



УДК 004.03:[347.78:343.71

ANALYSIS OF THE SYSTEMS OF TEXT PLAGIARISM DETECTION**АНАЛІЗ СИСТЕМ ВИЯВЛЕННЯ ТЕКСТОВОГО ПЛАГІАТУ****Zabolotna I.I. / Заболотна І.І.***c.med.s., as. prof. / к.мед.н., доц.*

ORCID:0000-0002-3284-0392

Makhnova A.V. / Махньова А.В.*c.med.s. / к.мед.н.*

ORCID:0000-0002-7809-4428

Komlev A.A. / Комлев А.А.*c.med.s., as. prof. / к.мед.н., доц.*

ORCID:0000-0002-5355-6331

*Donetsk National Medical University, Liman, Privokzalnaya, 27, 84404**Донецький національний медичний університет,**Лиман, Привокзальна, 27, 84404***Zabolotnyi O.S. / Заболотний О.С.***std. / студ.*

ORCID:0000-0003-0291-2727

*National Technical University «Kharkiv Polytechnic Institute»,**Kharkiv, Kyrpuchova, 2, 61002**Національний технічний університет «Харківський політехнічний інститут»,**Харків, Курпичова, 2, 61002*

Анотація. З кожним роком плагіат стає все більш поширеним явищем. Доступність інформації є підґрунтям для цього. Текстовий плагіат уявляє повне або часткове запозичення фрагментів тексту. Для його виявлення використовуються інформаційні технології і спеціалізоване програмне забезпечення, які постійно удосконалюються. Але це досить складне завдання, тому що існують різні види плагіату. У представленій роботі розглянуті найбільш популярні системи та онлайн ресурси перевірки текстів на збіг, проаналізовані їх ефективність і недоліки. При виборі програм необхідно враховувати здатність повноцінно підтримувати українську та англійську мови, їх зручність та доступність. Встановлено, що найбільш прийнятним варіантом для особистого використання є eTXT та аналогічні сервіси, а для великих обсягів перевірок доцільним є використання таких програм, як Unischek.

Ключові слова: плагіат, Інтернет, онлайн сервіс, програма

Вступ.

Доступність інформації і можливість її копіювання стали причиною безлічі проблем, пов'язаних із порушенням авторських прав [1]. Плагіат – видача чужого наукового тексту за власний або використання такого тексту у своїх роботах, не вказуючи посилання на автора. Текстовий плагіат уявляє повне або часткове запозичення фрагментів тексту (невидозмінених або модифікованих) [2]. Встановити його наявність достатньо складно, оскільки існує багато способів модифікувати початковий текст [2]. Програмні засоби допомагають ідентифікувати фрагменти тексту, які вже присутні в інших документах [1].

В Україні також має місце таке явище як плагіат, що негативно впливає на різні сфери наукової діяльності [3]. Тому проблема академічного плагіату на сьогодні дуже актуальна. Завданням даного дослідження є визначення ефективності та недоліків у використанні найпоширеніших систем виявлення текстового плагіату.



Основний текст

Перевірка та пошук співпадінь виконується методом шингл – розбивки тексту. Шингл – структурно-логічний фрагмент тексту, що складається з послідовності декількох слів. Пошук в Інтернеті здійснюється декількома пошуковими системами. У результаті візуалізується відсоток оригінальності тексту та список сайтів з відсотком збігу у відповідному кольорі в залежності від пошукових серверів [2, 3, 4]. Сервіси для перевірки текстів на унікальність працюють, переважно, за однаковим алгоритмом. Документ транслюється текстовим редактором в текст у форматі .txt і перевіряється [3].

AntiPlagiarism.NET – програмний продукт для перевірки тексту на наявність запозичень із різних джерел Інтернету. Пошук плагіату відбувається за індексом найбільших пошукових систем, включно Google, Yandex, Bing та інші. Є можливість окремого пошуку плагіату по базам даних Google Scholar та Google Books. Підсумковий звіт містить дані про обсяг введеного тексту, частку його оригінальності та перелік веб-сайтів, де знайдено співпадіня. Доступні пакетна перевірка і перевірка сайту, розширені налаштування. Підтримка форматів: doc, docx, pdf, odt, html, txt, rtf. і списку проксі. Перевірка достатньо швидка, результати відповідають дійсності. Недоліками є обмеження по об'єму тексту, що перевіряється, до 2000 символів, а також обмеження кількості перевірок на день до трьох. Для безлімітного користування необхідно придбати пакет Premium.

Програма eTXT Антиплагіат доступна як програмне забезпечення для установки на персональний комп'ютер, так і у форматі онлайн ресурсу. Можливо здійснювати перевірку текстових фрагментів, окремих файлів та пакетів файлів, а також сторінок сайтів. Передбачено дві функції перевірки тексту: на плагіат (наявність дослівних збігів) за допомогою опції «Метод виявлення копій» і на наявність рерайта (смыслових збігів) – опція «Метод виявлення рерайта». Недоліки: онлайн перевірка є безкоштовною з обмеженням до 3000 знаків без реєстрації і до 5000 знаків після реєстрації. Під час визначення того, був текст плагіатом чи ні, може бути проведений аналіз тексту на основі підрахунку частот слів і пропозицій. Більш висока частота слів і кількість посилань вказує на вищу ймовірність плагіату [2]. Крім того, програма призначена для пошуку плагіату серед робіт, доступних у мережі Інтернет, але не дає змоги аналізувати базу робіт окремих навчальних закладів [5]. Програмам eTXT Антиплагіат і AntiPlagiarism.net <http://plagiarisma.net/>, що має безкоштовну демо-версію, властивий зрозумілий і зручний Інтерфейс, неунікальний текст розпізнається правдоподібно, ресурс підтримує понад 190 мов [3].

Варіанти програми AntiPlagiarism та AntiPlagiarist за своєю структурою та функціональністю нерідко є платними комерційними модифікаціями «eTXT Антиплагіат» з акцентом на англomовний інтернет-контент. Установка повнофункціональної програми є платною, тому її використання не є легкодоступною.

Програма Advego Plagiatus призначена для пошуку в Інтернеті повних або частин копій текстових документів з інтуїтивним інтерфейсом. Вона показує ступінь унікальності тексту, джерела тексту, відсоток співпадіня тексту,



перевіряє унікальність вказаного URL. Програма перевіряє контент, візуалізує сайти, на яких були збіги матеріалу, показує унікальність тексту у відсотках. За принципом роботи майже не відрізняється від програми «Антиплагиат». Хоча «Антиплагиат» 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. Недоліками є ручне налаштування і обмеженість до 200000 символів.

Програма Плагіата.НЕТ призначена для пошуку в Інтернеті повних або частин копій текстових документів на платформах Windows 2000/XP/Vista/7. Вона 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 і підтримує формати doc, docx, rtf. – можливість роботи з документами Microsoft Word, Microsoft WordPad та інших редакторів. Недоліками є визначення плагіату на підставі даних лише двох пошукових систем Яндекс і Google; повільність; необхідність «ідеального» інтернет-з'єднання; безкоштовна версія є неадекватною щодо кінцевого результату.

У систему автоматичного виявлення плагіату в текстах Cognitive Text Analyzer вводиться досліджуваний текст. Далі Cognitive Text Analyzer проводить синтаксичний і структурний розбір і аналіз тексту, виявляє близько сотні властивих йому характеристик. Швидкість розбору тексту складає 450-500 сторінок за хвилину. Отже, програма працює дуже швидко. Недоліки: вимагає «ідеального» інтернет-з'єднання, працює адекватніше з англійськими, ніж з українськими 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. Відомо, що ці характеристики демонструють високу статистичну стійкість у застосуванні до текстів конкретного автора. Це дозволяє порівнювати текстові твори, отримувати кількісні оцінки їх близькості і з високою статистичною достовірністю судити про приналежність до того чи іншого автора.

Plagiarism Detector (<http://plagiarismdetector.net>) – безкоштовний онлайн ресурс, що автоматично сканує текстові документи, екстрагує дані та порівнює їх з мільйонами доступних джерел в мережі Інтернет. Перевірка здійснюється шляхом введення тексту у поле на сайті або завантаження файлу. Недоліки: програма розбиває текст на абзаци, для кожного з яких пропонується перевірити результат окремо, що вимагає значних витрат часу, особливо, якщо текст достатньо об'ємний. Програма Plagiarism-Detector Personal – персональний сервісний додаток, розроблений для ефективного виявлення, звітності та запобігання плагіату. Програма має простий інтерфейс, роботу можна почати відразу після установки Plagiarism-Detector на комп'ютер. Підтримує обробку документів у пакетному режимі [3].

PlagiarismChecker (<http://www.plagiarismchecker.com>) – позиціонується як



безкоштовний онлайн ресурс для викладачів, призначений для виявлення плагіату у наукових роботах шляхом перевірки коротких фраз або окремих речень. Пошук здійснюється по мережі Інтернет за допомогою пошукових машин Google and Yahoo! Додатково передбачена можливість перевірки веб-сторінок. Введення email адреси на сайті дозволяє отримувати сповіщення від Google про випадки плагіату його тексту у майбутньому. Недоліки: значні обмеження по об'єму введеного тексту [2].

Програма Compare Suite не лише виявляє схожість у текстах, але і порівнює файли і папки. За допомогою стандартної версії 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, послівного і за ключовими словами. Тому завжди можна дізнатися, в якому аспекті файли схожі, а в якому – відмінні. Програма вміє порівнювати також графічні та звукові файли, що є корисним доповненням до програми. Можливо працювати з файлами великих розмірів і в архівах. Суттєвим недоліком Compare Suite є надто висока ціна.

Viper є також 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 [2].

Найпростішим способом перевірки оригінальності тексту є використання веб-браузера, який 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.

В Україні на сьогодні найбільш популярним є онлайн-сервіс Unichек (попередня назва – unplug) <https://unichек.com/uk-ua> - інструмент для швидкої перевірки академічних робіт та будь-яких інших текстів на плагіат по Інтернет, а також по базах наукових робіт університетів чи репозитаріїв за рахунок можливості інтегруватися в електронні платформи закладів освіти (наприклад, Moodle, Google Classroom та ін.). Unichек працює з університетами за кордоном, а також використовується у багатьох українських закладах вищої освіти. Сервіс зручний та простий для роботи на кількох мовах, працює з необмеженою кількістю користувачів одночасно, тому пошук на плагіат відбувається протягом декількох секунд і отримуються найбільш точні результати в режимі реального часу. Серед вагомих переваг сервісу «Unichек» можна виділити: перевірку текстових документів на наявність запозичених частин тексту та цитат з відкритих джерел в мережі Internet або внутрішній базі закладу вищої освіти та ресурсів бібліотеки, генерування інформативного звіту про перевірку текстів у різних форматах, інтегрування з навчальними системами: LMS-Learning Management Systems, Moodle, Canvas, Sakai, Blackboard; одночасна перевірка до 100 робіт через Інтернет і до 35 робіт через внутрішню базу із одного акаунту;



перегляд документу та роботи із ним в системі без зміни його форматування; зручна робота та зберігання файлів в системі та можливість їх завантаження; знаходження текстових збігів у роботах, що написані українською, англійською, німецькою, іспанською та іншими мовами; надання списку посилань на оригінальні джерела, які можна переглянути за допомогою системи у зручному інтерфейсі; самостійне регулювання параметрів пошуку плагіату; знаходження та вилучення правильно процитованих речень та списку використаних джерел; формування бази робіт як у системі, так і з можливістю підключення сервісу до внутрішньої бази університету; автоматичне визначення заміни символів і літер у тексті, а також зворотна автоматична підстановка в текст правильних, і пошук на плагіат [6].

Польська компанія Plagiat.pl також реалізує принципи запобігання плагіату в освітньому та науковому середовищі і дає можливість українським закладам вищої освіти використовувати систему на сьогоднішній день безкоштовно. Перевагою цього сервісу є те, що існує можливість перевіряти тексти, які виконані різними мовами. При отриманні доступу до репозитарію закладу вищої освіти компанія Plagiat.pl гарантує недопущення копіювання і використання його змісту у комерційних цілях [6]. У системі Plagiat.pl простий і зручний для користування інтерфейс. Документи можна завантажувати в систему у форматах doc, odt, txt, pdf. Текст порівнюється з ресурсами Інтернету та базами даних, створеними університетом або передплаченими. У результаті перевірки система надає звіт подібності, основними елементами якого є: два коефіцієнти подібності, які показують процентну частку аналізованого документа, що є ідентичною виявленим джерелам; список джерел фрагментів, позначених як скопійовані; повний текст перевіреного документа, з чітко визначеними виявленими запозиченнями; тривожний сигнал, що інформує про викривлення в тексті в результаті використання різних алфавітів, що може вказувати на спробу приховати плагіат [3].

Таким чином, онлайн-ресурси та програми будуть корисними для науковців, студентів та інших осіб, кому доводиться працювати з академічними текстами [5]. До недоліків антиплагіатних онлайн-ресурсів можна віднести обмеженість їх функціональних можливостей у безкоштовних версіях, недостатня оптимізація для обробки україномовних робіт [5]. Також при виборі програм необхідно звертати увагу на їхню здатність повноцінно підтримували українську та англійську мови, зручність та доступність для використання.

Проведений аналіз засвідчив, що еТХТ Антиплагіат та інші аналогічні сервіси перевірки на плагіат (Anti-Plagiarism) є найбільш прийнятними для особистого використання, тоді як для використання на рівні закладів вищої освіти, де обсяги перевірок значно більші, доцільним є використання таких програм, як Unichек [2, 4, 6], Результатами дослідження [2] показали, що для використання в установах України та редакціях фахових видань доцільно застосовувати еТХТ Антиплагіат, Anti-Plagiarism за можливості їх переплати; Advego Plagiat, Плагіата.НЕТ – за відсутності дефіциту часу на виконання перевірки та Viper – для англійських текстів. Але загальна ефективність методів аналізу текстів залежить і від того, у якому саме вигляді був поданий до них



текст. Програмні засоби, які працюють на основі пошуку подібностей, дають достатньо точні результати, хоча виділено ряд факторів та обмежень підходу з пошуку подібностей. Для покращення роботи усіх методів на етапі нормалізації потрібно видаляти відомі словосполучення, враховувати синоніми, приводити літери до одного регістру та алфавіту. З цієї причини активно відбувається пошук інших підходів, одним із яких є ідентифікація плагіату на основі авторського стилю написання роботи. Він ще практично не реалізований, але має потенційні можливості використання [1].

Висновки.

Ефективне запобігання і виявлення плагіату передбачає використання програмного забезпечення для перевірки текстів на оригінальність. Сьогодні існує велика кількість алгоритмів і систем для цього. Хоча всі програми, за якими проведено порівняльний аналіз, продемонстрували наявність як переваг, так і недоліків, визначити, які з них кращі, вкрай важко. Тому вибір системи, перш за все, залежить від мети перевірки, а також переліку критеріїв та функціональних характеристик, важливих для вибору системи виявлення плагіату.

Література:

1. Главчева Ю.М., Каніщева О.В. Ідентифікація текстового плагіату в академічних документах // Системи управління, навігації та зв'язку. – 2018. – вип. 6 (52). – С. 83-86. DOI: 10.26906/SUNZ.2018.6.083.
2. Лупаренко Л.А. Інструментарій виявлення плагіату в наукових роботах: аналіз програмних рішень // Інформаційні технології і засоби навчання. - 2014. - Т. 40. - № 2. – С. 151-169.
3. Вергун А.Р., Савенкова Л.В., Чуканова С.О. Програмне забезпечення для перевірки наукових текстів на плагіат. - Київ: УБА, 2016. – 36 с.
4. Shkodkina Yu., Pacauskas D. Comparative Analysis of Plagiarism Detection Systems // Business Ethics and Leadership. - 2017. - № 1 (3). - С. 27-35. DOI: 10.21272/bel.1(3).27-35.2017.
5. Матасар Є.І. Проблема плагіату у вищих навчальних закладах та шляхи її вирішення. Збірник тез I Української конференції молодих науковців. Київ, 22–23 травня 2014 року. – С. 44-46.
6. Шабельник Т.В., Горбашевська М.О., Нетреба М.М., Тимофєєва І.Б. Аналіз інформаційно-технічного інструментарію виявлення академічного плагіату // ФМО. - 2019. - № 4 (22). – С. 130-136. DOI: 10.31110/2413-1571-2019-022-4-020.

References.

1. Glavcheva Yu., Kanishcheva O. Identification of text plagiates in academic documents. Systems of Control, Navigation and Communication. 2018;6(52):83-86. DOI: 10.26906/SUNZ.2018.6.083.
2. Luparenko L.A. Plagiarism detection tools for research works: analysis of software solutions. Information Technologies and Learning Tools. 2014;40,2:151-169.
3. Vergun A.R., Savenkova L.V., Chukanova S.O. Software for checking scientific texts for plagiarism. - Kyiv: UBA, 2016. – 36 p.
4. Shkodkina Yu., Pacauskas D. Comparative Analysis of Plagiarism Detection Systems. Business Ethics and Leadership. 2017;1(3):27-35. DOI: 10.21272/bel.1(3).27-35.2017.



5. Matasar E.I. The problem of plagiarism in higher education institutions and ways to solve it. Collection of theses of the 1st Ukrainian conference of young scientists. Kyiv, May 22-23, 2014.– P. 44-46.

6. Shabelnik T., Horbashevska M., Natreba M., Tymofieieva I. Information and technical tools for prevention and discussion of academic plagiarism in Ukraine. Physical and mathematical education. 2019;4(22):130-136. DOI: 10.31110/2413-1571-2019-022-4-020.

Abstract. *Plagiarism becomes more common every year. Availability of information is the basis for it. Textual plagiarism is full or partial borrowing of text fragments. To detect it there are used information technologies and specialized software that are constantly being improved. But it is quite a difficult task because there are different types of plagiarism. The presented work examines the most popular systems and online resources for checking texts for coincidence, their effectiveness and shortcomings are analyzed. When choosing programs, it is necessary to take into account the ability to fully support the Ukrainian and English languages, their convenience and availability. It has been established that the most acceptable option for personal use is eTXT and similar services, and it is advisable to use such programs as Unicheck for large volumes of checking.*

Key words: *plagiarism, Internet, online service, program*

Стаття відправлена: 15.01.2024 р.

© Заболотна І.І., Махньова А.В., Комлев А.А., Заболотний О.С.