

英語非母語話者のアカデミック・ライティングにおけるリアルタイム剽窃疑い提示の効果

佐藤 安理紗 ジェンジエラ^{1,a)} 柴田 健吾^{2,b)} 矢谷 浩司^{1,c)}

概要：英語が母国語でない話者は言語的な障壁により、剽窃、すなわち他の文献から不適切な形で言葉を「借用する」という行為を犯してしまうことがある。既存のシステムは既に執筆を終えた原稿に対して剽窃の可能性を提示するように設計されている。そのため、執筆途中においては剽窃の疑いを検出することができず、ユーザは各時点で瞬時に文章を修正することができない。本研究では、事後ではなくリアルタイムで剽窃の疑いを提示することが、英語非母語話者の執筆における剽窃行為を回避する上でより効果的であるという仮説を立てた。英語が母国語でない16名の参加者に学術論文を要約してもらった実験を実施し、剽窃の疑いに関する情報提示のタイミングの違いによる効果を比較した。本論文では、剽窃の疑い箇所をリアルタイムで提示することの利点と欠点を明らかにするとともに、将来の剽窃疑い検出ツールにおけるインターフェースデザインの検討について述べる。

1. はじめに

オンラインソースへのアクセスが増えるにつれて、コンピュータベースでの執筆における剽窃行為を犯しやすくなっている [5, 8, 14, 20, 21]。「剽窃」は、自分以外の者による作品を使用し、自分のものとして広めるという行為を指す。剽窃行為を犯す主な理由としては言語学的な障壁 [3, 7, 10, 11]、文化背景 [3, 11, 15]、執筆経験の欠如 [3] が挙げられる。非母語話者の場合にはさらなる言語学的な障壁に直面することもあるため、より剽窃行為を行いやすいことがある [3, 10]。

一般的な剽窃疑い検出ツールは、執筆者が文章のどの部分を修正すべきかを示し、それにより剽窃がもたらしうる重大な結果（退学処分の警告など）を執筆者は回避できる。しかし、このようなツールの多くは執筆作業を終えた後に使用されることを想定されており、執筆中にリアルタイムで剽窃の可能性を提示するには設計されていない。期限が迫っているなどの理由で執筆者に余裕がない場合には、剽窃の疑いについて確認する時間を取らないこともあ

る。筆者らは、現在のスペルチェック機能による検出と同様、剽窃の疑いについてリアルタイムで提示を行うことにより、執筆している文章の確認や修正を行うように動機づけられるという仮説を立てた。

本論文では、事後提示（執筆が完了した後に行われる確認）と比較し、リアルタイムでの剽窃疑い提示による効果を検討する。このために、16名の英語非母語話者の実験参加者に、両方の検出条件下で学術論文の内容を要約してもらったユーザ実験を実施した。本論文ではその実験結果を基に、リアルタイムでの提示に関する利点や欠点を明らかにし、および非母国語者向けの剽窃疑い提示システムにおいて検討すべきインターフェースデザインのついて述べる。

本研究の貢献は以下の通りである。

- ユーザ実験を通じて、事後の剽窃疑い提示と比較したリアルタイム剽窃疑い提示による効果の検証
- リアルタイム提示の利点と欠点の考察
- 剽窃疑い提示システムのインターフェースデザインの検討

2. 関連研究

2.1 剽窃疑い提示の効果

剽窃の疑いを提示することが執筆者に対するよいフィードバックになることが過去の研究から知られている。Akçapınar は学期の中間時点で指摘を受けた学生の83%において剽窃の割合が下がったことを明らかにした [1]。

¹ 東京大学大学院 工学系研究科
Interactive Intelligent Systems Laboratory,
Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

² ジュネーブ大学
University of Geneva
Geneva, Switzerland

a) arissa@iis-lab.org

b) kengo@iis-lab.org

c) koji@iis-lab.org

Schneider et al. は入力されたテキストに加えてユーザ行動データ（キーボードやマウスの動作、コピーや貼り付けなどの編集操作）も活用する剽窃検出手法を開発した [18]。プログラミングタスクを伴うユーザ実験を通じて、提案手法は多量の変更が施されている場合（大量のコードのリファクタリングなど）において剽窃行為を特定できることが分かった。Landau et al. は、剽窃に関するテストを受け、その結果を見た学生は、その後の調査において剽窃をより特定できるようになったこと報告している [12]。また、剽窃の例を見た学生の方が剽窃を行う可能性が低くなることも発見した。さらに、剽窃を回避するために罰を課すことは、剽窃行為を減らす上で最も効果が低いことが分かった。Lukashenko et al. は剽窃疑い提示と防止についてコンピューターベースの手法を検討し、それぞれについて短期的および長期的な良い効果をまとめている [13]。

上述の研究は剽窃疑い提示の良い効果を示している一方、これらの研究プロジェクトではどれも執筆活動と切り離して調査を行っている。本研究の主な目的は、執筆中に剽窃疑い提示を行うことによる効果の検証であり、それによってインタフェースデザインの指針を示すことにある。

2.2 剽窃疑い検出システム

剽窃の疑いを特定するアルゴリズム技術は計算言語学の研究にて多く研究されている [6,14]。Niezgoda et al. は文書をスキャンし剽窃の疑いがある候補箇所を特定するソフトウェアツール SNITCH [17] を開発した。用いられるアルゴリズムは素早く、かつ偽陽性率を低く抑え、高い精度を維持している。同様に、Sochenkov et al. は単純なコピーと貼り付けによる文や適度に修飾されただけの文言を発見するために深い構文解析を活用する Exactus Like [19] を開発した。Shochonkov らによる手法は、剽窃された文章がどの参考文献から引用された疑いがあるのかといった検出も行える。Mozgovoy et al. はファイルを絞り込む高速検出と執筆された文章を徹底的に分析する精密な検出機能を組み合わせることにより、剽窃疑い検出の速度を上げるアルゴリズムを開発した [16]。Wan et al. はハードウェアのプログラミング課題においてリアルタイムで剽窃を検出するため、トークンに基づいたアプローチを提案した [22]。Wan らはトークンに基づいた検出により、過剰なコストをかけずに増加していく剽窃の検出も実現可能であることを示唆している。

上述の研究による貢献は主にアルゴリズムに関するものであるのに対し、本研究の貢献は剽窃疑い提示ツールやシステムにおける新しいインターフェースデザインに関するものである。

2.3 剽窃防止のためのデザイン

適切な剽窃防止方法は学生が取り組む演習課題の種類

に依存することが知られている。Bakhtiyari et al. は剽窃を防止する上で倫理的な方法と非倫理的な方法について記述している [2]。アカデミック・ライティングに関して、Bakhtiyari らは多量の文献を読み込む、文献の閲読と執筆との間に時間を確保する、また頻繁に校正する、などを含む倫理的な方法をいくつか提案している。Sterngold は小分けにして出題される宿題、講座に特化したライティング課題、また研究について議論するための学生とのミーティングなど、剽窃を減らす課題のあり方を複数提案している [21]。Halak et al. は工学部の宿題について、学生ごとに独自に課される宿題と講座に基づいたプレゼンテーションといった2種類の手法を提案している [9]。Niezgoda et al. はユーザが執筆した内容を確認し修正するように動機づけるソフトウェアツールの実現は予防策の一つであることを強調している [17]。

本研究はリアルタイムで剽窃疑いを提示する利点を示すことで、これら上述の研究で示している演習課題設計のガイドラインを補完する。

3. 剽窃疑い提示手法の比較実験

リアルタイム（検出箇所を執筆中に提示）と事後（執筆完了後に検出箇所を提示）の剽窃疑い提示において、執筆行動とユーザエクスペリエンスに見られる違いを比較する実験を実施した。なお、以下の実験は大学における研究倫理審査委員会による承認を事前に受けている。

3.1 タスク

実験で行うタスクは、2通りの検出条件のうち片方の条件下において短い学術論文を読み、25分以内に最低300語で内容を要約するというものであった。実験のタスク設計は、複数回に渡る事前試行に基づき、ヒューマン・コンピューター・インタラクション (HCI) の分野で査読済みの会議論文として公開されている1,800から2,500語程度の短い論文を用いることにした。本実験の参加者の言語力は中級レベルであったことから、参加者が外部情報源（オンライン辞書など）を用いずに論文の内容を理解できることを担保するため、以下の方法で論文を選定した。

まず論文ごとの語彙の違いを考慮するため、lextutor^{*1}という語彙プロファイル機能を用いて語彙の分布を確認した。語彙プロファイルにはK1、K2、アカデミック・ワード・リスト、およびオフリスト・ワードの割合が含まれる。K1とK2は英語においてそれぞれ最もよく使われる上位1から1,000語と1,001から2,000語のことである。可読性と語数に関する基準を満たしている会議論文を抽出した後、K1スコアが67-78%の範囲である論文を数本選定した。その次に語数が2,100-2,200語程度となり、K1スコアが72%で

*1 <https://www.lex tutor .ca/vp/eng/>

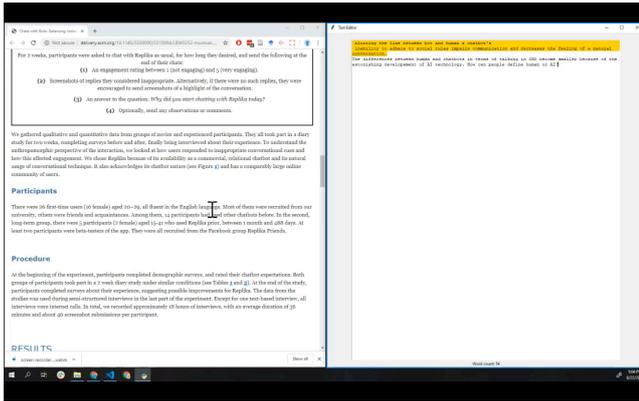


図 1: 実験に用いたエディタのインターフェース。実験参加者は左側の画面で参照文書を確認し、右側のテキストフィールドにおいて要約を行った。システムによって剽窃の疑いが検出された場合は、当該箇所を黄色でハイライトした。

ある HCI 論文を 2 本選定し、本実験で用いる文書とした。

3.2 実験に用いたシステム

2 通りの提示条件の違いを検討するため、剽窃疑い提示機能を搭載した独自のテキストエディタを開発した。

3.2.1 テキストエディタのインターフェース

Python の tkinter ライブラリ *2により開発した。実験に用いたエディタを図 1 に示す。このエディタは執筆に必要な基本的な操作や機能を有している (コピーと貼り付け、元に戻す、やり直す、ファイルの自動保存、文字数カウントなど)。

3.2.2 剽窃疑い検出機構

本研究では、剽窃疑い提示をリアルタイムで実行できるように、実装システムでは参照文書よりそのまま引用している場合のみを剽窃として扱った。システムはまずユーザが執筆した文章に Punkt tokenizer *3を適用し、ピリオドに基づいて各々の文を特定する。次に、ユーザが執筆した文 1 つ 1 つを参照論文と比較して、完全一致箇所を探す。完全一致箇所が見つからない場合は、検索に用いた各文に対して類似性スコアを算出する。文の類似性は difflib SequenceMatcher library *4を用いて計算する。現在の実装では文の類似性スコアが 0.7 以上であれば剽窃の疑いがあると見なしている。SequenceMatcher のドキュメンテーションによると、値が 0.6 であると概ね一致しているものと見なされる。偽陽性を回避するため、この閾値を手動で検証し、実装システムでは 0.7 に設定した。さらに SequenceMatcher は、多少の誤字にも対応できることが事前の調査でわかった。

剽窃の疑いがある全ての文には黄色のハイライトを施される (図 1)。これは、既存の剽窃疑い提示ツールで用い

られている視覚デザインを模したものである [4, 19]。ユーザが適切な修正を施し、剽窃の疑いがあると見なされなくなるまで、ハイライトは提示され続ける。リアルタイム提示の条件下では、ピリオドを入力 (文の終わりを示唆) するか、カーソルを移動 (ユーザの注意が文章の他の部分に移動したことを示唆) したことを執筆行為の区切りとして認識し、その度に剽窃疑い検出を実行する。事後提示の条件下では、ユーザが「確認」ボタンを明示的に押すことによるのみ剽窃疑い検出が行われる。

3.3 条件

リアルタイム提示と事後提示による効果の比較は同じ実験参加者に対して行った。3.1 節で述べた通り、参照文書には 2 本の学術論文を選定した。順序効果を回避するため、検出条件と節の組み合わせにはラテン方格法を用いた。また不要な変動を避けるため、提示するタイミングを除き、同じ検出アルゴリズム (3.2.2 を参照) を両条件において用いた。

3.4 実験参加者

実験参加者は、20 歳から 30 歳までの大学生を 16 名 (P1 - P16; 女性 7 名、男性 9 名) であった。全員英語を母国語としておらず、自己申告による英語の読み書きのスキルは中級レベルであった。各実験参加者に実験終了後 2,000 円の謝金を支払った。

3.5 実験手順

実験参加者に対して、まず実験説明書を読み、同意書に署名するよう求めた。次に、英語での読み書き経験について情報を得るため、実験参加者にインタビューを行い、さらに本研究のシステムと要約タスクについて説明した。タスクやシステムに慣れてもらった後、与えられた条件と文章に対して上述の要約タスクを実施してもらった。実験中は執筆に集中してもらうため、実験参加者によるインターネットの使用を禁止した。時間の経過を認識できるように、残り時間を 10 分、5 分、および 1 分の時点で告知した。

また、実験参加者には次のタスクを開始する前に 5 分間の休憩時間を与えた。両タスクを終えた後、インターフェースに関しての意見を聞き取るために再びインタビューを行った。なおこのインタビューは英語にて行なった。

3.6 データ収集

後で分析できるように、キーボードおよび剽窃疑い検出動作のログを全て保存した。さらに、ログが正しいことを担保するため、実験から得られたスクリーンキャプチャも記録した。また、全てのインタビューを録音し、書き起こしを行なった。本論文での報告のため、以下で引用するインタビューは、可能な限り正確な翻訳とするために、プロ

*2 <https://docs.python.org/3/library/tk.html>

*3 <https://www.nltk.org/api/nltk.tokenize.html#module-nltk.tokenize.punkt>

*4 <https://docs.python.org/2/library/difflib.html>

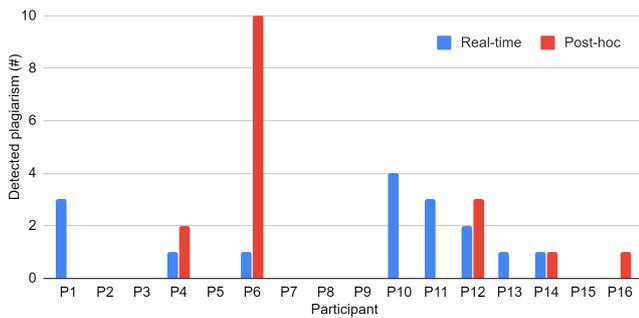


図 2: 各実験参加者における検出の発生回数。6 名の実験参加者 (P1、P9、P10、P11、P13、P15) は事後検出条件下で時間内に剽窃疑い提示機能を実行できなかった。

の翻訳サービスによる翻訳を行なったのち、さらに推敲したものである。

4. 結果

4.1 提示の発生回数

図 2 は本研究の実験における剽窃疑い検出の発生回数を示している。剽窃の疑いを指摘された実験参加者のうち、両条件において起きた提示の平均回数はそれぞれ 2.0 ($SD=1.19$) と 3.4 ($SD=3.78$) であった。事後検出条件下において、6 名の実験参加者 (P1、P9、P10、P11、P13、P15) は剽窃検出を実行する時間を確保できなかった。このことから、剽窃検出が容易に忘れられたり後回しにされてしまったりすることが確認できる。

4.2 リアルタイムでの提示の利点と欠点

2 通りの提示条件において観察されたユーザエクスペリエンスを分析するため、インタビューから取得したコメントに対して分類を行なった。特にリアルタイムでの提示における利点や欠点を示すコメントがいくつか得られた。

2 名の実験参加者は、リアルタイムでの提示により剽窃として検出された文章をすぐに確認し、文脈を見失う前に修正できたことに明示的に触れた。実験参加者 P3 は実験中にリアルタイムでの提示を受けることはなかったが、実験後のデモを見た後にその良さを確認し、以下のように述べた。

「リアルタイムでの提示だと自分が何を書いているのか考えている時点で黄色 [のハイライト] が現れるので、より良いと思います。それより後か前に現れると、考え直さなれないといけないかもしれません。」 [P13]

「リアルタイムの方が助かります。とても長い論文や文章を書く時のことを考えると、問題がある場合はできるだけ早く修正できるようにその時点で確実に知りたいからです。」 [P3 (インタビュー時に実験者によるデモを見た上での感想)]

本研究で観察されたもう一つのメリットは、執筆後に実験参加者が文章を再確認する必要がないという点であった。実験後に実験者によるデモを見た実験参加者 P7 は次のようにコメントしている。

「ウェブサイトなどにテキストを張り付けたりして確認する必要がなくなるので、[これらのシステムの] 方が便利だと思います。」 [P7 (インタビュー中に実験者によるデモを見た上での感想)]

一方で、実験参加者はリアルタイムでの提示が執筆過程における妨げになる可能性があることも感じていた。P15 は実験者によるデモを見た後にリアルタイムでの提示が執筆者にプレッシャーを与えてしまう可能性を指摘した。

「このシステムでは、執筆中に [ハイライトを見て] 突然、[ハイライトに] 目が奪われてしまい書き続けることができなくなりそうです。書き続けることができなくなってしまうと思います。『一旦別のことを書かなくてはいけない』と思うでしょう。どれだけ修正が必要か、またそれが頻繁に現れると [執筆者に] 影響するでしょう。[ハイライトが] たくさん現れれば、[執筆] を続けられないと感じるかもしれません。[P15 (インタビュー中に実験者によるデモを見た上での感想)]

また、実験参加者らは提示を遅延させたり非表示機能を追加することで、妨げといった悪影響を緩和できる可能性を示唆した。

「リアルタイムの方が便利ですが、私の思考プロセスの妨げになってしまうことがあります。両方の検出システムにメリットとデメリットがあると思います。例えば 3 から 5 分後に確認してくれるようなタイマーを設定できたら良いと思います。[ハイライトに] 集中してしまうと、何を書こうとしていたのか忘れてしまうので、その方が邪魔にならないと思います。」 [P4]

「上の方に小さく通知を出して問題があることを知らせてくれて、いったん無視して後で修正できるようにしたら良いでしょう。」 [P2 (インタビュー中に実験者によるデモを見た上での感想)]

4.3 事後提示機能の未実行

上述の通り、6 名の実験参加者が時間内に剽窃疑い提示機能を実行することができなかった。この条件で剽窃疑い検出を実行した実験参加者 P7 も、剽窃について確認することを忘れかけていたとコメントしている。実験参加者

P8 と P15 も、事後提示のデモを見た後に同様の意見を述べた。

「事後提示では毎回クリックしなくてはならず、忘れてしまうこともありました。ボタンを押さなきゃいけないことは事前に説明してもらっていましたが、すっかり忘れてしまいました。」 [P7]

実験参加者 P9 は論文執筆など時間的な制約が大きい状況下では、事後提示よりもリアルタイム提示の方が執筆者にとって助かるとはっきり述べた。

「自動的に提示をしてもらえると、その時点で言葉を変更できるので良いです。それで [ハイライト] されない (剽窃の疑いがない) 新しい文を書くことができます。なぜ助かるのかというと、論文を執筆しているときは、誰もが期限に追われています... なので、自動的に [提示を受けられること] で、考えていることと書いていることをより結び付けられます。私はその方がユーザにとってよいと思います。」 [P9 (インタビュー中に実験者によるデモを見た上での感想)]

5. 考察とインタフェースデザインの検討

本実験を通じて、リアルタイムで剽窃の疑いを提示することについて以下の利点と欠点を見出した。

- リアルタイムでの提示は、執筆者が文脈を見失う前に即座に修正を行うことを可能にする。
- リアルタイムでの提示は、時間的な制約が大きい条件下において、剽窃の確認を促すのに効果的である可能性がある。
- 既存の剽窃疑い提示ツールにおいて使われているハイライト方法は、リアルタイムでの提示では執筆の妨げとなりうる。

本研究の実験参加者は一般的にリアルタイムでの提示について肯定的であったが、気が散ってしまう点に関しては慎重な検討が求められる。特に、執筆と剽窃確認の2状態を交互に行き来することを不要にすることが上記に要約した2つの点に集約されている。非母語話者は、執筆においてどのように書くべきかわからず、文章を写してしまうことなどを容易におこなうことがある。リアルタイムでの剽窃疑い提示はどのように文章を書くか執筆者に考えるように促し、執筆活動を支援するものとなりえる。

本研究で確認されたリアルタイム提示の主な欠点は気が散ってしまう点であった。定性的な結果に見られるように、気が散ってしまう主な原因はテキストのハイライトという視覚デザインであった。本研究では既存の剽窃疑い提示ツールにおいて一般的に利用されている、ハイライトという視覚デザインを採用した。しかし、今後のインター

フェースにおいてはより視覚的に邪魔しないデザインを採用すべきであることが実験結果からわかる。例えば、検出した剽窃回数のカウンター (語数カウントに似た機能)などを提示し、ユーザがフィードバックの内容を見たいというインタラクションを明示的にした場合にのみ (例えばカーソルのホバリングなど)、詳細を提示する。

本研究における定性的な結果は、提示のタイミングも執筆の妨げとなる可能性を減らすために重要であることを示唆している。P4 が提案した一定の時間遅延の他、執筆者がまとまった章を書き終えたことを検知したり、執筆者が修正可能なタイミングを自動的に検出したりするといった方法も妨げを緩和する改善策となりうる。

6. 実験および考察に関する注意点

本研究ではリアルタイムでの応答性 (例えば全く同じ内容を剽窃している場合のみを対象としている、など) を実現するため、限定的な剽窃疑い提示機能のみを実装した。そのため、リアルタイムでの剽窃を搭載したテキストエディタの開発は、本研究による貢献とするものではない。同様に、本研究の主な焦点は検出アルゴリズムの速度や精度の向上でもない。本研究の結果はリアルタイムでの提示の利点を示し、リアルタイムでの剽窃疑い検出のアルゴリズムの研究を促すものである。

実験結果において提示された剽窃の数が少なかった理由の一つとしては、タスクの設定が剽窃を必ずしも十分に促すものではなかったことがある。本研究における実験参加者のうち、明らかな剽窃が検出されたのは約半数のみに留まった。全般的に実験参加者らは剽窃がアカデミック・ライティングにおいて深刻な不正行為であることを理解しており、実験参加者の剽窃行為を促すことは根本的に難しい。これまでの研究においては、剽窃に関するユーザ実験を対象とした標準的な手法が確立しておらず、効果的な実験計画も今後の研究において探究されるべきである。

7. 結論

剽窃はアカデミック・ライティングにおいて深刻な課題となっているなか、執筆者を教育し、剽窃の可能性を回避するように促す効果をもたらさう新しい執筆インターフェースが求められる。本研究では剽窃の疑いがある箇所をリアルタイムで提示することの効果と比較実験により検証した。実験結果からは、リアルタイムでの提示による重要な利点を2点、改善事項を1点確認した。また、自然言語処理とインターフェースデザインの両方において、重要な研究の方向性をいくつか見出した。

謝辞

本研究、および本論文に助言を頂いた Carla F. Griggio、D. Antony Chacon、Zefan Sramek、Zhongyi Zhou、Xu

Hao, 佐野 翔子、正木 博明、松井 秀憲、Toby Chong Long Hin に感謝申し上げます。また実験参加者の皆様に感謝申し上げます。

参考文献

- [1] Akçapmar, G.: How automated feedback through text mining changes plagiaristic behavior in online assignments, *Computers Education*, Vol. 87, pp. 123–130 (online), DOI: 10.1016/j.compedu.2015.04.007 (2015).
- [2] Ale Ebrahim, N., Bakhtiyari, K., Salehi, H., Embi, M., Shakiba, D. M., Zavvari, A., Moghdam, M. and Mohammadjafari, M.: Ethical and Unethical Methods of Plagiarism Prevention in Academic Writing, *International Education Studies*, Vol. 7, pp. 52–62 (online), DOI: 10.5539/ies.v7n7p52 (2014).
- [3] Amsberry, D.: Deconstructing Plagiarism: International Students and Textual Borrowing Practices, *The Reference Librarian*, Vol. 51, No. 1, pp. 31–44 (online), DOI: 10.1080/02763870903362183 (2009).
- [4] Barrie, J.: Turnitin (2019).
- [5] Bin-Habtoor, A. S. Y. and Zaher, M.: A Survey on Plagiarism Detection Systems, *International Journal of Computer Theory and Engineering*, pp. 185–188 (online), DOI: 10.7763/IJCTE.2012.V4.447 (2012).
- [6] Chowdhury, H. A. and Bhattacharyya, D. K.: Plagiarism: Taxonomy, Tools and Detection Techniques, *CoRR*, Vol. abs/1801.06323 (online), available from <http://arxiv.org/abs/1801.06323> (2018).
- [7] Dong, Y. R.: Non-native Graduate Students' Thesis/Dissertation Writing in Science: Self-reports by Students and Their Advisors from Two U.S. Institutions, *English for Specific Purposes*, Vol. 17 (online), DOI: 10.1016/S0889-4906(97)00054-9 (1998).
- [8] Goddard, R. and Rudzki, R.: Using an Electronic Text-Matching Tool (Turnitin) to Detect Plagiarism in a New Zealand University, *Journal of University Teaching and Learning Practice*, Vol. 2, p. 8 (2005).
- [9] Halak, B. and El-Hajjar, M.: Plagiarism Detection and Prevention Techniques In Engineering Education, (online), DOI: 10.1109/EWME.2016.7496465 (2016).
- [10] Hayes, N. and Introna, L.: Systems for the Production of Plagiarists? The Implications Arising from the Use of Plagiarism Detection Systems in UK Universities for Asian Learners, *Journal of Academic Ethics*, Vol. 3, pp. 55–73 (online), DOI: 10.1007/s10805-006-9006-4 (2005).
- [11] Heitman, E. and Litewka, S.: International Perspectives on Plagiarism and Considerations for Teaching International Trainees, *Urologic oncology*, Vol. 29, No. 1, pp. 104–108 (online), DOI: 10.1016/j.urolonc.2010.09.014 (2011).
- [12] Landau, J. D., Druen, P. B. and Arcuri, J. A.: Methods for Helping Students Avoid Plagiarism, *Teaching of Psychology*, Vol. 29, No. 2, pp. 112–115 (online), DOI: 10.1207/S15328023TOP2902_06 (2002).
- [13] Lukashenko, R., Šakele, V. and Grundspenkis, J.: Computer-based plagiarism detection methods and tools: An overview, Vol. 285, p. 40 (online), DOI: 10.1145/1330598.1330642 (2007).
- [14] Maurer, H., Kappe, F. and Zaka, B.: Plagiarism - A Survey, *Journal of Universal Computer Science*, Vol. 12, No. 8, pp. 1050–1084 (online), available from http://www.jucs.org/jucs12s/plagiarism_a_survey (2006).
- [15] Mcgowan, U.: *International students: A conceptual framework for dealing with unintentional plagiarism*, pp. 92–107 (online), DOI: 10.4018/978-1-59904-801-7.ch007 (2007).
- [16] Mozgovoy, M., Karakovskiy, S. and Klyuev, V.: Fast and reliable plagiarism detection system, pp. S4H-11–S4H-14 (online), DOI: 10.1109/FIE.2007.4417860 (2007).
- [17] Niezgoda, S. and Way, T. P.: SNITCH: a software tool for detecting cut and paste plagiarism, *SIGCSE* (2006).
- [18] Schneider, J., Bernstein, A., Brocke, J. v., Damevski, K. and Sheppard, D.: Detecting Plagiarism based on the Creation Process, *IEEE Transactions on Learning Technologies*, Vol. 11, pp. 348–361 (online), DOI: 10.1109/TLT.2017.2720171 (2017).
- [19] Sochenkov, I., Zubarev, D., Tikhomirov, I., Smirnov, I., Shelmanov, A., Suvorov, R. and Osipov, G.: Exactus Like: Plagiarism Detection in Scientific Texts, Vol. 9626, pp. 837–840 (online), DOI: 10.1007/978-3-319-30671-1.76 (2016).
- [20] Šprajc, P., Urh, M., Jerebic, J., Trivan, D. and Jereb, E.: Reasons for Plagiarism in Higher Education, *Organizacija*, Vol. 50 (online), DOI: 10.1515/orga-2017-0002 (2017).
- [21] Sterngold, A.: Confronting Plagiarism: How Conventional Teaching Invites Cyber-Cheating, *Change: The Magazine of Higher Learning*, Vol. 36, pp. 16–21 (online), DOI: 10.1080/00091380409605575 (2010).
- [22] Wan, H., Liu, K. and Gao, X.: Token-based Approach for Real-time Plagiarism Detection in Digital Designs, *2018 IEEE Frontiers in Education Conference (FIE)*, pp. 1–5 (online), DOI: 10.1109/FIE.2018.8658531 (2018).