



УДК 004.42

ДОСЛІДЖЕННЯ МЕТОДІВ АВТОМАТИЗОВАНОГО ПОШУКУ ОЗНАК ПЛАГІАТУ У НАУКОВИХ ТВОРАХ

Студ. О.В. Голікова, гр. МгЗАК (н)-16 (л)
Науковий керівник доц. В.Б. Дроменко
Київський національний університет технологій та дизайну

Мета і завдання. Аналіз існуючих автоматизованих систем пошуку ознак плагіату (запозичень) у творах наукового характеру, виявлення їх недоліків з метою формування оптимальної моделі автоматизованої системи пошуку ознак плагіату (запозичень) у наукових творах, яка б максимально задовольняла потреби користувачів.

Об'єкт та предмет дослідження. Автоматизована система пошуку ознак плагіату (запозичень) у творах наукового характеру, що проводить пошук шляхом аналізу текстового масиву даних. Предметом дослідження є методи і засоби пошуку ознак плагіату (запозичень).

Методи та засоби дослідження базуються на розробці алгоритмів лінгвістичної ідентифікації текстових даних з метою визначення приналежності твору до певного виду. Для більш глибокого аналізу тексту, орієнтованого на пошук у ньому ознак плагіату (запозичень) використовуються методи порівняння, побудовані на алгоритмах пошуку ключових ознак.

Наукова новизна та практичне значення отриманих результатів. Не зважаючи на існування низки спеціалізованих програм для пошуку ознак плагіату (запозичень) у наукових творах, дана тема потребує подальшого вдосконалення. В даному дослідженні наведено рекомендації для вдосконалення існуючих систем пошуку ознак плагіату (запозичень) у наукових творів, шляхом введення нових алгоритмів пошуку, доступу до існуючих баз даних наукових текстів та створення нових.

Практичне значення отриманих результатів полягає у наступному:

1. Розроблена автоматизована система дозволить більш детально досліджувати наукові твори на пошук ознак плагіату (запозичення) у ньому;
2. Розширення доступу до існуючих баз даних наукових творів та створення нових баз даних дозволить використовувати для порівняння більшу кількість творів.

Результати дослідження. Для реалізації поставленої мети визначено наступні напрямки дослідження:

1. Аналіз існуючих систем пошуку ознак плагіату (запозичень) у наукових творах, визначення їх особливостей, алгоритмів роботи, переваг та недоліків.
2. Розробка алгоритмів пошуку ознак плагіату (запозичень) у наукових творах. Первинний пошук (текстуальні запозичення), пошук перефразованих фрагментів тексту, перекладених фрагментів тексту, слів, у яких є заміна символів тощо).
3. Розробка екранних форм для виведення звіту про результати проведеного пошуку. Загальний звіт з даними у відсотковому еквіваленті. Детальний звіт.
4. Розробка алгоритмів доступу до існуючих баз даних наукових творів та рекомендації по створенню єдиної бази даних наукових творів в Україні.

Пошук ознак плагіату (запозичень) у наукових творах проводиться шляхом послідовного порівняння вихідного тексту (Текст 1) з іншими текстами, що містяться у базі даних/бібліотеці (Тексти 2-н). При цьому тексти порівнюються фрагментарно, тобто фрагмент Тексту 1 послідовно перевіряється з Текстами 2-н, з поточною фіксацією результатів. Результати перевірки позначаються певними маркерами. В



залежності від маркування, фрагменти текстів або зберігаються у буфері даних, або перевіряються повторно, але вже за іншими алгоритмами. За результатами перевірки формується звіт, у якому відображається наявність запозичень у кількісному чи відсотковому еквіваленті. При цьому середній відсоток виявлених запозичень виводиться за результатом аналізу за усіма етапами (повні текстуальні збіги, часткові текстуальні збіги, перекладений текст, неналежно оформлене цитування тощо), з ігноруванням елементів, що не повинні мати знак оригінальності (форма твору, що має регламентований характер, належно оформлене цитування, використання загальновідомих понять і формул там, де це необхідно тощо).

В процесі досліджень за вищезазначеними напрямками отримані наступні результати:

1. Проведено аналіз існуючих систем пошуку ознак плагіату у наукових творах, виявлено їх основні недоліки.
2. Проведено аналіз основних недоліків існуючих систем пошуку плагіату у наукових творів та визначено напрямки їх усунення.
3. Розроблено алгоритм пошуку ознак плагіату (запозичень) у наукових творів із забезпеченням необхідних етапів пошуку.
4. Проаналізовано можливі способи використання локальних та глобальних баз даних наукових текстів.
5. Розроблено алгоритми доступу до існуючих баз даних наукових творів та створено рекомендації по створенню єдиної бази даних наукових творів в Україні.

Висновки. За результатами проведеної роботи сформовано такі висновки:

1. Основною проблемою для впровадження автоматизованої системи на пошук ознак плагіату у наукових творах є доступ до обмеженої кількості баз даних/бібліотек наукових та науково-навчальних творів. Вирішення цієї проблеми полягає у створенні єдиної бази наукових робіт, доступ до якої б мала система, або надання системі доступу до баз даних/бібліотек, що знаходяться на місцях.
2. Розробка автоматизованої системи пошуку ознак плагіату у наукових творах має базуватися на декількох алгоритмах, які передбачають не тільки пошук повних текстуальних збігів, але й пошук перефразованих фрагментів тексту, перекладених фрагментів тексту, слів, у яких є заміна символів тощо.
3. Система повинна накопичувати у собі результати пошуків для їх подальшого аналізу, а також мати внутрішню базу даних найбільш популярних фраз та зв'язних конструкцій, для визначення ознак плагіату у перефразованих фрагментах тексту.
4. Результатом роботи системи стає звіт щодо виявлених ознак плагіату у відсотковому еквіваленті. При цьому враховується відсоток запозичень на кожному етапі перевірки та виводиться середнє значення.

Ключові слова: ознаки плагіату, текстуальні запозичення, база даних наукових творів, плагіат у наукових творах.

ЛІТЕРАТУРА

1. Литвиненко О. Є. Система порівняльного аналізу тексту / О. Є. Литвиненко // Інформаційні технології в системі управління освітою України : міжн. наук.-техн. конф, 29 - 30 верес. 2005 р. : тези доп. – Х., 2005
2. Петренко І. Питання виявлення плагіату літературного твору / І. Петренко // Теорія і практика інтелектуальної власності. – 2009. – № 4.
3. Закон України Про авторське право і суміжні права // Відомості Верховної Ради України (ВВР). – 1994. – № 13. – С. 64