のメイルの行数で割ると、類似度が算出される。

4　まとめ

文章からキーワードを抽出し、Yahoo!検索 WEB サービスを用いウェブ検索を行うプログラムを作成した。このシステムを利用し、ウェブサイトの文章と類似している場合にも警告文が送信される。