УДК 004.03+001.9

Лупаренко Лілія Анатоліївна
провідний інженер
Інститут інформаційних технологій і засобів навчання НАПН України, м. Київ, Україна
lisoln@rambler.ru

ІНСТРУМЕНТАРІЙ ВИЯВЛЕННЯ ПЛАГІАТУ В НАУКОВИХ РОБОТАХ: АНАЛІЗ ПРОГРАМНИХ РІШЕНЬ

Анотація. В статті розглянуто поняття плагіату, самоплагіату та їх основні види. Приведено класифікацію інструментів автоматичного відстеження плагіату у текстових документах та основні методи, що лежать в основі їх функціонування. Розглянуто низку популярних програмних засобів та он-лайн ресурсів для перевірки текстів на збіг, проведено їх ґрунтовний аналіз, тестування та оцінювання за визначеними автором програмно-технічними характеристиками, продуктивністю функціонування та отриманими в процесі перевірки результатами. Представлено рекомендації щодо оптимального інструментарію, що може бути використаний в роботі наукових установ та редакцій фахових видань.

Ключові слова: плагіат; засоби автоматичного пошуку плагіату; програмне забезпечення; он-лайн ресурс.

1. ВСТУП

Постановка проблеми. Мірою ефективності дослідницької діяльності вченого нині прийнято вважати його *публікаційну активність*. Відкритий доступ до світового наукового досвіду в мережі Інтернет, а також переваги електронного формату, що дозволяє легко застосовувати принцип «скопіювати-вставити» (Copy & Paste), призвели до появи численного масиву робіт, що дублюють одна одну. Така ситуація характерна, переважно, хоча і не виключно, для недосвідчених авторів.

Фізичні та інтелектуальні ресурси вчених, на яких окрім дослідницької і викладацької роботи покладено додаткове адміністративне навантаження (наприклад, керівників установ, інститутів, факультетів), також небезмежні. Тиск, що здійснюється на дослідників з метою збільшення кількості їх наукової продукції та рейтингів, які прямо впливають на перспективи подальшого кар'єрного росту та можливість отримання наукового ступеня, призводить до зниження якості рукописів та частішої появи в їх публікаціях *плагіату*.

Історично склалося, що запозичення коротких висловів або невеликих частин тексту є прийнятним, поширеним, і, навіть, обов'язковим явищем: процес генерації нового знання ґрунтується на ідеях попередніх робіт, на основі яких автор може описати загальну концепцію, запропонувати новий спосіб розв'язання певної проблеми або аргументувати отримані результати. Безперечно, будь-яка наукова праця має ґрунтуватись на результатах роботи попередників, однак не копіювати їх!

В сучасній західній науковій культурі розвинувся активний *рух опору явищу плагіату*, що пропагує і підносить такі цінності, як індивідуальність, креативність та оригінальність. Однак, незважаючи на зусилля наукової спільноти, кількість етичних порушень продовжує зростати. Особливо яскраво це виражено в тих країнах світу (Китай, Індія, країни СНД), де щорічно генерується величезна кількість статей, що викликає сумніви: чи містяться в них дійсно нові наукові відкриття чи це – чергове переписування загальновідомих фактів?

© Л. А. Лупаренко, 2014.

З експоненціальним зростанням обсягів наукової продукції, плагіат перетворився у проблему і для редакцій наукових видань, оскільки достатньо складно одразу виявити його наявність у рукописі. Згідно положень Комітету з видавничої етики (Committee on Publication Ethics (COPE)) редактори відповідальні за весь опублікований контент, тому основним з їх обов'язків є перевірка рукописів, що готуються до публікації, на наявність порушень етичних стандартів проведення досліджень, оскільки це суттєво впливає не лише на репутацію видання або окремого автора, а й на розвиток галузі науки в цілому. Однак, як свідчить практика, в нашій країні така профілактична робота видавництвами не проводиться, і рукописи приймаються без перевірки.

Із розвитком інформаційних технологій створено клас програмних продуктів, призначених для автоматичного відстеження текстових запозичень в корпусах текстів, що в перспективі може мінімізувати поширення плагіату на шпальтах фахових видань України.

Аналіз останніх досліджень і публікацій. В науковій літературі дослідження в галузі автоматичного пошуку плагіату у текстових документах сфокусовані в трьох основних напрямках:

а) аналіз існуючих алгоритмів пошуку текстових збігів [12; 5] і розробка та технічний опис нових методів [14; 26; 22; 17; 28; 29; 21; 27];
б) класифікація програмних засобів пошуку плагіату [23; 24];
в) порівняльний аналіз існуючих програмних засобів [13; 16; 19; 20; 3; 6; 10] та проектування і розробка нових [1; 8; 13].

В низці вищезазначених робіт вітчизняних та закордонних вчених, а також на чисельних веб-сайтах і персональних блогах неодноразово здійснювались спроби проаналізувати наявні засоби пошуку плагіату, вибірка яких не перевищувала 5-10 варіантів, однак не обґрунтовано, за якими критеріями та з яких мотивів обрано ті чи інші ресурси. Досі не було здійснено повномасштабного дослідження такого інструментарію за єдиним набором характеристик та з чітким поділом окремо на програмне забезпечення та он-лайн ресурси.

З огляду на це **метою статті** є аналіз програмних рішень для виявлення плагіату в наукових роботах з метою визначення найоптимальнішого ресурсу, що може бути використаний у роботі наукових установ та у процесі підготовки фахових видань НАПН України.

2. ТЕОРЕТИЧНІ ОСНОВИ

2.1. Поняття плагіату та самоплагіату, їх види.

У нормативній та науковій літературі поняття ***плагіат*** визначають наступним чином:

- ✓ «оприлюднення (опублікування), повністю або частково, чужого твору під іменем особи, яка не є автором цього твору» [11];
- ✓ «привласнення авторства на чужий твір науки, літератури, мистецтва або на чуже відкриття, винахід чи раціоналізаторську пропозицію, а також використання у своїх працях чужого твору без посилання на автора» [15, 796-797]
- ✓ «вид порушення прав автора або винахідника, що полягає у незаконному використанні під своїм ім'ям чужого твору (наукового, літературного, музичного) або винаходу, раціоналізаторської пропозиції (повністю або частково) без вказівки джерела запозичення...» [4, 601];

Плагіатом також можуть бути визнані неправомірні дії з примусу до співавторства [9], а в електронному середовищі це явище може розглядатись як один з

видів електронних злочинів, накшталт хакерства, розповсюдження вірусів, спаму та фішингу.

Глобально поширеним нині в науковому середовищі є ***самоплагіат*** (англ. self-plagiarism) – *повторна* публікація автором значних за обсягом та ідентичних за формою і змістом своїх же наукових текстів без зазначення факту їх попередньої або *одночасної* публікації.

Самоплагіат може бути реалізований як у формі повторної публікації однієї і тієї ж статті із незначними змінами у назві і змісті («*подвійна*» або «*багаторазова публікація*»), так і у вигляді статті, що містить значні за обсягом фрагменти раніше опублікованих наукових робіт без зазначення посилань на останні («*нарізка*») [6]. З правової точки зору самоплагіат є актом обману, а не злочину щодо інтелектуальної власності.

Виділяють такі дії, що **характеризують процес** плагіату [18]:

- видавання чужої роботи за власну;
- копіювання слів або ідей іншої особи без посилання на її праці;
- умисне упущення посилання зі списку джерел;
- надання невірних даних про джерело (наприклад «бите» посилання);
- зміна порядку слів зі збереженням загальної структури речення та без посилання на джерело;
- копіювання великої кількості тексту або ідей із зазначенням посилань на джерела, що в сукупності складають більшу частину статті.

Плагіат класифікують за наступними **категоріями** [16]:

1. точне копіювання без змін (Copy & Paste) та без належного бібліографічного оформлення запозичених фрагментів;
2. копіювання із змінами у мовній, лексичній і технологічній інтерпретації (із перестановкою слів, заміною літер, цифр);
3. наслідування стилю;
4. переклад з іншої мови;
5. запозичення ідеї.

Вищезазначені категорії описують *умисний* плагіат. Додатково виділяють ще два **види**: а) *ненавмисний* (випадкова подібність, збіг ідей чи відкриттів, зроблених різними авторами незалежно один від одного, і т.п.) та б) *завуальований* (викладення вже відомих фактів як власної ідеї без зазначення першоджерел) [9].

З огляду на специфіку наукової діяльності та причин його появи, плагіат буває [9]:

- ✓ *професійний* – присвоєння інтелектуальних, творчих, професійних здобутків інших у професійних цілях (підвищення авторитету, заробітку, отримання нагороди, визнання, тощо);
- ✓ *освітньо-науковий* – присвоєння чужого інтелектуальної власності виключно у процесі здобуття наукового ступеню або освітньої кваліфікації.

В освіті та науці плагіат зустрічається трьох **типів**:

а) ***Текстовий плагіат*** – повне або часткове запозичення фрагментів тексту (не видозмінених або модифікованих), що присутній у статях, тезах, звітах, монографіях, рукописах кваліфікаційних робіт, тощо.

б) ***Плагіат програмних кодів*** – повне або часткове копіювання програмного коду, написаного іншою особою (наприклад, студентами) та поданого як власна розробка.

в) ***Плагіат в нетекстових джерелах*** **–** копіювання даних у електронних таблицях, діаграмах, даних наукових експериментів, фото- та медіаресурсах.

Основна відмінність між вищезазначеними типами – це **методи**, що використовуються для пошуку збігів у кожному з цих випадків. Встановити наявність першого типу плагіату дещо важче, оскільки існує безліч способів модифікувати початковий текст. Нині актуальними є дослідження в сфері *внутрішнього пошуку плагіату* [26] в документі за умови відсутності списку джерел та на основі аналізу зміни авторського стилю в окремих частинах тексту (див. рис. 1).

Рис 1. Таксономія видів неправомірного запозичення текстів в поєднанні з відповідними їм методами пошуку плагіату [26].

Щодо можливих шляхів виявлення наявності плагіату – існує два способи:

а. *ручний пошук*, що здійснюється безпосередньо викладачами, науковцями, редакторами, читачами журналів. Наявність в редакційній колегії наукових часописів добросовісних, чесних, свідомих рецензентів, що є експертами в своїй галузі та ґрунтовно аналізують рукописи, може значно мінімізувати проблему, однак не усунути її цілком.

б. *автоматичний пошук* за допомогою комп'ютерної техніки та програмних засобів.

2.2. Класифікація інструментів автоматичного відстеження плагіату.

Деякі з вищезазначених типів плагіату достатньо легко виявити за допомогою доступного в мережі Інтернет спектру пропрієтарних і вільнопоширюваних програмних рішень, що можна коротко класифікувати наступним чином:

I. Класифікація за корпусом (колекцією джерел, з якими механізм порівнює заданий документ):
   а) за типом бази даних:
      – внутрішня – наповнюється власником і/або користувачами;
      – зовнішня – електронний репозиторій, всі доступні в мережі Інтернет матеріали;
      – внутрішня + зовнішня.
   б) за типом даних, якими оперує:

- текстові матеріали;
- програмні коди;
- нетекстові матеріали (таблиці, схеми та організаційні діаграми, медіа, тощо).

II. Класифікація за характеристиками ресурсу (однак не обмежується лише поданими нижче):
   а) за типом:
      - програмне забезпечення для локального встановлення на робочу станцію користувача;
      - он-лайн ресурс.
   б) за формою власності:
      - пропрієтарне;
      - вільнопоширюване.
   в) за спектром доступних мов (оптимально якомога ширший набір):
      - для текстових матеріалів – наявність іноземних мов, наприклад, англійської, китайської, російської, тощо із врахуванням лінгвістичних особливостей кожної мови;
      - для програмних кодів - C, Java, php, Prolog, тощо.

III. Класифікація за складністю метрик, що використовуються:
   а) поверхневі показники (наприклад, кількість однакових фрагментів довжиною у п’ять слів у документах, що порівнюються);
   б) структурні показники (міра подібності документів, що порівнюються, із врахуванням їх структури).

Для спеціалістів (викладачів, науковців, редакторів фахових видань та організаторів наукових конференцій), специфіка професійної діяльності яких вимагає активного використання таких засобів, необхідний лише якісний інструментарій, що дозволить здійснювати глибинний аналіз рукописів на збіг та 100%-не відстеження запозичень.

3. МЕТОДИ ДОСЛІДЖЕННЯ

У процесі дослідження нами було здійснено аналіз вітчизняної та зарубіжної наукової літератури, а також широкого спектру джерел в мережі Інтернет, у результаті чого визначено і відібрано для подальшого розгляду низку популярних он-лайн ресурсів та програмних засобів, призначених для виявлення неправомірних «запозичень» у текстових документах. (див. табл. 1)

Таблиця 1

Інструментарій для пошуку текстового плагіату

Тип	Назва ресурсу	
Програмне забезпечення	✓ Anti-Plagiarism ✓ eTXT Антиплагиат ✓ Advego Plagiatus ✓ Double Content Finder	✓ Praide Unique Content Analyser ✓ Viper ✓ Плагиата НЕТ

Он-лайн ресурси	✓ DupliChecker ✓ PaperRater ✓ Plagiarisma.net ✓ PlagiarismChecker ✓ Plagium, ✓ PlagTracker ✓ SeeSources ✓ PlagScan ✓ Plagiarism Detector ✓ Защита уникальности контента ✓ FindCopy (Miratools) ✓ Grammarly ✓ Docoloc ✓ Text.ru ✓ Антиплагиат

Приведені ресурси були проаналізовані за наступними групами характеристик:

I. **Вихідні та організаційні дані:** ім'я розробника, веб-адреса, тип ресурсу (пропрієтарний/вільнопоширюваний), можливість безкоштовного тріал-доступу, необхідність обов'язкової реєстрації на сайті та укладання договору з розробником, кількість користувачів однієї наукової установи, що матимуть доступ до програми.

II. **Функціональні та програмно-технічні вимоги:**
   а) кількість мов, що підтримуються;
   б) формати файлів, якими оперує ресурс;
   в) база пошуку (мережа Інтернет, інституційний репозиторій, локальна системна БД) та інтеграція з пошуковими машинами;
   г) типи елементів, що перевіряються (фрагменти введеного вручну тексту / файли / веб-сторінки);
   д) обмеження щодо об'єму введеного тексту;
   е) обмеження щодо кількості перевірок;
   ж) алгоритми, що використовуються системою для порівняння документів;

III. **Продуктивність роботи:**
   а) час виконання перевірки (від завантаження до звіту про результат);
   б) форма представлення кінцевого результату (статистичний звіт);
   в) дані про результати пошуку;
   г) особливості використання / інсталювання;
   д) особливості інтерфейсу (інтуїтивна зрозумілість, наявність інструктивних та довідкових матеріалів).

В процесі порівняння оцінювалось ступінь того, наскільки функціональні можливості даних ресурсів задовольняють специфіці наукової діяльності в нашій країні, а їх продуктивність – з огляду на отримані результати перевірки.

За текстовий зразок для тестування було обрано наукову статтю Бикова В. Ю. [2] (26613 символів = 0,67 др.арк), опубліковану в №1 (2006) електронного фахового видання «Інформаційні технології і засоби навчання», що за даними Google Scholar за весь період виходу журналу була процитована найбільше разів – 50. Цей факт потенційно має забезпечити велику кількість збігів в результатах пошуку, до того ж встановлено, що дана робота неодноразово була скопійована недобросовісними сайтами і поширюється як студентський реферат. За англомовний зразок було обрано статтю авторів Dave E. Marcial, Mitzi S. Fortich, Jeambe B. Rendal [25] (36007 символів = 0,9 др.арк), що опублікована в останньому номері того ж видання (№1 (39) 2014 р.) та дублюється лише на персональному сайті автора. Перевірка здійснювалась наступними способами: шляхом введення довільної частини тексту статті в спеціальне поле на сайті, завантаження файлу статті та за URL-адресою статті в журналі.

4. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

4.1. Аналіз програмних засобів для виявлення текстового плагіату

В результаті тестування 7 програмних засобів пошуку текстового плагіату, було одержано наступні результати:

а) ***Double Content Finder (DCFinder), Praide Unique Content Analyser II –*** не рекомендується для використання;

б) ***Advego Plagiatus***, ***Viper*** та ***Плагиата.НЕТ*** – рекомендується для персонального використання окремими науковцями. З огляду на достатньо тривалий час, необхідний для перевірки одного документа, дані програмні засоби не зручно використовувати в процесі підготовки наукових видань, де необхідно виконувати перевірку масиву рукописів;

в) ***Anti-Plagiarism*** – придатна для використання за умови її передплати;

а) ***eTXT Антиплагиат*** – рекомендовано для застосування в наукових установах НАПН України та редакціях фахових видань.

Докладні дані порівняльного аналізу подано в Таблицях 1-3 додатку до статті (Comparative_Tables.pdf), що доступний за веб-адресою: http://journal.iitta.gov.ua/index.php/itlt/article/view/1050/771

Коротко приведемо опис та отримані результати тестування кожного ПЗ.

eTXT Антиплагиат (http://www.etxt.ru/antiplagiat/) - сервіс перевірки текстів на унікальність, що доступний як програмне забезпечення для установки на персональний комп'ютер, так і у форматі он-лайн ресурсу. Можливо здійснювати перевірку текстових фрагментів, окремих файлів та пакетів файлів, а також сторінок сайтів. Передбачено 2 функції перевірки тексту: *на плагіат* (наявність дослівних збігів) за допомогою опції «Метод виявлення копій», та на наявність *рерайта* (смислових збігів) – опція «Метод виявлення рерайта».

Тестування: підсумковий звіт містить підкреслені частини тексту заданого документу, які були запозичені з інших джерел, а також список доменів, на яких знайдені збіги, із визначеною їх часткою (%) по кожному домену та загальним показником унікальності всього тексту. Унікальність українського тексту – 6%, англійського – 73%. Підсумкові звіти нашого тестування (див. файл Test_reports.rar) доступні для перегляду за адресою: http://journal.iitta.gov.ua/index.php/itlt/article/view/1050

Недоліки: он-лайн перевірка є платною з обмеженням 10000 знаків (1.5 руб./1000 симв. з пробілами).

Висновки: рекомендовано для використання.

Advego Plagiatus (http://advego.ru/plagiatus/) - програма пошуку в мережі Інтернет часткових або повних копій текстового документа, що показує ступінь унікальності тексту, список джерел тексту, відсоток збігу текстів. Також доступна перевірка унікальності за URL-адресою. Найновіша версія Advego Plagiatus 1.3.0.92., що тестувалась, знаходиться в стадії публічного бета-тестування. Пошук плагіату доступний в режимі «Швидкої перевірки» та «Глибинної перевірки».

Тестування: в режимі швидкої перевірки встановлено (копірайт/рерайт): 4%/2% унікальності для україномовної статті (за 33 хв.) та 77%/72% – для англомовної статті (58 хв.);

Недоліки: сканування, навіть у режимі швидкої перевірки, триває досить довго.

Висновки: підходить для персонального використання. Не може бути рекомендована для редакцій журналів, що потребують перевірки великої кількості текстів.

Double Content Finder (DCFinder) (http://textbroker.ru/main/dcfinder.html) – безкоштовна програма для перевірки тексту на унікальність по мережі Інтернет за допомогою пошукової машини Яндекс. Не потребує встановлення на комп'ютер та запускається при подвійному клацанні. Робота з програмою можлива будь-яким з трьох способів: додаванням тексту з буферу обміну, завантаженням текстового файлу або зазначенням веб-адреси, за якою розташована потрібна стаття. У підсумковому звіті виводиться список адрес, за якими розташовані копії статті (кількість адрес обмежено 50–ма). На даний момент підтримка програми припинена.

Тестування: в обох випадках 100% унікальності тексту, що не є достовірним результатом.

Недоліки: підтримка лише російської мови.

Висновки: не рекомендується для застосування.

Praide Unique Content Analyser II (http://best-soft.npps.biz/files.php?f=259) - програма для перевірки тексту на унікальність в мережі Інтернет щляхом розбору вихідних даних на пасажі або шингли. Програма має наступні переваги: три варіанти введення вихідних даних (по url, з файлу або фрагмент тексту з буфера обміну), розбір вихідних даних на пасажі або шингли, а також можливість вибору пошукових машин користувачем. Чим більше заданих користувачем параметрів, розмірів тексту і менше розмір пасажу або шинглу – тим довший час пошуку.

Тестування: підсумковий звіт не містить статистичних даних, а лише список сайтів, де знайдено подібні пасажі.

Недоліки: занадто довгий час на виконання перевірки.

Висновки: не може бути рекомендований.

Viper (http://www.scanmyessay.com) – простий, точний та безкоштовний інструмент виявлення плагіату, що допомагає здійснити пошук по широкомасштабній базі документів (10 мільярдів одиниць), призначений для сканування лише англомовних текстів.

Тестування: вдалось протестувати лише одну статтю, перевірка тривала 17 хвилин. Підсумкові звіти нашого тестування (див. файл Test_reports.rar) доступні для перегляду за адресою: http://journal.iitta.gov.ua/index.php/itlt/article/view/1050

Недоліки: обмеження кількості запитів/на день; англомовний інтерфейс, не зрозумілий на перший погляд.

Висновки: рекомендовано для персонального користування та пошуку плагіату в англомовних текстах.

Плагиата.НЕТ (http://www.mywebs.ru/plagiatanet.html) - безкоштовна програма для пошуку часткових або повних копій текстових документів в мережі Інтернет. Програма підтримує роботу з файлами у форматі .rtf і .txt. Після вибору і завантаження файлу програма в автоматичному режимі розбиває текст на фрази. Якщо в тексті зустрічаються слова з вживанням букв декількох мов (наприклад, російської та англійської), з'явиться діалогове вікно, в якому необхідно вказати подальші дії. Рекомендується вибирати «ТАК», тобто відбудеться заміна англійських букв на аналогічні російські.

В підсумковому звіті відображаються всі інтернет-посилання на файли, з якими порівнювався заданий документ, та ступінь унікальності тексту у відсотках. Окремо виділені джерела, збіги з якими перевищує 3%. Результат роботи програми можна зберегти у форматі .rtf в меню «Файл».

Тестування: для україномовної статті виявлено 100% збігів з іншими джерелами; для англомовної - 31%.

Недоліки: занадто довгий час на виконання перевірки (29 хв. – укр.; 60 хв. – англ.).

Висновки: може бути рекомендований для персонального використання.

4.2. Аналіз он-лайн ресурсів перевірки текстів на унікальність

В результаті тестування 15 он-лайн ресурсів перевірки текстів на унікальність, було отримано наступні результати:

г) сервіси ***PlagiarismChecker, Plagium, PlagTracker*** протестувати не вдалось через помилку в роботі ресурсу;

д) ***Plagiarisma.Net*** та ***Антиплагиат*** не можуть бути рекомендовані з огляду на обмеження у використанні;

е) ***Duplichecker, PaperRater, SeeSources, Plagiarism Detector*** можна застосовувати для швидкого поверхневого аналізу текстових документів;

ж) ***PlagScan, Docol©c, Grammarly*** – придатні для використання за умови їх передплати;

б) ***FindCopy(MiraTools), Защита уникальности контента*** та ***Text.ru*** – рекомендовано для застосування в наукових установах НАПН України.

Докладні дані порівняльного аналізу подано в Таблицях 1-3 додатку до статті (Comparative_Tables.pdf), що доступний за веб-адресою: http://journal.iitta.gov.ua/index.php/itlt/article/view/1050/771

Коротко приведемо опис та отримані результати тестування кожного ресурсу.

Duplichecker (http://www.duplichecker.com/) – безкоштовний он-лайн інструмент для перевірки на плагіат у вільному доступі. На головній сторінці даного ресурсу міститься поле для вводу текстового фрагменту з обмеженням до 1500 слів, а також поле для завантаження текстових файлів (максимум до 50 кБ). Підтримуються формати *.docx та *.txt. Кількість перевірок для незареєстрованих користувачів обмежена – 1 пошук/день. Реєстрація на сайті дозволяє виконувати до 50 перевірок на день. Після реєстрації з'являється додаткове поле для вводу веб-посилання на необхідний файл, якщо він розміщений в мережі.

В профілі користувача з'являється меню з додатковими можливостями, а саме:

- порівняльний пошук плагіату в двох введених користувачем текстових фрагментах або за URL посиланнями на дві веб-сторінки;
- можливість отримувати щомісячного або щотижневого звіту про плагіат на вказаний сайт.

Тестування шляхом завантаження файлів здійснити не вдалось – на сайті з'являється помилка, що розмір файлу занадто великий, однак при завантаженні менш об'ємних файлів, результату також не було отримано. Перевірка подібності тексту веб-сторінки шляхом введення її URL-адреси працює, результати з'являються миттєво – не довше ніж за хвилину; отримано до 20 збігів текстових фрагментів. Швидкий пошук за невеликими частинами скопійованого і вставленого в поле тексту показав до 40 збігів для україномовної статті та 15 – для англомовної; отримані джерела вірні.

Недоліки: ресурс не працює з текстовими файлами, допустимий обсяг тексту для введення (1500 слів) не задовольняє специфіці наукових робіт об'ємом до 40000 символів.

Висновок: ресурс можна використовувати частково, для окремих абзаців або для перевірки вже опублікованих html-статей, розміщених на сайтах.

PaperRater (http://www.paperrater.com/plagiarism_checker) – безкоштовний он-лайн сервіс, що містить автоматизований літ. редактор для перевірки граматики, правопису та стилістики рукописів, а також інструмент виявлення плагіату, шляхом використання потужностей основних пошукових машин (Yahoo!, Bing, Google). Даний ресурс простий у використанні та дозволяє здійснити комплексну перевірку текстів під час підготовки до друку. Інтерфейс інтуїтивно зрозумілий. Для пошуку плагіату необхідно скопіювати та ввести текст рукопису у відповідне поле на сайті. Не має жодних обмежень щодо розмірів тексту. Результатом перевірки є повідомлення по наявність плагіату, частка оригінального тексту рукопису та посилання на веб-сторінки, що можуть містити подібний текст.

Тестування: для обох статей, що перевірялись, знайдено збіги з іншими документами в мережі Інтернет; у випадку україномовної статті таких збігів значно більше, що й треба було довести.

Недоліки: підсумковий звіт досить стислий та не містить статистичних даних, не відображено запозичені фрагменти тексту, приведено лише 3 зовнішні джерела.

Висновок: ресурс може бути використаний для швидкого аналізу рукопису, однак звіт не містить достатньо детальний даних для підтвердження наявності плагіату.

Plagiarisma.Net (http://plagiarisma.net) – інструмент пошуку плагіату, що доступний як он-лайн так і у вигляді програмного забезпечення для інсталювання у вигляді клієнта на робочу станцію користувача, що підтримує ОС Windows 7,8, Vista та XP. Перевірку можливо здійснювати за фрагментами тексту, що вводяться в спеціальне поле, за URL веб-сторінки та шляхом завантаження файлів формату TXT, HTML, RTF, DOC, DOCX, ODT, PDF. Підтримується більше 190 мов. Є можливість окремого пошуку плагіату по базам даних Google Scholar та Google Books. Для безлімітного користування необхідно придбати пакет Premium. Підсумковий звіт містить дані про обсяг введеного тексту, частку його оригінальності та перелік веб-сайтів, де знайдено спів падання.

Тестування: перевірка достатньо швидка, результати відповідають дійсності, весь задекларований функціонал працює.

Недоліки: обмеження по об'єму тексту, що перевіряється, до 2000 символів, а також обмеження кількості перевірок на день, а саме – 3 перевірки/день.

Висновок: з огляду на обмеження не може бути рекомендований для використання.

PlagiarismChecker (http://www.plagiarismchecker.com) – позиціонується як безкоштовний он-лайн ресурс для викладачів, призначений для виявлення плагіату в наукових роботах студентів 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 надає користувачеві ряд преференцій, в тому числі, дозволяє отримувати сповіщення від Google про випадки плагіату його тексту у майбутньому.

Тестування: перевірку як текстового фрагмента, так і веб-сторінки виконати не вдалось; ресурс повертає помилку щодо максимально допустимої довжини запиту

пошукової машини Google (до 32 слів), хоча введений нами текст даний ліміт не перевищував.

Недоліки: значні обмеження по об'єму введеного тексту, що не задовольняє об'ємам наукової літератури.

Висновок: не підходить для використання з огляду на обмеження та помилки в роботі ресурсу.

Plagium (http://www.plagium.com) – он-лайн ресурс, що підтримує всі типи перевірки на наявність запозичень (по текстовим фрагментам, файлам та URL). Перевірка здійснюється для рукописів на 6 європейських мовах, при чому використовується авторська технологія, за якою вхідний текст розбивається на дрібніші "фрагменти", що порівнюються з веб-контентом. Передбачена можливість створення акаунтів наукових установ та інститутів, що скеровуватимуться відповідальними адміністраторами.

Тестування: хоча розробник декларує можливість безкоштовної перевірки текстів об'ємом до 25000 символів, жодних операцій на сайті зробити не вдалось, в тому числі і зареєструватись. З'являється повідомлення, що локація користувача не дозволяє проводити операції на сайті.

Недоліки: платний доступ та помилки в роботі ресурсу.

Висновок: не може бути рекомендований для використання.

PlagTracker (http://www.plagtracker.com) – безкоштовний он-лайн ресурс пошуку текстового плагіату та інструмент автоматичного коригування, що може слугувати студентам, викладачам та власникам веб-сайтів у процесі підготовки текстових матеріалів, а також редакторам фахових видань для перевірки поданих в редакцію рукописів на оригінальність. В основі даного проекту є авторський алгоритм для пошуку плагіату в контенті, розроблений двома **українськими студентами** в рамках їхньої наукової роботи під час навчання в університеті. Перевірка здійснюється по мережі Інтернет та власній внутрішній базі документів (20 млн. наукових статей). Текстові фрагменти об'ємом до 5000 слів перевіряються безкоштовно. Завантаження файлів формату .doc та .txt, а також текстів більше 5000 слів доступно за умови реєстрації та придбання Premium акаунта (7,5 $ /місяць).

У підсумковому звіті текстові збіги виділені кольором, із вказівками, де необхідно проставити цитати, та посиланням на оригінальні джерела. Додатково приводиться частка (%) запозиченого тексту.

Тестування: здійснити не вдалось – ресурс повертає помилку 504 Gateway Time-out.

Недоліки: платний доступ для здійснення перевірки великих за об'ємом документів та завантаження файлів, а також помилки роботи ресурсу в тріал-режимі.

Висновок: рекомендовано за умови можливості придбання Premium акаунта та налагодження роботи сервісу.

SeeSources (http://www.plagscan.com/seesources/analyse.php) – безкоштовний, автоматичний і ефективний он-лайн інструмент контент-аналізу на наявність скопійованого тексту. Перевірка відбувається в два етапи:

- *первинний пошук* повертає список Інтернет джерел, що відповідають заданому документу разом із оцінкою ступеню їх відповідності;
- в процесі *глибинного пошуку* результати додатково відфільтровуються; підраховуються подібні записи та відображаються подібні фрази.

Тестування: здійснено одну перевірку, що доступна безкоштовно. Знайдено близько 20 джерел, однак їх релевантність досить низька.

Недоліки: обмеження за розміром завантажуваних файлів до 300kB та текстових фрагментів до 1000 слів.

Висновок: ресурс може бути використаний для швидкого аналізу рукопису. Для більш професійного пошуку доцільно використати професійну версію даного ресурсу – PlagScan.

PlagScan (http://www.plagscan.com/) – автоматичний он-лайн сервіс пошуку плагіату, що вимірює якість тексту шляхом тестування документа на оригінальність.

Є можливість одночасної перевірки декількох документів та завантаження їх одним .zip архівом та перевірка за URL адресами файлів.

Звіт відображається в браузері та доступний для завантаження у форматі .pdf. Відображається перелік запозичених фрагментів, перераховані їх джерела та кількість збігів на джерело. Визначається загальна частка плагіату в документі (%).

Тестування:

- для україномовної статті знайдено 192 збігів із 17 джерел, з яких всі 17 розміщенні в Інтернеті (98,5% плагіату);
- для англомовної статті: 6 збігів з 2 джерел, з яких 2 розміщенні в Інтернеті (1,4% плагіату).

Підсумкові звіти нашого тестування (див. файл Test_reports.rar) доступні для перегляду за адресою: http://journal.iitta.gov.ua/index.php/itlt/article/view/1050

Недоліки: не виявлено.

Висновок: ресурс рекомендується для використання в наукових установах за умови його передплати.

Plagiarism Detector (http://plagiarismdetector.net) – безкоштовний он-лайн 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.

Звіт, що з'являється у результаті процедури аналізу, містить підкреслені частини тексту заданого документу, які були запозичені з інших джерел.

Тестування і 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. Окрім цього, користувач обов'язково повинен проаналізувати результати, оскільки виникають запитання щодо їх релевантності.

Висновок: ресурс може бути використаний для швидкого аналізу рукопису або його окремих частин.

Защита уникальности контента (http://www.content-watch.ru/text) – безкоштовний он-лайн сервіс, що дозволяє здійснювати перевірку унікальності текстів та контенту сайтів.

Пошук реалізується шляхом введення необхідного фрагменту тексту в спеціальне поле або URL адреси сайту, де опубліковано певний контент. Додатково передбачено можливість заповнити поле "Ігнорувати", де можна вписати домен сайту, на якому текст розміщено, що дозволяє виключити із результатів пошуку джерело тексту. Нещодавно розробниками поліпшено та впроваджено модуль роботи з українськими текстами, для яких рівень якості перевірки відтепер відповідає рівню перевірки російськомовних текстів.

Тестування: Перевірялись фрагменти тексту об'ємом 10 тис. символів. Для україномовної статті знайдено 5 джерел, що містять 90% збігів в мережі Інтернет, унікальність тексту 10%. Для англомовної статті знайдено 2 джерела збігів, одне з яких - персональний сайт автора, та 71% унікальності.

Недоліки: максимально допустимий обсяг тексту, що можуть перевірити зареєстровані користувачі, рівний 10000 символам, що замало для обсягу наукової статті; також обмеження щодо кількості перевірок – 20/день.

Висновок: ресурс рекомендується для використання в наукових установах, авторами наукових статей та редакціями видань.

FindCopy (http://www.miratools.ru/) – проект холдингу Mirafox (Росія), що являє собою безкоштовний он-лайн ресурс для перевірки унікальності текстів. Перевірка здійснюється шляхом введення тексту або URL-адреси веб-сторінки у спеціальне поле на сайті, що містить текст для перевірки. Для зняття обмеження щодо об'єму тексту та кількості перевірок на день необхідна реєстрація на сторонньому сайті Copylancer.ru в ролі замовника або виконавця (копірайтера). Інтерфейс інтуїтивно зрозумілий, жодних інструкцій на сайті нема. Перевірка збігів здійснюється по всій мережі Інтернет. Додатково зберігається історія всіх пошуків користувача.

Тестування: В процесі тестування знайдено 45 джерел збігів для україномовної статті, в основному сайти, на яких запозичено наш оригінальний текст (що й треба було довести), а також 3 джерела – для англомовної статті (2 з яких – сайт видання, в якому стаття опублікована, та 1 – персональний сайт автора). Процес пошуку зайняв загалом 3-4 хвилини в обох випадках.

Недоліки: не виявлено.

Висновок: Ресурс може бути використаний в роботі наукових установ та редакцій.

Docol©c (https://www.docoloc.de) – інструмент пошуку плагіату в текстових фрагментах та файлах шляхом перевірки по комп'ютерним програмам та базам даних, в результаті чого користувачеві доступні дані щодо наявності запозиченого тексту, цитат та джерел документів в мережі Інтернет. Для виконання професійної перевірки необхідна реєстрація на сайті та передплата ліцензійного доступу, розмір якої залежить від запланованої кількості сторінок, що перевірятимуться за рік. Ключ доступу може використовуватись лише даним користувачем, однак університети, наукові установи, видавництва та інші юридичні особи можуть оформити доступ на 25 осіб. Безкоштовно можна працювати в демо-режимі, попередньо зареєструвавшись на сайті, однак точність такого пошуку обмежена – показані лише деякі випадкові частини запозиченого тексту (як приклад використання даного ресурсу).

Тестувались файли у форматах .PDF, .doc, .docx, odt, RTF, HTML та.txt у демо-режимі. Результатом перевірки є звіт у форматі html, що містить інформацію про кількість перевірених речень та кількість запозичених речень, що виділені різними кольорами відповідно до ступеню їх подібності із знайденими джерелами, на які також подаються посилання. Звіт можна завантажити на ПК, зберегти в акаунті (до 60 днів у демо-режимі) або надіслати на потрібну електронну адресу.

Недоліки: не вдалось протестувати в професійному режимі з огляду на платний доступ.

Висновок: ресурс придатний для використання за умови передплати його науковою установою.

Grammarly (http://www.grammarly.com) являє собою автоматизований он-лайн коректор, що дозволяє здійснити миттєвий пошук і правку більше 250 типів

граматичних помилок, навіть тих, що залишилися невідстеженими іншими популярними текстовими процесорами, а також провести вдосконалення стилістики відповідно до типу документу (наукові статті, звіти, ессе, листи, книжки, дисертації або бізнес-документи). Додатково доступний для завантаження безкоштовний інструмент Grammarly Lite (http://www.grammarly.com/download/chrome) для веб-браузерів Chrome, Firefox і Safari, що дозволяє здійснювати поточну перевірку орфографії під час роботи в мережі Інтернет (наприклад у Facebook, Gmail, Twitter, LinkedIn, блогах, тощо). Grammarly також є інструментом відстеження неоригінальних текстів, що не лише виявляє запозичені тексти, а й робить підказки на їх джерела та дозволяє оформити їх в цитати.

Тестування: результатом перевірки є статистичний звіт щодо кількості граматичних, орфографічних, стилістичних помилок та помилок пунктуації, містить попередження про наявність неоригінального тексту, а також загальну шкала якості рукопису (максимально 100 балів).

Недоліки: перегляд результатів перевірки можливий лише за умови оплати; ресурс придатний для роботи лише з англомовними текстами, не підтримується формат .pdf, є обмеження по об'єму тексту.

Висновок: ресурс придатний для підготовки англомовних рукописів за умови реєстрації та придбання права доступу до повного функціоналу.

Text.ru (http://text.ru/) – безкоштовний онлайн-сервіс перевірки тексту на унікальність, що дозволяє розпізнавати плагіат, навіть якщо в тексті було здійснено перестановку слів і фраз, змінено відмінки, часи та інші граматичні категорії слова та додані нові слова. На сайті розміщено поле для вводу тексту, що перевіряється, допустимий обсяг якого від 100 до 15 000 символів. Є можливість запустити перевірку унікальності будь-якої кількості текстів. Забезпечується можливість після запуску перевірки закрити вікно браузера, а потім у зручний для час зайти в архів текстів (за посиланням "Архів моїх текстів") і подивитися результати перевірки. Після реєстрації перевірка відбуватиметься швидше і без обмежень по кількості перевірених текстів.

Тестування: краще введений текст ставиться в чергу, тому користувачеві необхідно зачекати 1-2 хвилини поки перевіряються тексти, завантажені іншими користувачами. Перевірка відбувається швидко. Підсумковий звіт містить список доменів, на яких знайдені збіги, із визначеною їх часткою (%) по кожному домену та загальним показником унікальності всього тексту. Є функції перегляду подібних фрагментів з підсвічуванням. Для україномовної статті унікальність рівна 0.30%, для англомовної – 23.74% (із посиланнями на журнал, в якому стаття опублікована).

Недоліки: необхідність очікування в черзі та ліміт 15000 символів.

Висновок: даний сервіс рекомендується до використання.

Антиплагиат (http://www.antiplagiat.ru/QuickCheck.aspx) - інтернет-рішення для виявлення запозичень в текстових документах, призначене для використання студентами та викладачами, вузами і науковими організаціями.

Пошук запозичень здійснюється по російському сегменту мережі Інтернет, а також реферати, наукові публікації, статті російськомовної «Вікіпедії» та колекції дипломних робіт Академії ВЕГУ. Підключення додаткових колекцій джерел (колекція Російської державної бібліотеки та нормативно-правової документації) можливе за умови передплати аккаунту «Стандарт». Алгоритми системи побудовані таким чином, що на перевірку не впливає ряд типових маніпуляцій, таких як заміна кириличних букв схожими латинськими, перестановка абзаців, перегрупування речень в абзацах, перестановка слів у реченнях, розбиття і злиття речень, заміна

пробілів точками та заміна слів синонімами. Можливо здійснювати перевірку завантаженням файлів у форматах HTML, Plain Text (txt), DOC (MS Word), RTF, PDF, ZIP и RAR та текстових фрагментів до 5000 символів.

Існує локальна версія системи «Антиплагіат», що розміщується на обладнанні ВНЗ та функціонує незалежно від наявності доступу до Інтернету Зареєстровані користувачі, які мають безкоштовний доступ, після перевірки отримують звіт, в якому вказано відсоток оригінальності і список джерел запозичення. Користувачі, які мають платний аккаунт, отримують можливість перегляду повного звіту, де кольором виділені блоки тексту, що запозичені, із посиланням на джерело.

Тестування: система не розпізнає українську мову, в підсумковому звіті знайдено лише 3 джерела збігів на сайтах рефератів, підсумкова оцінка оригінальності: 96,83% . Для англомовної статті знайдено 5 джерел збігів, однак не тих, на яких стаття власне опублікована, підсумкова оцінка оригінальності: 96,2%.

Недоліки: ліміт допустимого об'єму для текстових фрагментів (5 тис. символів), обмежена база даних, проблеми з підтримкою української мови та необхідність передплати.

Висновок: не рекомендується.

5. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

За результатами дослідження можемо зробити висновок, що для використання в установах НАПН України та редакціях фахових видань доцільно застосовувати такі програмні засоби:

- ✓ ***eTXT Антиплагиат, FindCopy(MiraTools), Защита уникальности контента*** **та** ***Text.ru****;*
- ✓ ***Anti-Plagiarism, PlagScan, Docol©c, Grammarly*** – за можливості їх передплати;
- ✓ ***Advego Plagiatus***, ***Плагиата.НЕТ*** – за відсутності дефіциту часу на виконання перевірки та ***Viper –*** для англомовних текстів.

Потрібно зважати на те, що такий інструментарій лише виявляє подібність текстів та знаходить джерела запозичення, однак не визначає, чи є той чи інший фрагмент тексту першоджерелом або правильно оформленою цитатою. Іноді цитата оцінюється як дублювання, а отже у кінцевих підсумкових звітах користувачеві повідомляється лише про фрагменти, що *потенційно* можуть бути неавторськими. Остаточне рішення про наявність факту плагіату залишається на розсуд користувача, тому потреба в ручній перевірці та людському оцінюванні (рецензуванні) матеріалу все одно існує.

Оскільки кожен такий ресурс має свої переваги і недоліки згідно програмно-технічних характеристик, не можна очікувати від машинної обробки стовідсотково точних результатів. Основна ж мета використання даного класу програмних засобів – максимально загальмувати тенденцію до стрімкого поширення плагіату та вивести вітчизняну науку на новий якісний рівень.

Перспективи подальших досліджень вбачаємо у розробленні організаційних умов, технічного забезпечення, технології та специфікації впровадження цих засобів як сервісу в корпоративному середовищі НАПН України.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Автоматизована система перевірки текстів на плагіат / В. Б. Мокін, В. В. Войтко, С. В. Бевз [та ін.] // Вісник Вінницького політехнічного інституту. – 2010. – № 5.— С. 12—17.

2. Биков В. Ю. Теоретико-методологічні засади моделювання навчального середовища сучасних педагогічних систем [Електронний ресурс] / В. Ю. Биков // Інформаційні технології і засоби навчання. – 2006. – № 1 (1). – Режим доступу : http://journal.iitta.gov.ua/index.php/itlt/article/view/285/271
3. Болілий В.О. Перевірка унікальності тексту при оцінюванні студентських робіт творчого або дослідницького характеру/ В.О. Болілий, В. В. Копотій // Наукові записки НДУ ім. М. Гоголя. – 2011. – № 7 (34).— С. 134—145.
4. Большая Советская энциклопедия [Текст] : в 30 томах / гл. ред. А.М. Прохоров. - изд. 3-е. - М. : «Советская энциклопедия», 1975. - Т. 19. - 647 с.
5. Квашина Ю. А. Методы поиска дубликатов скомпонованных текстов научной стилистики [Электронный ресурс] / Ю. А. Квашина // Технологический аудит и резервы производства. – 2013. – № 3 (3).— С. 16-20 – Режим доступа : http://journals.uran.ua/tarp/article/viewFile/14893/12698
6. Котляров И. Д. Самоплагиат в научных публикациях. / И. Д. Котляров // Научная периодика: проблемы и решения. – С. 6-12.
7. Ліннік І. Програмне забезпечення для виявлення плагіату: практичний аспект [Електронний ресурс] / І. Ліннік // Науковий блог НаУ «Острозька Академія». – 2013.– Режим доступу: http://naub.oa.edu.ua/2013/prohramne-zabezpechennya-dlya-vyyavlennya-plahiatu-praktychnyj-aspekt/
8. Михайловський Ю.Б. Система Anti-Plagiarism як інструмент запобігання плагіату в навчальній та науковій діяльності / Ю.Б. Михайловський, Н.А. Длугунович // Вісник Хмельницького національного університету. Технічні науки.. – 2013. – № 3.— С. 162—168.
9. Петренко В.С. Поняття та види плагіату. / В.С. Петренко // Часопис цивілістики.. – 2013. – Вип. 14.— С. 128—131.
10. Поповський О. І. Огляд програм порівняльного аналізу на збіг / О. І. Поповський // Збірник тез доповідей 2-го Кіровоградського соціально-економічного форуму «Інформаційне суспільство і влада». – Кіровоград. – 2013. – С. 99—100.
11. Про авторське право і суміжні права: Закон України від 23.12.1993 № 3792-XII // Відомості Верховної Ради України від 29.03.1994 року. – № 13. – Ст. 64.
12. Чижова А. А. Алгоритми пошуку плагіату [Електронний ресурс] / А. А. Чижова // Східно-Європейський журнал передових технологій. – 2010. - 4,№ 2 (46), С. 13-16. – Режим доступу: http://journals.uran.ua/eejet/article/view/2921
13. Шарапова Е. В. “Универсальная система проверки текстов на плагиат «Автор.net»” / Е. В. Шарапова, Р. В. Шарапов // Информатика и её применения – 2012. – № 3 (6).— С. 52–58
14. Шинкаренко В. І. Система контролю плагіату в студентських роботах [Електронний ресурс] / В. І. Шинкаренко, О. С. Куроп’ятник // Східно-Європейський журнал передових технологій. – 2012. – Том. 4. - № 2(58). - С. 32-36, 2012. – Режим доступу: Ihttp://journals.uran.ua/eejet/article/view/4214
15. Энциклопедический словарь Ф.А. Брокгауза и И.А. Ефрона [Текст] / Ф.А. Блокгауз, И.А. Ефрон. - в 82 основ.и 4 допол. полутомах. - СПГ, 1898. - Т. XXIII, кн. 48. - 961 с.
16. Ali A. M. El Tahir Overview and Comparison of Plagiarism Detection Tools [online] / A. M. El T. Ali, H. M. D. Abdulla, V. Snasel// CEUR Workshop Proceedings. – 2011. – Vol. 706.— P. 161-172.– Available from: http://ceur-ws.org/Vol-706/poster22.pdf
17. Anzelmi D. Plagiarism Detection Based on SCAM Algorithm [online] / D. Anzelmi, D. Carlone, F. Rizzello, R. Thomsen, D. M. Akbar Hussain// Proceedings of the International MultiConference on Engineers and Computer Scientists 2011. – 2011. – № 6 (1).— P. 272-277– Available from: http://www.ijsce.org/attachments/File/v2i5/E0984092512.pdf
18. Berlinck R.G. S. The academic plagiarism and its punishments - a review / R.G. S. Berlinck // Brazilian Journal of Pharmacognosy. – 2011. - № 21(3) — P. 365-372.
19. Bull J. Technical Review of Plagiarism Detection Software Report [online] / J. Bull, E. Coughlin, C. Collins, D. Sharp – Luton. –2000. – p. 36 – Available from:
20. Hunes C. Examining Anti-Plagiarism Software: Choosing the Right Tool [Електронний ресурс] / C. Hunes, J. Stiffler, M. Malsed. – Clermont. –2003. – 60 p. – Available from: http://pdf.brand2.net/examining-anti-plagiarism-software-choosing-the-right-tool-w2556/
21. Hariharan Sh. Automatic Plagiarism Detection Using Similarity Analysis [online] / Sh. Hariharan // The International Arab Journal of Information Technology. – 2012. – № 4 (9).— P. 322-326 – Available from : http://www.ccis2k.org/iajit/PDF/vol.9,no.4/2796-4.pdf
22. Kharat R. Semantically Detecting Plagiarism for Research Papers [online] / R. Kharat, P. M. Chavan, V. Jadhav, K. Rakibe // International Journal of Engineering Research and Applications (IJERA) – 2013. – № 3 (3).— P. 077-080– Available from: http://www.ijera.com/papers/Vol3_issue3/P33077080.pdf
23. Lancaster, T. Effective and Efficient Plagiarism Detection : PhD thesis / Lancaster Thomas. – London, 2003. – 228 p.

24. Lancaster T. Classifications of Plagiarism Detection Engines [online] / T. Lancaster, F. Culwin // Innovation in Teaching and Learning in Information and Computer Sciences. 2005. – №2 (4). – 16 p. – Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.184.2406&rep=rep1&type=pdf
25. Marcial D. E. ICT Skills Enhancement Training in Teacher Education: The Case in Central Visayas, Philippines [online] / D. E. Marcial, M. S. Fortich, J. B. Rendal// Information Technologies and Learning Tools. – 2014. – № 1 (39).— P. 230-240 – Available from: http://journal.iitta.gov.ua/index.php/itlt/article/view/964/749
26. Meyer zu Eissen S. Intrinsic Plagiarism Detection [online] / S. Meyer zu Eissen, B. Stein// Advances in Information Retrieva: Proceedings of the 28th European Conference on IR Research, ECIR. – 2006. – P. 565-569 – Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.5366&rep=rep1&type=pdf
27. Shenoy M. Automatic Plagiarism Detection Using SimilarityAnalysis [online] / M. Shenoy, K. C.Shet, U. D. Acharya// Advanced Computing: An International Journal (ACIJ). – 2012. – № 3 (3).— P. 59-62 – Available from: http://airccse.org/journal/acij/papers/0512acij06.pdf
28. Singh R. Duplicity Detection System for Digital Documents [online] / R. Singh, C. Dutta// International Journal of Soft Computing and Engineering (IJSCE). – 2012. – № 5 (2).— P. 24-28 – Available from: http://www.iaeng.org/publication/IMECS2011/IMECS2011_pp272-277.pdf
29. Tschuggnall M. Detecting Plagiarism in Text Documents through Grammar-Analysis of Authors [online] / M. Tschuggnall, G. Specht// 15th GI-Symposium Database Systems for Business, Technology and Web, 11th March - 15th March, 2013. – 2013. — P. 241-259 – Available from: http://www.btw-2013.de/proceedings/Detecting%20Plagiarism%20in%20Text%20Documents%20through%20Grammar Analysis%20of%20Authors.pdf

Матеріал надійшов до редакції 01.04.2014р.

ИНСТРУМЕНТАРИЙ ПОИСКА ПЛАГИАТА В НАУЧНЫХ РАБОТАХ: АНАЛИЗ ПРОГРАММНЫХ РЕШЕНИЙ

Лупаренко Лилия Анатольевна
ведущий инженер
Институт информационных технологий и средств обучения НАПН Украины, г. Киев, Украина
lisoln@rambler.ru

Аннотация. В статье рассмотрено понятие плагиата, самоплагиата и их основные виды. Приведена классификация инструментов автоматического отслеживания заимствований в текстовых документах, а также основные методы, лежащие в основе их функционирования. Рассмотрен ряд популярных программных средств и онлайн ресурсов для проверки текстов на совпадение, проведен их детальный анализ, тестирование и оценка по заранее определенным программно-техническим характеристикам, производительности функционирования, а также полученным в процессе проверки результатам. Представлено рекомендации относительно оптимального инструментария, который может быть использован в работе научных учреждений и редакций научных изданий.

Ключевые слова: плагиат; средства автоматического поиска плагиата; программное обеспечение; онлайн ресурс.

PLAGIARISM DETECTION TOOLS FOR RESEARCH WORKS: ANALYSIS OF SOFTWARE SOLUTIONS

Liliia A. Luparenko
lead engineer
Institute of Information Technologies and Learning Tools of NAPS of Ukraine, Kyiv, Ukraine
lisoln@rambler.ru

Abstract. The article deals with the concept of plagiarism, self-plagiarism and their basic types. It is described a classification of plagiarism automatic detection tools for text documents and the basic methods that underlie their functioning. It is considered a number of popular software and online resources to check text similarity, carried out their detailed analysis, testing and evaluation according to defined software specifications, productivity of functioning and obtained results during the checking. There are presented recommendations of the best tools that can be used by the research institutions and scientific journals' editors.

Keywords: plagiarism; plagiarism detection software; software; online resources.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Automated system of checking text for plagiarism / V. B. Mokin, V. V. Voitko, S. V. Bevz [and others] // Visnyk Vinnytskoho politekhnichnoho instytutu. – 2010. – № 5.— P. 12—17. (in Ukrainian)
2. Bykov V. Yu. The Theoretical and Metodological Grounds for Modeling of Educational Environment of Modern Pedagogic Sysytems [online] / V. Yu. Bykov // Informatsiini tekhnolohii i zasoby navchannia. – 2006. – № 1 (1). – Available from: http://journal.iitta.gov.ua/index.php/itlt/article/view/285/271. (in Ukrainian)
3. Bolilyi V. O. Check the Uniqueness of the Text in the Assessment of Student Work Creative or Exploratory / V.O. Bolilyi, V. V. Kopotii // Naukovi zapysky NDU im. M. Hoholia. – 2011. – № 7 (34).— P. 134—145. (in Ukrainian).
4. The Great Soviet Encyclopedia [text] : 30 volumes / chief editor. A. M. Prohorov. - M. : «Sovetskaja Enciklopedija». - 1975. - Vol. 19. - 647 p. (in Russian)
5. Kvashina Ju. A. Methods of Finding Duplicate Texts Arranged Scientific Style [online] / Ju. A. Kvashina // Tehnologicheskij audit i rezervy proizvodstva. – 2013. – № 3 (3).— P. 16-20 – Available from: http://journals.uran.ua/tarp/article/viewFile/14893/12698. (in Russian)
6. Kotljarov I. D. Plagiarism in scientific publications. / I. D. Kotljarov // Nauchnaja periodika: problemy i reshenija. – P. 6-12.
7. Linnik I. Software to Detect Plagiarism: Practical Aspects [online] / I. Linnik // Naukovyi blog NaU «Ostrozka Akademiia». – 2013.– Available from: http://naub.oa.edu.ua/2013/prohramne-zabezpechennya-dlya-vyyavlennya-plahiatu-praktychnyj-aspekt/. (in Ukrainian)
8. Mykhailovskyi Yu. B. Anti-Plagiarism System as a Tool for Plagiarism Preventing in Educational and Research Activities/ Yu. B. Mykhailovskyi, N. A. Dluhunovych // Visnyk Khmelnytskoho natsionalnoho universytetu. Tekhnichni nauky. – 2013. – № 3.— P. 162—168. (in Ukrainian)
9. Petrenko V. S. The concept and types of plagiarism. / V.S. Petrenko // Chasopys tsyvilistyky. – 2013. – Vol. 14.— P. 128—131.
10. Popovskyi O. I. Review of Applications for the Comparative Analysis of Matches / O. I. Popovskyi // Zbirnyk tez dopovidei 2-ho Kirovohradskoho sotsialno-ekonomichnoho forumu «Informatsiine suspilstvo vlada». – Kirovohrad. – 2013. – P. 99—100.
11. On Copyright and Related Rights: Law of Ukraine on 23.12.1993 № 3792-XII // Vidomosti Verkhovnoi Rady Ukrainy vid 29.03.1994 roku. – № 13. – 64 p.
12. Chyzhova A. A. Algorithms of Finding Plagiarism [online] / A. A. Chyzhova // Skhidno-Yevropeiskyi zhurnal peredovykh tekhnolohii. – 2010. - 4, № 2 (46), P. 13-16. – Available from: http://journals.uran.ua/eejet/article/view/2921. (in Ukrainian)
13. Sharapova E. V. Universal System of text checking for plagiarism «Avtor.net» / E. V. Sharapova, R. V. Sharapov // Informatika i ejo primenenija. – 2012. – № 3 (6).— P. 52–58. (in Russian)
14. Shynkarenko V. I. The System for Control of Plagiarism in Student Papers [online] / V. I. Shynkarenko, O. S. Kuropiatnyk // Skhidno-Yevropeiskyi zhurnal peredovykh tekhnolohii. – 2012. – Том. 4. - № 2 (58). - P. 32-36, 2012. – Available from: http://journals.uran.ua/eejet/article/view/4214. (in Ukrainian)
15. F. A. Brokgauz and I.A. Efron Collegiate Dictionary [text] / F. A. Blokgauz, I A. Efron. - 82 main and 4 additional volumes. – St. Petersburg, 1898. - Vol. XXIII, book. 48. - 961 p. (in Russian)
16. Ali A. M. El Tahir Overview and Comparison of Plagiarism Detection Tools [online] / A. M. El T. Ali, H. M. D. Abdulla, V. Snasel// CEUR Workshop Proceedings. – 2011. – Vol. 706.— P. 161-172.– Available from: http://ceur-ws.org/Vol-706/poster22.pdf. (in English)
17. Anzelmi D. Plagiarism Detection Based on SCAM Algorithm [online] / D. Anzelmi, D. Carlone, F. Rizzello, R. Thomsen, D. M. Akbar Hussain// Proceedings of the International MultiConference on Engineers and Computer Scientists 2011. – 2011. – № 6 (1).— P. 272-277– Available from: http://www.ijsce.org/attachments/File/v2i5/E0984092512.pdf. (in English)

18. Berlinck R.G. S. The academic plagiarism and its punishments - a review / R.G. S. Berlinck // Brazilian Journal of Pharmacognosy. – 2011. - № 21(3) — P. 365-372. (in English)
19. Bull J. Technical Review of Plagiarism Detection Software Report [online] / J. Bull, E. Coughlin, C. Collins, D. Sharp – Luton. – 2000. – 36 p.– Available from: http://www.jisc.ac.uk/uploaded_documents/luton.pdf. (in English)
20. Hunes C. Examining Anti-Plagiarism Software: Choosing the Right Tool [online] / C. Hunes, J. Stiffler, M. Malsed. – Clermont. –2003. – 60 p. – Available from: http://pdf.brand2.net/examining-anti-plagiarism-software-choosing-the-right-tool-w2556/. (in English)
21. Hariharan Sh. Automatic Plagiarism Detection Using Similarity Analysis [online] / Sh. Hariharan // The International Arab Journal of Information Technology. – 2012. – № 4 (9).— P. 322-326 – Available from : http://www.ccis2k.org/iajit/PDF/vol.9,no.4/2796-4.pdf. (in English)
22. Kharat R. Semantically Detecting Plagiarism for Research Papers [online] / R. Kharat, P. M. Chavan, V. Jadhav, K. Rakibe // International Journal of Engineering Research and Applications (IJERA) – 2013. – № 3 (3).— P. 077-080– Available from: http://www.ijera.com/papers/Vol3_issue3/P33077080.pdf. (in English)
23. Lancaster, T. Effective and Efficient Plagiarism Detection : PhD thesis / Lancaster Thomas. – London, 2003. – 228 p. (in English)
24. Lancaster T. Classifications of Plagiarism Detection Engines [online] / T. Lancaster, F. Culwin // Innovation in Teaching and Learning in Information and Computer Sciences. 2005. – №2 (4). – 16 p. – Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.184.2406&rep=rep1&type=pdf. (in English)
25. Marcial D. E. ICT Skills Enhancement Training in Teacher Education: The Case in Central Visayas, Philippines [online] / D. E. Marcial, M. S. Fortich, J. B. Rendal// Information Technologies and Learning Tools. – 2014. – № 1 (39).— P. 230-240 – Available from: http://journal.iitta.gov.ua/index.php/itlt/article/view/964/749. (in English)
26. Meyer zu Eissen S. Intrinsic Plagiarism Detection [online] / S. Meyer zu Eissen, B. Stein// Advances in Information Retrieva: Proceedings of the 28th European Conference on IR Research, ECIR. – 2006. – P. 565-569 – Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.5366&rep=rep1&type=pdf. (in English)
27. Shenoy M. Automatic Plagiarism Detection Using SimilarityAnalysis [online] / M. Shenoy, K. C.Shet, U. D. Acharya// Advanced Computing: An International Journal (ACIJ). – 2012. – № 3 (3).— P. 59-62 – Available from: http://airccse.org/journal/acij/papers/0512acij06.pdf. (in English)
28. Singh R. Duplicity Detection System for Digital Documents [online] / R. Singh, C. Dutta// International Journal of Soft Computing and Engineering (IJSCE). – 2012. – № 5 (2).— P. 24-28 – Available from: http://www.iaeng.org/publication/IMECS2011/IMECS2011_pp272-277.pdf. (in English)
29. Tschuggnall M. Detecting Plagiarism in Text Documents through Grammar-Analysis of Authors [online] / M. Tschuggnall, G. Specht// 15th GI-Symposium Database Systems for Business, Technology and Web, 11th March - 15th March, 2013. – 2013. — P. 241-259 – Available from: http://www.btw-2013.de/proceedings/Detecting%20Plagiarism%20in%20Text%20Documents%20through%20Grammar Analysis%20of%20Authors.pdf. (in English)