

UDC 004.89:347.78OSADCHYI V. I.^{1*,2*}^{1*} Scientific Library, Ukrainian State University of Science and Technologies (Dnipro, Ukraine),^{2*} Translation.in.ua (Dnipro, Ukraine),

e-mail: expert@translation.in.ua, ORCID 0000-0003-4242-6328

OSADCHA O. V.

Oles Honchar Dnipro State University (Dnipro, Ukraine), e-mail: elena.osadcha@gmail.com,

ORCID 0000-0003-1414-6345

Artificial Intelligence and Machine Learning Algorithms for Assessing the Authenticity of a Scientific Article in Scopus: Translator's Experience

Objective. This paper examines ways to solve the problem of cross-language plagiarism in scientific works written in Ukrainian, which are to be translated and published in English. Considering that Ukrainian university libraries are directly involved in the practices of improving the level of awareness of lecturers and scientists, as well as their support of a large number of new digital tools, we draw attention to the emergence of new opportunities in the practices of supporting academic integrity. **Methods.** Big Data mining techniques and analysis of algorithms underlying machine translation software were employed to identify the cases of cross-language plagiarism in scientific articles originally written in the Ukrainian language. **Results.** Based on the analysis of 4000 translated manuscripts, it was established that the standard Microsoft Word 2022 software, typically used to write an article, identifies with a very high accuracy those parts of the text that had been earlier published and stored in a digital format. **Conclusions.** With the advent of Microsoft Office 365 software (released in 2022), it becomes possible to check any article originally written in Ukrainian or Russian, while being translated into English, for similarities with previously published academic papers. This allows for an instantaneous correction check that may prove useful in preventing the intended or unintended occurrence of cross-language plagiarism in scientific papers. It is advisable to more actively involve librarians of Ukrainian universities in using the powerful potential of digital support for the research activities of their users, including writing papers and checking them for signs of plagiarism.

Keywords: cross-language plagiarism; university libraries; academic paper in the Ukrainian language; Scopus; Microsoft Office 365 software; translation from Ukrainian into English

Introduction

Scientific libraries of universities of Ukraine, as well as other countries, are directly involved in the practices of improving the level of awareness of lecturers and scientists, as well as their support of a large number of new digital tools that appear in new forms of conducting scientific research and disseminating their results. Practices for evaluating the level of publishing activity and maintaining academic integrity are also in the focus of attention of libraries.

Pressed by state-initiated requirements, a University lecturer's career in Ukraine strongly depends on the number of scientific papers published annually in prestigious international journals. The content of such papers (even peer-reviewed) is typically not discussed publicly, which leaves some aspects related to writing a manuscript immune to thorough analysis in terms of the dubious practice of borrowing a sentence/paragraph from foreign literary sources without including the author in the reference list.

Given the mandatory checks for plagiarism at any higher educational establishment, manuscripts in Ukrainian are subject to such testing and are presumably devoid of cases of academic integrity violations. However, once a paper is to be submitted to an international journal and is to be translated from Russian/Ukrainian into English, a simple trick of borrowing a paragraph in English from any source and translating it into Russian/Ukrainian to innocently insert it into a manuscript for a peer review (by Russian/Ukrainian reviewers) and then, after it is

accepted, back to English, may work miracles in terms of compiling a text of any length required by a publisher. All that the "author" should do is change some words to cheat specialized software designed to identify an exact match or similarity.

Literature analysis

The adoption of the Law of Ukraine "On Education" in 2017 formally introduced the concept of "academic integrity" (Verkhovna Rada of Ukraine, 2017). According to Clause 4 of Article 42 of this Law, a violation of academic integrity includes academic plagiarism; self-plagiarism; fabrication; falsification; cheating; deception; bribery; biased assessment.

Since then, the endless theorizing about academic integrity has not, unfortunately, yielded any real examples of identified violations, which can create the impression of absolute transparency in Ukrainian science. A rudimentary Google search for the keywords "example of academic integrity", "violation of academic integrity", and "academic dishonesty" does not give a single local example.

In the IT industry, the specified issue is referred to as external duplicate content. As noted by Horst (2022), external duplicate content implies that the same text is found on multiple domains. The Google search engine has a dedicated resource to tackle this issue with developers by introducing the concept of a canonical website locator (Google Developers, 2022), which essentially traces the primary source of information.

As regards scientific papers, there is the concept of cross-language plagiarism, which, according to Ouriginal (2021), refers to the kind of plagiarism or cheating where the source content is in one language while the plagiarized content is in another. In other words, the authors argue, it is plagiarism by translation. There is a growing body of research on the topic, specifically, when detecting cross-language English-Arabic plagiarism (Hattab, 2015; Alaa, Tiun, & Abdulameer, 2016), or when identifying duplicate publication (Rohrich & Sullivan, 2009). There is also a very detailed study reported by Pryimak (2019) about plagiarism in English-, French-, and Italian-Ukrainian dictionaries.

IEEE defines plagiarism as the reuse of someone else's prior processes, results, or words without explicitly acknowledging the original author and source (IEEE, 2022). According to this definition "plagiarism in any form is unacceptable and is considered a serious violation of professional behavior with ethical and legal consequences".

Another aspect of this shameful phenomenon, as noted in Plagiarism in Research (2020), is cross-lingual plagiarism, where the original content is in a language different from the language of the plagiarized text. In scholarly publishing, this is a growing concern, the authors claim, since auto-translation tools make it easy to copy ideas/text from an already published paper and translate it into a different language.

Aims

We assume that the corporate ethics of Ukrainian science encourages unscrupulous practice, particularly in scientific publications. It is this aspect of scientific activity that is addressed in the current paper.

The hypothesis under consideration is as follows: when writing a paper using Internet search engines, it is tempting to include, for example, in a literature review not just a primary source but a phrase (sentence, paragraph) of another author while passing it off as your own. The situation is extremely simplified if the original manuscript is written in Russian or Ukrainian and

the article is translated into English, or vice versa. Indirectly, this is confirmed when the author's usual style of presentation suddenly changes to exquisite academic English.

Therefore, this study aims to identify ways to detect cases of cross-language plagiarism when a paper originally written in Ukrainian/Russian is translated into English.

Methods

The above hypothesis is based on our personal practical experience of translating 4,000 peer-reviewed articles by Ukrainian, Kazakh, and Azerbaijani scientists into English for confirmed publications in journals indexed by Scopus in the period from 2016 to 2022.

Given the specificity of this activity, we have read all the papers from beginning to end, including abstracts, references, and assurances of the full reliability of the results.

Taking into account the conditions for prolonging lecturers' full-time employment contracts at Ukrainian universities, academic staff are forced to report annually on the number of papers they publish in internationally acclaimed scientific journals. Note that the current study excludes articles written "for a report on science" (endless repetitions, scientific clichés, a precisely adjusted number of pages) and commissioned articles (lengthy fiction about concepts that often simply do not exist). This is a promising area for further research into revealing dishonest practices given that such papers, in our rough estimation, account for at least 5 % of manuscripts submitted.

Results and Discussion

Until August 2022, we did not have the opportunity to confirm the above hypothesis, even after repeatedly contacting the relevant services at Microsoft Translator and Google Translate for clarification. Our concern is absolutely similar to the view expressed in *Plagiarism in Research* (2020) that "some plagiarism detectors do offer cross-lingual plagiarism checks while it is unclear if they are as effective at detecting this form of plagiarism as they are at detecting plagiarism across articles in the same language". This is because of the complexity, the authors note, involved in detecting similarity across articles that may be written in languages whose grammatical structures are very different (exactly what our experience confirms).

The newest Ukrainian version of the Microsoft Office 365 cloud application package working on the Microsoft Windows 11 operating system (released in 2022) (Microsoft, 2022) includes, as far as we know for the first time ever, a multilingual text authentication in the standard Microsoft Word text editor.

Now one can test our assumptions on real examples of those papers that we have translated recently (September–October 2022). The new standard function "Correction" demonstrates, in parallel with translation, how many sources in the text coincide with papers found on the Internet and provides a direct active link to the journal, article, and/or original page.

The artificial intelligence algorithm underlying this new feature operates seamlessly with texts in the Ukrainian, Russian, and English languages. This significantly facilitates anti-plagiarism checks in real time. An additional benefit is that the software proposes adding a link in the manuscript to the source detected.

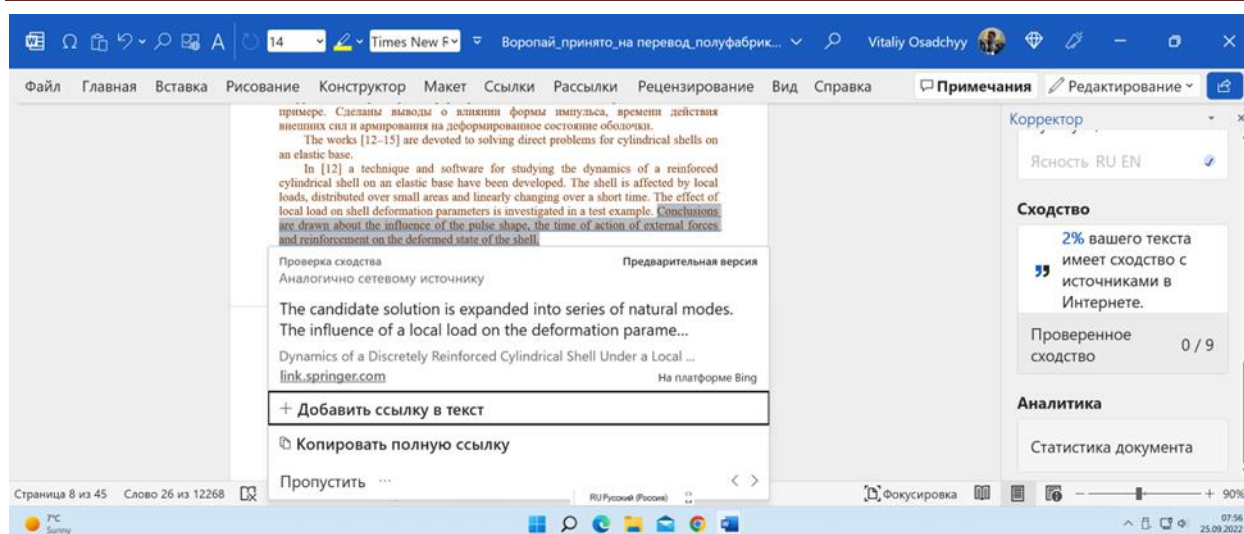


Fig. 1. Example of the Microsoft Word page showing a Correction summary (specifically, 9 sentences in the original manuscript fully coincide with Internet sources)

So, from now on an author cannot blame the lack of specialized software or costly applications for not knowing where a sentence or a paragraph was borrowed from. The caveats worth mentioning are the newest Microsoft Office package, available commercially but not common at local Universities, and the technical aspect of anti-plagiarism checks remaining to be clarified: who exactly is responsible for identifying the malpractice? In the primary triangle author–reviewer–translator, common sense suggests that the latter is most likely to reveal such cases. This assumption leaves many legal and ethical questions open for the time being.

Conclusions

1. On the one hand, the public results of anti-plagiarism checks can adversely affect the reputation of the author, on the other hand, hypothetically, Scopus specialists will now be able to check the Russian and Ukrainian versions of the paper, making appropriate conclusions about the author, reviewer, journal, educational institution, and this country as a whole.

2. In our view, the first step that must be initiated by University administrations is public discussion of a cross-language plagiarism issue, which is, unfortunately, common but rarely considered. There is no doubt that such a proposal will meet strong opposition and even rejection as it may reveal some negative aspects related to writing academic papers.

3. It is advisable to more actively involve librarians of Ukrainian universities in using the powerful potential of digital support for the research activities of their users, including writing papers and checking them for signs of plagiarism.

4. Academic integrity must be endorsed at all levels so that Ukrainian scientists cannot be reprimanded for any dishonest practices and activities.

REFERENCES

- Alaa, Z., Tiun, S., & Abdulameer, M. (2016). Cross-language plagiarism of Arabic-English documents using linear logistic regression. *Journal of Theoretical and Applied Information Technology*, 83(1), 20-33. Retrieved from <http://www.jatit.org/volumes/eightythree1.php> (in English)

- Google Developers. (2022). *Help Google choose the right canonical URL for your duplicate pages*. Google Search Central. Retrieved from https://developers.google.com/search/docs/crawling-indexing/consolidate-duplicate-urls?hl=en&visit_id=638031636483722597-103696988&rd=1 (in English)
- Hattab, E. (2015, December). Cross-language plagiarism detection method: Arabic vs. English. *Proceedings of 2015 International Conference on Developments of E-Systems Engineering (DeSE)* (pp. 141-144). Dubai, United Arab Emirates. doi: <https://doi.org/10.1109/DeSE.2015.25> (in English)
- Horster, J. (2022). *Why is it important to prevent duplicate content?* Retrieved from <https://www.seoreviewtools.com/duplicate-content-checker/#:~:text=As%20mentioned%20above,and%20search%20results> (in English)
- IEEE [web-site]. (2022). Retrieved from <https://www.ieee.org/publications/rights/plagiarism/plagiarism.html> (in English)
- Microsoft [web-site]. (2022). Retrieved from <https://www.microsoft.com/uk-ua/>
- Ouriginal. (2021). *What is cross-language plagiarism?* Retrieved from <https://www.ouriginal.com/cross-language-plagiarism-and-challenges/> (in English)
- Plagiarism in Research [web-site]. (2020). Retrieved from <https://www.editage.com/insights/does-a-plagiarism-checker-work-for-plagiarism-between-different-languages> (in English)
- Pryimak, D. M. (2019). Vtorynnist i plahiat yak peredumovy nyzkoi yakosti inshomovno-ukrainskykh slovnykiv (na materialy anhliisko-, frantsuzko- ta italiisko-ukrainskykh vydan) [Lack of originality and plagiarism as preconditions of Ukrainian bilingual dictionaries' inferior quality (a study of English-, French-, and Italian-Ukrainian dictionaries)]. *Visnyk Kyivskoho natsionalnoho lnhvystychnoho universytetu. Seriya Filolohiia*, 22(1), 123-152. doi: <https://doi.org/10.32589/2311-0821.1.2019.170210> (in Ukrainian)
- Rohrich, R. J., & Sullivan, D. (2009). Plagiarism and dual publication: review of the issues and policy statement. *Plastic and Reconstructive Surgery*, 124(4), 1333-1339. doi: <https://doi.org/10.1097/prs.0b013e3181b59d42> (in English)
- Verkhovna Rada of Ukraine. (2017). *On education: Law of Ukraine*. № 2145-VIII. Retrieved from <https://zakon.rada.gov.ua/laws/show/en/2145-19#Text> (in English)

OSADCHYI V. I.^{1*,2*}

^{1*} Наукова бібліотека, Український державний університет науки і технологій (Дніпро, Україна),

^{2*} Translation.in.ua (Дніпро, Україна),

e-mail: expert@translation.in.ua, ORCID 0000-0003-4242-6328

OSADCHA O. V.

Дніпровський національний університет імені Олеся Гончара, (Дніпро, Україна),

e-mail: elena.osadcha@gmail.com, ORCID 0000-0003-1414-6345

Алгоритми штучного інтелекту та машинного навчання для оцінки автентичності наукової статті в Scopus: досвід перекладача

Мета. У статті розглядаються шляхи вирішення проблеми міжмовного плагіату в наукових роботах, написаних українською мовою, які підлягають перекладу та публікації англійською мовою. **Методика.** Для виявлення випадків міжмовного плагіату в наукових статтях, написаних українською мовою, було використано методи опрацювання великих даних та аналіз алгоритмів програмного забезпечення машинного перекладу. **Результати.** На основі аналізу 4000 перекладених рукописів встановлено, що стандартне програмне забезпечення Microsoft Word 2022, яке зазвичай використовується для написання статті, з дуже

LIBRARY SERVICES FOR SCIENCE AND EDUCATION SUPPORT

високою точністю визначає ті частини тексту, які раніше були опубліковані та збережені в цифровому форматі. **Висновки.** З появою програмного забезпечення Microsoft Office 365 (випущеного у 2022 році) з'явилася можливість перевірити будь-яку статтю, оригінально написану українською чи російською мовою, з одночасним перекладом на англійську, на схожість із раніше опублікованими науковими статтями. Це дозволяє миттєво перевірити наукову працю, що може виявитися корисним для запобігання навмисному чи ненавмисному виникненню міжмовного плагіату в наукових статтях. Доцільно активніше залучати бібліотекарів українських університетів до використання потужного потенціалу цифрової підтримки науково-дослідницької діяльності своїх користувачів, у т.ч. написання робіт та перевірки їх на ознаки плагіату.

Ключові слова: міжмовний плагіат; наукова робота українською мовою; Scopus; програмне забезпечення Microsoft Office 365; переклад з української мови на англійську

Received: 10.07.2022

Accepted: 20.11.2022