

キー入力に基づいた剽窃検知可能な テキストエディタの設計と実装

D-6

Design and Implementation of a Text Editor
Associated with Key Input for Detecting Plagiarism

山崎 和也[†] 高橋 淳二[†] 戸辺 義人[†]Kazuya YAMAZAKI[†] Junji TAKAHASHI[†] Yoshito TOBE[†][†]青山学院大学理工学部情報テクノロジー学科[†]Department of Integrated Information Technology, Aoyama Gakuin University

1. はじめに

本研究では、コピー&ペーストの使用を検知することによって、文字単位で盗用を検出することができるシステム、CODOR(Copy Detection Editor)を設計、実装し、その性能を評価した。

2. 関連研究

本研究の関連研究として[1]が挙げられる。[1]では、与えられた文書を検索にかけ、オンライン上の他文書と照合することで、内容の類似度を調べ、盗用を検出する。本研究は、文書の内容ではなく、文字の入力方法から盗用を検出するという点において異なる。

3. CODOR の設計

CODOR は、文字の入力方法を取得し、各文字と対応付けることができるテキストエディタである。キー操作を監視することで、キーボード入力された文字かコピーされた文字であるかを判定する。

3.1 リストの設計

CODOR は、文字と入力方法を対応付けるために 2 つのリストを保持している。1 つ目は入力された文字を要素とする文字列リスト、2 つ目が各文字のステータスを要素とするステータスリストである。2 つのリストを対応させることで、文字と入力情報の対応付けを行う。

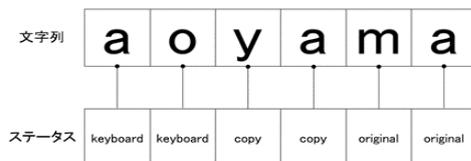


図 1 文字と入力情報の対応付け

3.2 ステータスの種類

入力方法を表すステータスとして、3つの状態を定義する。各ステータスの詳細を表 1 に示す。

表 1 文字ステータスの種類

ステータス名	役割
keyboard	キーボード入力された文字。
copy	他人の文書からコピー&ペーストされた文字。
original	自分で入力した文字を再使用する目的でコピーした文字。

CODOR では、コピー&ペーストの使用を検知して、盗用か否かを判定する。しかし、盗用とは、他人の文書をコピーすることであり、エディタ内で自分が入力した文字を再使用する目的でコピーを行った場合は盗用とは言えない。そこで本研究では、外部からのコピーなら copy、再使用なら original と 2 つのステータスに分け、copy のみを盗用と定義することにした。

4. 実装

CODOR では、コピー検知を行うと、文字のステータスごとに文字の色を変化させ、テキストを画面に再表示する。ステータスと文字の対応は、keyboard が緑、copy が赤、original が黄色とした。また、画面下のステータスバーにテキストの全文字数とコピーされた文字数、コピー率を表示し、どの程度コピーが使用されたかがユーザに一目でわかるようにした。

5. 実験・評価

コピー検知に要する時間を計測した。文字数の異なる 8 種類のテキストを用意し、それぞれ 10 回ずつ計測し、平均を求めた。結果を図 2 に示す。すべてのテストパターンにおいて 1 秒以下で検出を行うことができた。

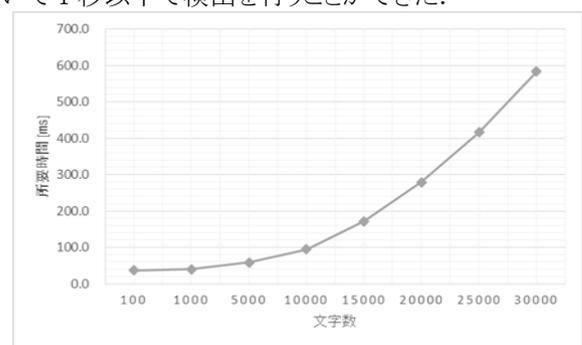


図 2 コピー検知の所要時間

6. むすび

本項では、コピー&ペーストの使用を検知することで、文字単位で盗用を検出することができるシステム、CODOR を提案した。今後は、ファイル操作の高速化と従来技術との組み合わせによる精度向上を目指す。

参考文献

[1] 馬青, 松山秀人, 村田真樹, “Web サイトからの盗作の自動検出システム”, 情報処理学会研究報告(2009)