

ПІДХІД ДО АНАЛІЗУ НА УНІКАЛЬНІСТЬ КУРСОВИХ РОЗРОБОК

Вінницький національний технічний університет

Анотація

За результатами проведеного аналізу виявлено проблеми традиційних способів аналізу курсових робіт на плагіат та запропоновано покращений підхід до аналізу курсових робіт на плагіат, що базується на використанні семантичного аналізу тексту. Такий підхід надає можливість враховувати більшу кількість випадків копіювання або переробки тексту.

Ключові слова: плагіат, алгоритм аналізу тексту, семантичний аналіз тексту, інтелектуальний аналіз даних.

Abstract

According to the results of the analysis, the problems of traditional methods of course works analysis for plagiarism identified and an updated approach to analysis for plagiarism based on the use of a semantic text analysis was proposed. Such an approach provides an opportunity to take into account a greater number of accidents of copying or changing an existing text.

Keywords: plagiarism, text analysis algorithm, semantic text analysis, intellectual analysis of data.

Стрімкий розвиток мережі Інтернет, поряд зі зростаючою комп'ютерною грамотністю, сприяє проникненню плагіату в різні сфери людської діяльності: плагіат є гострою проблемою в освіті, промисловості та науковому співтоваристві [1].

Плагіат є оприлюдненням (частково або повністю) наукових (творчих) результатів, отриманих іншими особами, як результатів власного дослідження (творчості) та/або відтворення опублікованих текстів (оприлюднених творів мистецтва) інших авторів без зазначення авторства [2].

Перевірка підозрілих документів в цій ситуації вимагає ефективного порівняння документів з різних інформаційних джерел, в тому числі що є у базі викладача, що викладає відповідну дисципліну. При цьому, слід зауважити, що всі алгоритми базуються на попарній перевірці документу, що підлягає аналізу (в даному випадку курсової розробки), з документами, що є в інформаційних джерелах. Як правило, використовують моделі, такі як суфіксне дерево або суфіксний масив, які були адаптовані для виконання цього завдання в контексті комп'ютерного виявлення плагіату. Однак співставлення підстрічки є неефективним рішенням для перевірки великих колекцій документів, адже алгоритм відпрацьовує в середньому за $2h$ -порівнянь, де h - довжина рядка, в якій ведеться пошук. Із значною збільшенні обсягу даних тривалість перевірки стає неприйнятною [3]. За означеним алгоритмом, аналіз передбачає розбиття тексту на фрагменти однорідності, які приймають різні послідовності частин мови з подальшим аналізом фрагментів, виділенням фрагментів неоднорідності, що мають різні частоти знаходження обраної послідовності частин мови. Це вказує на можливий плагіат в даному місці [2].

З метою усунення означеного недоліку при аналізі курсових робіт на унікальність пропонується удосконалити відповідний алгоритм за рахунок введення попереднього зберігання хешованих фрагментів.

Діаграму діяльності процесу аналізу курсової роботи на плагіат зображено на рисунку 1.

Таким чином, запропоновано покращений підхід до аналізу курсових робіт на плагіат, що базується на використанні аналізу фрагментів та семантичного аналізу тексту. Визначений підхід надасть можливість враховувати більшу кількість випадків копіювання та видозмінення існуючого тексту, незалежно від обсягу даних, що аналізуються.



Рисунок 1 - Діаграма діяльності процесу аналізу курсової роботи на плагіат

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Stein B., Koppel M. Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection — ACM SIGIR Forum, 2011. — 144 с.
2. Положення про академічну доброчесність студентів та науково-педагогічних працівників ВНТУ — <https://vntu.edu.ua/uploads/2019/2.PDF>, 2019. — 16 с.
3. Monostori K., Zaslavsky A., Document Overlap Detection System for Distributed Digital Libraries. ACM, 2000. — 2 с.
4. Савчук Т. О. Концептуалізація моделювання процесу аналізу проблемних ситуацій / Т.О. Савчук, О.В. Смирнова — Вісник ВПІ. — 2011. — № 1, С.96–101

Савчук Тамара Олександрівна — PhD, професор кафедри комп'ютерних наук Вінницький національний технічний університет, м. Вінниця, e-mail: savchtam@gmail.com.

Кучевський Юрій Андрійович — студент кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця, e-mail: yurii.kuchevskiy@gmail.com

Savchuk Tamara Oleksandrivna. — PhD, Professor of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia, e-mail: savchtam@gmail.com.

Yurii A. Kuchevskiy — student of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia, e-mail: yurii.kuchevskiy@gmail.com