

レポートの剽窃防止技術についての一考察 ～LMS のツール「Turnitin」の活用事例～

水谷 晃三

帝京大学理工学部情報電子工学科

概要

既存の著作物や他の受講者のレポートから転記したりして、自身のレポートとして提出するような剽窃が行われることがある。剽窃検出を支援する Turnitin と呼ばれるツールが本学 LMS に導入され、試用が開始されている。一方、剽窃検出の技術は発展途上であり、Turnitin などのツールの導入効果を得るにはノウハウの蓄積が欠かせない。そこで本稿では剽窃検出技術の概観を述べるとともに Turnitin の活用事例を述べる。

1. はじめに

課題のレポートをワードプロセッサなどにより作成し、その文書ファイルをレポートとして提出する方法がとられることがある。特に近年は、情報リテラシー教育の重要性も高まっており、学生が作成したレポートファイルをそのまま提出させることは日常的になりつつある。一方、レポートファイルを提出させることは、容易に複製できる性質上、他の学生が作成したレポートの一部を変更して提出したり、インターネット上の著作物をそのまま自身のレポートに転記したりするような剽窃行為の常態化を招き、教育上の問題となる。

そのような背景を踏まえ、本論文では剽窃検出技術について俯瞰するとともに、本学の LMS に導入された Turnitin と呼ばれる剽窃検出の支援ツールの活用事例について述べる。

2. 剽窃検出技術について

Offutt, J. は 8 つの剽窃 (Plagiarism) の型があることを指摘した[1]。そのうち、近年の剽窃検出技術が主な対象とするのは著作物同士を比較することによるものである。

ある著作物に剽窃が含まれるかどうかを検出するには、対象となる著作物中の要素(文, 図, 表など)が他の著作物の要素と同じまたは似て

いることを検出する技術が必要である。近年、図, 表, 式などの類似性を考慮した剽窃検出の研究も行われているが[2], 文の意味的な類似性や, 文章内における文の意味的な役割をラベルづけしていった, これを比較することによる方法など, 文を対象とした研究が主要な観点となっている[3].

文の意味に注目する理由は, 文の構造を単純に比較するだけでは剽窃を検出できないためである。オリジナルの文に対して文意が変わらないように表現を変えたり, 文を分割・結合したりといった剽窃行為に対しては, 文の構造だけではなく文が持つ意味的な類似性を踏まえる必要がある。文の意味を分析したりこれを活用したりする技術は, 情報技術分野としてはチャレンジングかつ過渡的な研究課題の一つである。特に日本語を対象にした研究は英語を対象にしたものに比べて少なく, 更なる発展が期待される。

3. Turnitin の活用事例

3.1 Turnitin の概要

Turnitin は本学の LMS で使用できる剽窃検出の支援ツールである。Word ファイルなどのテキスト部分について, 既存の著作物や Web ページなどの文章との類似性を可視化することができ, この結果を参考に教授者は剽窃の有無を判断することができる。ブラウザ上で可視化結果を閲覧できるほか, 文章内にコメントを加えて学習者にフィードバックするなどの機能が使用できる(図 1)。

3.2 Turnitin の活用事例

① 活用方法の概要

筆者が担当する情報科学科通信課程の科目において Turnitin を使用した。本科目は授業期間中に 2 つの課題レポートの提出を義務付けている。従来の紙ベースのレポート提出と同様に運用できるよう、Turnitin の機能的制限を踏まえて以下のような方策をとった。

- 受講者がレポートファイルを複数回提出した場合、Turnitin は類似性の可視化結果や教授者のフィードバックを含む提出済みのレポートファイルを上書きしてしまう。可視化結果やフィードバックをその後のレポートの作成や教授者側の管理に役立てられるよう、複数の課題提出先の画面を LMS に用意した。
- 通信課程の運用上の取り決めとして、提出されたレポートの採点やフィードバックに期限が設けられている。そこで、複数の課題提出先の画面は 2 週間ごとに提出期限を設けて用意し、それぞれ該当期間中のみ表示されるように設定した。

以上の方策により、ファイルの再提出により上書きが発生する可能性が減り、なおかつ従来の紙でのレポート提出と同様の運用が可能となる。

② 得られた成果、考察

従来の紙ベースの運用ではフィードバックを赤ペンで紙面上に記述していた。Turnitin の画面には印刷イメージを再現した形で提出されたファイルが表示される。画面上に直接コメントを記入できるため従来と同様のフィードバックが可能である。Web ベースのシステムであるため、インターネットに繋がっていればどこでも採点作業が可能である。大量の紙レポートを持ち歩くことが不要になる利点も大きい。

類似性の可視化機能の効果は限定的であった。課題の性質上、インターネット上の記事からの転記がレポートに含まれる可能性が高いが、それが疑われる箇所でも検出されていないケースがあった。しかしながら、他の学生が提出した

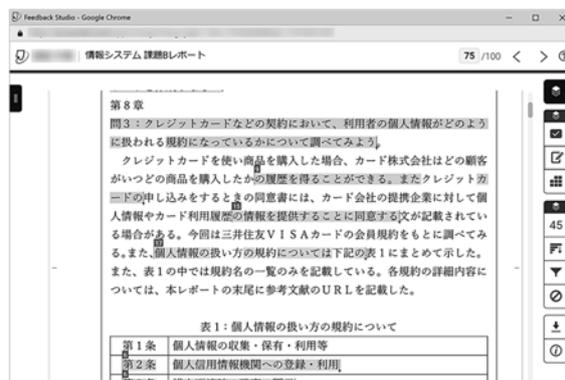


図1 Turnitin による剽窃検出支援の画面例

レポートの一部分をコピーして作成されたものや、ファイル名だけ変えて他の学生と同じファイルを提出してきたような場合については確実に検出された。それを踏まえたコメントを受講者へフィードバックすることで、剽窃行為を未然に防いだり、自ら作文する意欲を喚起したりする効果が期待できたと考えられる。本機能の精度は更なる改良が求められるが、現状でも Turnitin 活用の効果はあり、今後も利用を継続したいと考えている。

4. おわりに

本稿では課題レポートなどにおける剽窃行為に関して、剽窃検出技術を概観するとともに、LMS に導入されている剽窃検出の支援ツール Turnitin の活用事例を紹介した。本技術は過渡的なものであり更なる改良が求められるが、現状でも Turnitin を有効活用できていることを述べた。

参考文献

- [1] Jeff Offutt, “ Editorial: Plagiarism Is For Losers ”, Software Testing, Verification and Reliability, Vol.25, pp1-3, 2018
- [2] Norman Meuschke, Vincent Stange, Moritz Schubotz and Bela Gipp, “ HyPlag: A Hybrid Approach to Academic Plagiarism Detection”, The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp.1321-1324, 2018
- [3] Christina Kraus, “ Plagiarism Detection - State-of-the-Art Systems (2016) and Evaluation Methods”, CoRR, abs/1603.03014, 2016