

ЗМЕНШЕННЯ ХИБНОПОЗИТИВНИХ РЕЗУЛЬТАТІВ АНТИПЛАГІАТ СИСТЕМ

*канд. техн. наук, проф. О.А. Козіна, аспірант Д.А. Пасько
Національний технічний університет "Харківський політехнічний
інститут", м. Харків*

У науковому товаристві жорстко поставлено питання авторського права і унікальності робіт. Унікальність, є одним з важливих критеріїв при проходженні перевірок для публікації нових наукових праць. І найчастіше, автори можуть витратити більшу частину часу на боротьбу з системами антиплагіату, а не на проведення додаткових експериментів або удосконалення теоретичної бази роботи. Такі системи шукають схожі слова або словосполучення, але жодним чином не розуміють сенсу тексту, тобто вони навряд чи здатні стверджувати про факт плагіату і швидше розраховані на відловлювання недобросовісних студентів, ніж на наукових працівників.

Обґрунтовано необхідність удосконалення існуючих систем антиплагіату шляхом додавання визначення контексту роботи. За рахунок використання алгоритмів семантичного аналізу [1, 2] для виділення сенсу з тексту, можна підвищити точність систем для боротьби з плагіатом. Визначивши контекст роботи, можна сильно звужити межі, в яких буде вже виконуватися перевірка на запозичення. Завдяки цьому досягається зменшення знаходження словосполучень з інших незв'язаних з темою областей [3].

Показано основні переваги підвищення точності перевірки текстів за рахунок виділення контексту, що може бути досягнуто за рахунок зменшення хибнопозитивних спрацьовувань систем перевірки на унікальність. Для визначення контексту запропоновано використовувати один із способів при природній обробці мови (NLP) - LDA (латентне розміщення Дирихле) [4], при якому слова розбиваються на групи, на підставі яких проводиться вже перевірка на плагіат. Для підвищення точності і захисту від перефразування, пропонується такі групи піддавати синонімічній заміні, і перевіряти знову.

Список літератури: 1. *Jurafsky D.* Speech and Language Processing / *D. Jurafsky, J. H. Martin* // Pearson, 2014. Режим доступу: <https://web.stanford.edu/~jurafsky/slp3/19.pdf>. – Дата звернення: 30 жовтня 2020. 2. *Greene Spin S.C.* Lexical Semantics, Transitivity, and the Identification of Implicit Sentiment / *S.C. Spin* // ProQuest, – 2007. 3. Сервіси для перевірки якості тексту. [Електронний ресурс]. Режим доступу: <https://bit.ly/3osqKty>. 4. *Blei D.M.* Latent Dirichlet Allocation / *D.M. Blei, Y. Andrew, M. Jordan*, – 2003. – 996 с. Режим доступу: <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>.