

類似レポート検出のための採点支援システム*

嶋村 章子† 矢吹 太朗 佐久田 博司

青山学院大学 理工学部 情報テクノロジー学科‡

1 はじめに

本研究は、課題を与えて受講者がレポートを執筆・提出し評価を受けるという、学校教育でおこなわれることの多いプロセスに関する支援システムの設計・開発、および実データによるシステム評価に関するものである。

近年、情報技術の発達により、課題の提出をウェブ上でおこなうことが多くなった。同時に、他人のレポートやウェブ上の文章を剽窃したレポートも増えている。そのような不正な解答を容易に可能とするネットワーク環境が一般的になったのである。不正レポートは、社会規範からみても、効果的な学習プロセスからみても排除すべきであるが、そのような剽窃レポートを判別する作業は、非常に煩雑である。剽窃レポートをあらかじめ判別することができれば、レポートの質をチェックする実質的な評価作業に時間を費やすことが可能となるであろう。

現在、類似レポートを検出するためのシステムは数多くある。小河らによるシステムでは、ベクトル空間モデルによる類似レポートの一括判定を行っている[1]。また、高橋らによるシステムでは、ウェブサイトからの剽窃を検出することができる[2]。

これらのシステムは類似レポートの検出自体が目的であり、検出結果を判断する教員の作業を補助するものではない。

そこで本研究では、類似レポートの検出に加え、その後の確認や修正作業を容易にする統合的な採点支援システムを提案する。システムは、視覚的にわかりやすい表示や教員の直感的な操作を可能とする仕様であり、より効率的な採点作業の環境を提供する。

2 提案するシステム

本システムは次の要素で構成される(図1)。

1. データベース：レポートデータなどを格納するデータベース

2. サーバーサイドプログラム：類似レポートを検出するサーバーサイドプログラム
3. ブラウザ：類似レポート対を表示する
4. ブラウザ：ユーザの操作をフィードバックする

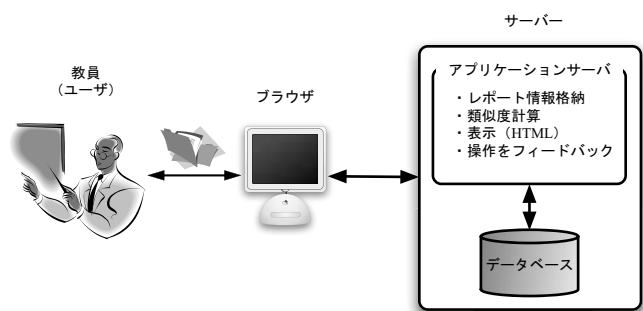


図1 本システムの概要図

2.1 レポート情報を格納する

ブラウザを通してレポート情報をデータベースに格納する。レポート情報とは、課題名や受講者名などの識別情報、レポートの文全体、一文ごとに分けられた文章、品詞ごとに分けられた文章構成要素などである。

2.2 類似レポートを検出する

格納されたレポート情報を用いてレポート間の類似度を計算し、類似レポートを検出する。

類似度の計算には、N-品詞の考え方をを用いる。N-品詞とは、文を一品詞ずつ移動させながら、N 品詞ずつ切り出し比較参照する手法である(表1)。本研究で用いる類似度計算式を下式1[3]に示す。

$$R = 1 - \frac{1}{N} \sum_{i=1}^N \left\{ \frac{P_A(X_i) - P_B(X_i)}{P_A(X_i) + P_B(X_i)} \right\}^2 \quad (1)$$

- R ：類似度
- N ：文字パターンの数
- X_i ： i 番目の文字パターン
- P_A ：文章 A 中の文字パターン X_i の出現頻度
- P_B ：文章 B 中の文字パターン X_i の出現頻度

また、移動させる品詞の数は2とする。これは、再現率、適合率を用いて予備実験によって検証した結果、最適と判断したためである。

* Grading Support System with Similar Reports Detector

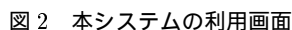
† Akiko SHIMAMURA (shimamura@idea.aoyama.ac.jp)

‡ Department of Integrated Information and Technology, College of Science and Engineering, Aoyama Gakuin University

文章 A:	今日の天気は晴れ
X_1	今日の
X_2	の天気
X_3	天気は
X_4	は晴れ
文章 B:	今日の天気は雨
マッチ	x

課題名や受講者名に対応するレポートおよび類似レポートを表示する．本システムの利用画面を図 2 に示す．

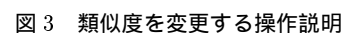
類似している文章は同色で表示され、類似箇所が一目でわかるようになっている。また、文章をクリックことで、類似箇所を水平に並べて表示させることができる。



1. 左上の課題名をクリックすると、その課題のレポートを提出した受講者が表示される。
2. 受講者の名前をクリックすると、受講者のレポートが表示される。その際、類似レポートがあれば、左側に表示される。

ユーザの操作によって類似度の値更新し、表示を変更する。ユーザは、以下の操作を行うことができる。

- 類似度を変更したい文章がある場合、その文章をダブルクリックすることにより、類似候補を排除することができる(図3)。



また、類似度の数値が適切でないと感じられた際には、図 2 における画面左下のスライダーを用いて、類似レポートと判断する基準を変更することができる。

本研究では、教員・採点者の正しい評価を助けることができる、類似レポート検出のための採点支援システムを開発した。機能として、レポートの格納、類似レポートの検出、視覚的にわかりやすい表示、直感的な操作によるフィードバックを実現した。このシステムにより、教員のレポート採点時に剽窃レポートを判別する作業の効率化を図ることができると考えられる。

提出されたレポート内の類似レポート検出は可能となったが，ウェブ上の文章を剽窃したレポートも検出できることが必要と考えられる．コース管理システムと連携させ，利便性を高めることも今後の課題である．

- [1] 小河貴博, 岩堀祐之, 岩田彰: 情報メディア教育における類似レポート判定システムの構築, 平成 13 年度電気関係学会東海支部連合大会講演論文集, 604, p. 304 (2001).
- [2] 高橋勇, 宮川勝年, 小高知宏, 白井治彦, 黒岩丈介, 小倉和久: Web サイトからの剽窃レポート発見支援システム, 電子情報通信学会論文誌 D, Vol. J90-D, No. 11, pp. 2989-2999 (2007).
- [3] 松浦司, 金田康正: 近代日本小説家 8 人による文章の n-gram 分布を用いた著者判別, 情報処理学会研究報告, Vol. 2000, No. 53(NL-137), pp. 1-8 (2000).