

ІНФОРМАЦІЙНІ СИСТЕМИ ПОШУКУ ПЛАГІАТУ В СТУДЕНТСЬКИХ НАУКОВИХ ПРАЦЯХ В УМОВАХ ІНТЕРНЕТ-ПРОСТОРУ: ТЕХНОЛОГІЇ, МЕТОДИ, МОДЕЛІ

Вінницький національний технічний університет

Анотація

У цій статті розглянуто проблеми плагіату у вищій школі в умовах Інтернет-мережевого простору. Характеризуються методи та технології, що використовуються для запобігання плагіату в наукових працях студентів. Пропонується узагальнена модель автоматизованої системи перевірки текстів на плагіат та її пошукові можливості.

Ключові слова: Наукова праця, вищий навчальний заклад, студент, технологія, алгоритм, метод, автоматизована система пошуку плагіату (АСПП), плагіат.

Abstract

This article discusses the problems of plagiarism in high school on the Internet. Methods and technologies used to prevent plagiarism in students' research are described. A generalized model of an automated plagiarism check system and its search capabilities are proposed.

Key words: scientific work, higher educational institution, student, technology, algorithm, method, automated system of search of plagiarism (ASSP), plagiarism.

Вступ

Плагіат став невід'ємною частиною журналістського існування в мережі Інтернет. Копіювання текстів стало звичайною справою. Навіть школяр може викачати з мережі Інтернет стільки, скільки йому знадобиться матеріалу і говорити надалі, що він має повне право на його авторство.

Плагіат (Від лат. Plagio – викрадати) – привласнення чужого авторства, видача чужого твору або винаходу за власне.

Безумовно, боротьба з цією серйозною проблемою ведеться. На даний момент вже створено систему AntiPlagiat.ru, яка зробила перший крок у розвитку технологій, що допомагають відслідковувати незаконне копіювання журналістських матеріалів в мережі Інтернет.

З появою, багатьох Інтернет видань, мережа стала заражатися понівеченими текстами, які не мають першоджерела. Часом, незаконним копіюванням грішать і провідні Інтернет-видання, а все через те, що в нинішньому, суворому світі журналістики доводиться брати інформацію з усіх боків і не замислюватися про її походження. Мало хто став дивитися на якість, тепер, всім потрібні щоденні сенсації і плітки, якими кишить Інтернет.

Такі дії є неприпустимими і повинні бути зупинені. Звісно ж, для цього має існувати правове підґрунтя, кожен університет повинен мати кодекс честі, молоде покоління потрібно зі школи навчати академічній доброчесності, а перевірка на плагіат повинна стати невід'ємним елементом при виконанні будь-яких домашніх завдань, чи у школі, чи в університеті.

Рівень свідомості учнів, студентів, аспірантів повинен бути настільки високим, щоб вони самі, у першу чергу, були зацікавлені у перевірці їхніх робіт на плагіат з метою уникнення ненавмисного плагіату. Задля цього програмне забезпечення для перевірки робіт на наявність плагіату повинно увесь час удосконалюватись та ставати доступним для кінцевого користувача[1].

Одним з найбільш актуальних питань для науковців є перевірка академічних текстів на плагіат. Онлайн-сервіси та програми, які дозволяють провести таку перевірку, будуть корисними для науковців, студентів та інших осіб, тому доводиться працювати з академічними текстами. Адже, навіть якщо автор пише свій твір самостійно, може виникнути ситуація, коли його частина буде збігатися зі статтею, тезами конференції чи іншою вже оприлюдненою працею. У такому випадку цю частину тексту краще переписати.

Перевірити оригінальність тексту можна за допомогою веб-браузера, онлайн сервісів або спеціальних антиплагіатних програм.

Перевірка за допомогою веб-браузера є ефективною та зручною, коли потрібно визначити оригінальність невеликої частини тексту (наприклад, абзацу чи речення), з'ясувати, звідки ця частина тексту була запозичена[2].

Сервіси для перевірки текстів на унікальність працюють, переважно, за однаковим алгоритмом. Документ транслюється текстовим редактором в текст у форматі .txt і перевіряється. Перевірка та пошук співпадінь виконується по шинглах (методом шингл – розбивки тексту). Шингл – структурно-логічний фрагмент тексту, що складається з послідовності декількох слів. Пошук в Інтернеті здійснюється декількома пошуковими системами. В результаті візуалізується відсоток оригінальності тексту та список сайтів з відсотком збігу у відповідному кольорі в залежності від застосованих пошукових серверів[3].

Однак, у разі виникнення потреби перевірити оригінальність великої за обсягом праці, краще використовувати спеціальні онлайн-платформи або програми. Вони можуть бути як платними, так і безкоштовними та дозволяють перевіряти документи в різних форматах: DOC, ODT, TXT чи PDF.

Найпростішим способом перевірки оригінальності тексту є використання веб-браузера. Для цього достатньо виділити частину тексту лапками, скопіювати його і ввести в рядок введення. У цьому разі веб-браузер запропонує кілька посилань, за якими можна знайти такий текст. Якщо ж текст оригінальний, то з'явиться повідомлення про те, що не знайдено жодного документа чи результату. Для користувачів Google Chrome існує також можливість перевірити, звідки було запозичене використане в тексті фото. Для цього потрібно лише виділити відповідне зображення, натиснути праву клавішу мишки та вибрати пункт «Шукати зображення в Google»[4].

Огляд безкоштовного програмного забезпечення:

Програми Etxt Антиплагіат, як і AntiPlagiarism.net, що має безкоштовну демо-версію, дозволяють перевіряти написані тексти, всі сторінки потрібного сайту та вести пакетну перевірку всіх файлів з папки. Для того, щоб не здійснювати завантаження файлів по одному, існує пакетна перевірка. Пакетна перевірка дозволяє завантажити до програми Etxt Антиплагіат папку з файлами для одночасної перевірки. Для перевірки тексту потрібно вставити його у віконце сервісу і натиснути кнопку «Перевірити на унікальність». Інтерфейс зрозумілий і зручний, не унікальний текст розпізнається правдоподібно. Програма дуже зручна тим, що має різні типи перевірок, також великі замовники можуть отримати послугу у вигляді виділеного сервера і за допомогою інтерфейсу API перевіряти до 1 мільйона символів на добу.

До недоліків, за відгуками користувачів, відносяться:

Проблема з цифровими статистичними даними, вкрай складно перевіряти таблиці, в яких внесені дані, наприклад, по ВВП країни по роках, всі цифри будуть полічені плагіатом.

Програма Advego Plagiatus перевіряє контент, візуалізує сайти, на яких були збіги матеріалу, показує унікальність тексту у відсотках. За принципом роботи майже не відрізняється від програми «Антиплагіат». Хоча «Антиплагіат» має більше можливостей, Advego Plagiatus працює дещо швидше і стабільніше, об'єктивність пошуку залежить від адекватності налаштувань. Текст, який перевіряється у Advego Plagiatus, ділиться на фрази, а потім відбувається пошук подібних фраз в Інтернеті через пошукові сервери. Якщо фраза знайдена в Інтернеті, то вона вважається неоригінальною й підсвічується жовтим кольором. Недоліки: вимагає ручних налаштувань, обмеженість до 200000 символів.

Програма «Плагіата.НЕТ» розрахована для пошуку в Інтернеті часткових або повних копій текстових документів. Програма показує ступінь унікальності тексту у відсотках і джерела тексту. Це дає змогу одразу перевірити ресурси, звідки були взяті запозичення. Програма включає можливість перевіряти сторінки в Інтернеті. Неунікальний текст, тобто плагіат, по закінченню перевірки

виділяється іншим кольором. Після завершення перевірки програма «Плагіата.НЕТ» показує час, протягом якого здійснювалася перевірка. Програма підтримує формати RTF, DOC, DOCX – можливість роботи з документами Microsoft Word, Microsoft WordPad та інших редакторів. Недоліки: визначення плагіату відбувається на основі даних лише двох пошукових систем Яндекс і Google; повільна; вимагає «ідеального» інтернет-з'єднання; безкоштовна версія є неадекватною щодо кінцевого результату[5].

Система автоматичного виявлення плагіату в текстах Cognitive Text Analyzer. У систему вводиться досліджуваний текст. Далі Cognitive Text Analyzer проводить синтаксичний, семантичний і структурний розбір і аналіз тексту, усуває синтаксичні і семантичні неоднозначності і виявляє близько сотні характеристик, властивих даному тексту. Точна кількість характеристик залежить від обсягу тексту і різноманітності конструкцій, що зустрічаються в тексті. Швидкість розбору тексту на сучасному офісному комп'ютері складає 450-500 сторінок тексту в хвилину. В порівнянні з іншими програмами ця програма працює дуже швидко.

Недоліки: вимагає «ідеального» інтернет-з'єднання, працює адекватніше з англійськими, ніж з українсько- та російськомовними текстами. На відміну від існуючих систем перевірки на плагіат, що використовують стандартні ознаки, Cognitive Text Analyzer враховує структурні зв'язки, підлеглість одних конструкцій іншим в рамках пропозиції. Виділені зв'язки інваріантні до перестановки абзаців і пропозицій усередині тексту, контекстній заміні слів. Як показали 20 дослідження, ці характеристики демонструють високу статистичну стійкість в застосуванні до творів конкретного автора. Це дозволяє порівнювати текстові твори, отримувати кількісні оцінки їх близькості і з високою статистичною достовірністю судити про приналежність перу того або іншого автора; застосовується в англо- та російськомовній журналістиці.

Огляд комерційного програмного забезпечення:

AntiPlagiarism.NET – програмний продукт для перевірки тексту на наявність запозичень з різних джерел Інтернету. Пошук плагіату відбувається за індексом найбільших пошукових систем, включно Google, Yandex, Bing та інші. Визначає відсоток унікальності тексту. Неунікальні фрагменти у тексті виділяються різними кольорами. Пакетна перевірка. Перевірка сайту. Розширені налаштування. Підтримка форматів: doc, docx, pdf, odt, html, txt, rtf. Підтримка списку проксі. Вартість програми близько 20\$. Недоліком програми є те, що вона дуже погано функціонує з Windows XP та Windows 7.

Програма Compare Suite не лише виявляє схожість в текстах, але порівнює файли і папки. Найкраща з існуючих програм. За допомогою стандартної версії програми можливе порівняння таких типів файлів, як Adobe PDF, файлів Word, Excel та веб-сторінок. CompareSuite – не типовий інструмент для порівняння, тому що він може бути використаний для різних типів порівняння – посимвольного, послідовного і за ключовими словами. Одна з головних відмінностей в тому, що програма дозволяє використовувати різні методи порівняння. Тому завжди можна дізнатися, в якому аспекті файли схожі, а в якому – відмінні. Програма вміє порівнювати також графічні та звукові файли, що безумовно є корисним доповненням до програми. Програма порівнює файли великих розмірів і навіть в архівах.

Compare Suite випускається у вигляді стандартної і професійної версії. Професійна версія додатково підтримує елементи, необхідні для порівняння файлів і папок, розташованих на FTP, дає можливість створення коментарів до файлів і документів, роблячи зручною, наприклад, спільну роботу над написанням тексту в команді. Серйозним недоліком Compare Suite є надто висока ціна: 70\$ за стандартну і 140\$ за професійну версію. Єдиний недолік програми це її висока ціна[6].

Viper є платною програмою. Додаток спочатку сканує документи, а потім перевіряє їх на унікальність. За допомогою Viper можна перевіряти текст на багатьох сайтах та у пресі, яка опублікована в мережі Інтернет. Пошук плагіату може бути здійснений ще на локальному диску комп'ютера. Також програма може перевіряти одночасно два документи і виявити їх рівень схожості. Однаковий текст, який є в обох документах, підсвічується іншим кольором, тобто візуалізується компіляція. Єдиний недолік програми це реклама та банери під час роботи з нею[5].

При виборі програм необхідно звертати увагу на їхню здатність повноцінно підтримували українську, російську та англійську мови, зручність та доступність для використання.

Висновок

В якості узагальнення хотілося б зауважити, що кожне програмне забезпечення перевірки академічних текстів на плагіат має свої переваги та недоліки. Але в ході аналізу найкращою програмою для перевірки на плагіат є Compare Suite. Вона найкраще виконує свої функції, саме головне точно і без помилок. Я вважаю, що потрібно побільше таких програм тому, що це дуже актуальна проблема на сьогоднішній день. В майбутньому, можливо, буде набагато легше знайти плагіат, тим самим буде набагато легше перевіряти наукові роботи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Антиплагіатна первинна технічна перевірка наукових праць (методичні вказівки). – Львів, 2016. – 38 с.
2. Артамонов Є. Б. Аналіз методів протидії автоматичним системам визначення плагіату в електронних документах / Є. Б. Артамонов // Проблеми інформатизації та управління. - 2012. - №4(40). - С. 12-18.
3. Плагіат у студентських роботах: методи виявлення та запобігання : методичний посібник / Дніпропетровський національний університет ім. О. Гончара ; кол. авт.: Н. В. Стукало, К. В. Ковальчук, М. В. Литвин [та ін.]. – Дніпропетровськ, 2013. – 44 с.
4. Шишка Р. Б. Плагіат та його прояви і небезпеки / Р. Б. Шишка // Часопис Київського університету права. – 2014. – Вип. 4. – С. 170-175.
5. Технічна експертиза наукових праць на наявність академічного плагіату (методичні вказівки) / Авт.: Чоп'як В. В., Надрага О. Б., Вергун А. Р. – Львів, 2016. – 49 с.
6. Лупаренко Л. А. Інструментарій виявлення плагіату в наукових роботах: аналіз програмних рішень / Л. А. Лупаренко // Інформаційні технології і засоби навчання. – 2014. – Том 40, №2. – С. 151-169.

Галицький Артур Валентинович – студент групи КІТС-19Б, факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, Вінниця, e-mail artur.galitskiy.2019@gmail.com

Науковий керівник: **Шелепало Галина Василівна** – кандидат фізико-математичних наук, доцент кафедри захисту інформації, Вінницький національний технічний університет, Вінниця, e-mail nv.shelepalo@vntu.edu.

Galitsky Arthur Valentinovich - student of kits-19B group, Faculty of management and information security, Vinnytsia National Technical University, Vinnytsia, e-mail artur.galitskiy.2019@gmail.com

Supervisor: **Shelepalo Halyna Vasylivna** - Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Information Protection, Vinnytsia National Technical University, Vinnytsia, e-mail hv.shelepalo@vntu.edu.ua