

УДК 681.3.07

**Тесьолкін О.І.**  
*кандидат технічних наук,  
доцент кафедри економічної кібернетики  
Полтавського національного технічного університету  
імені Юрія Кондратюка*

**Кузьменко В.В.**  
*студентка  
Полтавського національного технічного університету  
імені Юрія Кондратюка*

## СПОСОБИ ТА МЕХАНІЗМИ ПІДБОРУ УНІКАЛЬНОГО КОНТЕНТУ ПІД ЧАС ПЕРЕВІРКИ ТЕКСТУ НА ПЛАГІАТ

### THE WAYS AND MECHANISMS OF UNIQUE CONTENT SELECTION FOR PLAGIARISM CHECKING

#### АНОТАЦІЯ

У статті проведено дослідження програмного забезпечення, за допомогою якого можна здійснювати перевірку тексту на плагіат, та виявлено найкращі програмні продукти. Розроблено принципи, методи та алгоритми пошуку ідентичних символічних масивів у текстових документах із метою виявлення плагіату, а також комп'ютерна програма, яка дає змогу комплексно здійснювати перевірку текстів на наявність плагіату. Створено і протестовано власну програму для вирішення цього завдання. Даний програмний продукт може бути використаний навчальними закладами як інструмент для перевірки на наявність плагіату рефератів, курсових та дипломних робіт, наукових статей тощо.

**Ключові слова:** унікальність, перевірка, плагіат, комп'ютерна програма.

#### АННОТАЦИЯ

В статье проведено исследование программного обеспечения, с помощью которого можно осуществлять проверку текста на плагиат, и выявлены лучшие программные продукты. Разработаны принципы, методы и алгоритмы поиска идентичных символьных массивов в текстовых документах с целью выявления плагиата, а также компьютерная программа, которая позволяет комплексно осуществлять проверку текстов на наличие плагиата. Создана и протестирована собственная программа для решения этой задачи. Разработанный программный продукт может быть использован учебными заведениями как инструмент для проверки на наличие плагиата рефератов, курсовых и дипломных работ, научных статей и тому подобное.

**Ключевые слова:** уникальность, проверка, плагиат, компьютерная программа.

#### ANNOTATION

Due to the fast Internet technology development, different types of the Information has become much more accessible. Modern society is living in information vacuum. Writing the texts, author presents not the only his personal opinion, but also a variety of information sources. The information technology development has greatly simplified the process of information exchange. Therefore, the actual problem is the data identification in the analysis of information flows. For writing a scientific text, one of the main requirements is uniqueness, which can be verified by special software. Through this program we can check, if the text is original or was copied from another source. Therefore, the actual topic today is the development of a computer program to check the texts for plagiarism. Evaluating a scientific work on uniqueness, using such kind of software, will clearly indicate which parts of the text have been used and from which sources they were taken. Selection of the unique Web-resources content is not a minor task of the modern search engines. For this purpose the most important is, the analysis of the basic, most frequently used programs and online services, which will be the object of our study, and choose

the best of them, which can be used to verify the text. Was made a research about different programs, which can be used to check the text on plagiarism and identify the best software product for this purpose. The principles, methods and algorithms for finding identical character are set in text documents, to identify plagiarism. The computer program allows checking complex texts for plagiarism. In conclusion, was created and tested own program to solve this problem. Developed software can be used at schools, as a tool to check for plagiarism essays and dissertations, research works etc.

**Keywords:** uniqueness, test, plagiarism computer program.

**Постановка проблеми.** Сучасний розвиток інформаційних технологій, а саме розвиток мережі Інтернет, значно спростили процес обміну інформацією. Доступність інформаційних ресурсів, безперечно, сприяє розвитку особистості та суспільства в цілому. У зв'язку з цим актуальним є завдання ідентифікації даних у процесі аналізу інформаційних потоків, що є головним завданням сучасних пошукових систем, спрямованих на підбір унікального контенту Web-ресурсів. Перевірка текстового контенту здійснюється за допомогою програмних засобів [1, с. 25]. Для перевірки наукових робіт на запозичення потрібно використовувати таке програмне забезпечення, яке буде чітко вказувати, які частини запозичення тексту було використано та з яких джерел, щоб дійсно можна було перевірити, чи був плагіат.

Для цього важливим є аналіз основних, найбільш вживаних програм та онлайн-сервісів, які будуть об'єктом нашого дослідження.

**Аналіз останніх досліджень і публікацій.** Питання авторських прав досліджувалися багатьма авторами. Провідними роботами із цієї тематики є праці В.Я. Іонаса, О.С. Іюффе, В.І. Серебровського, О.Д. Святоцького, Г.Ф. Шершеневича. Грунтовними працями сучасних дослідників є роботи: С.О. Глотова, В.В. Коноваленка, Л.Г. Максимова, А.П. Панкєєва, І.В. Порало, І.М. Рассолова, Т.І. Стексової, Р.В. Шарапова, Е.В. Шарапової, О.О. Штефан та ін.

**Мета статті** полягає в аналізі програмного забезпечення, що надає можливість перевіряти текстовий матеріал на унікальність.

**Виклад основного матеріалу дослідження.** Плагіат у студентському середовищі становить загрозу з цілої низки причин. Зокрема, він не сприяє розвитку креативного мислення і творчих здібностей студента, спотворює реальну картину знань, вмінь і навичок, ускладнює процес оцінювання результатів викладачем, призводить до виникнення недобросовісної конкуренції у студентському середовищі. Крім того, це явище вкрай негативно сприймається світовою науковою спільнотою, а отже, його поширення створює перешкоди ефективної інтеграції у міжнародне освітнє й наукове середовище.

Загальнодоступна електронна форма викладення матеріалів, досліджень та наукових праць не лише дає змогу багаторазово використовувати і розповсюджувати інформацію, а й полегшує процес виявлення плагіату. Так, виділяють декілька основних ознак плагіату:

- написання роботи складною мовою, наявність професійних термінів, висновків, що виходять за рівень знань студента;

- стилістична неоднорідність дослідження;
- невідповідність роботи вимогам викладача, поставленому завданню, лекційному матеріалу чи роздатковим матеріалам;

- бібліографія, що викликає сумніви (занадто довгий перелік, посилання на неофіційні Інтернет-ресурси (наприклад, «банки рефератів»); цитати із книжок більш пізнього видання, або, навпаки, застарілі дані; відсутність у бібліотеках зазначених джерел).

Існує декілька видів технічних засобів, які шляхом автоматичного порівняння текстових документів в електронній формі з базою даних інших текстів створюють умови для швидкого виявлення плагіату [2, с. 10].

Програми, які були обрані для тестування: Copyscape, [www.antiplagiat.ru](http://www.antiplagiat.ru), Etxt Антиплагіат, AdvegoPlagiatus, DupeFreePro, Viper, CognitiveTextAnalyzer.

Програма [www.copyscape.com](http://www.copyscape.com) [7] – західний сервіс для перевірки на унікальність, але може перевіряти і український текст. Перевіряється контент, який уже був викладений на якомусь ресурсі в Мережі. Для перевірки наукових робіт, які є в електронному варіанті, недоцільно використовувати даний сервіс. Для перевірки на унікальність наукові роботи потрібно викладати на веб-сайті, створити електронний блог.

Тобто щойно написаний текст потрібно залити на хостинг і у віконці вставити адресу (URL) сторінки. На сторінці з'являться фрази з тексту, які зустрічаються на інших сайтах.

Програма [www.antiplagiat.ru](http://www.antiplagiat.ru) [6] – безкоштовний сервіс, за допомогою якого можна перевірити унікальність тексту. Перевірку можна провести двома способами: перший – це скачати програму і встановити її на свій комп'ютер; і друга – скористатися онлайн-сервісом. Скопіювавши текст у віконце сервісу, натискаємо кнопку «Проверить», і текст буде перевірятись на унікальність. Тестування онлайн-сервісу

показало на 15% більше унікального тексту, ніж програми. Отже, можна зробити висновок про низьку якість роботи онлайн-сервісу.

Програма *Etxt Антиплагіат* – програма для перевірки унікальності тексту, яка дає змогу провести детальний аналіз тексту і визначити оригінальність статті в процентному співвідношенні. *Etxt Антиплагіат* має також онлайн-версію для перевірки коротких текстів (до 5 000 знаків).

Пакемна перевірка дає змогу завантажити до програми *Etxt Антиплагіат* папку з файлами для одночасної перевірки. Для того щоб не завантажувати файли по одному, існує пакемна перевірка. Для перевірки тексту потрібно вставити його у віконце сервісу і натиснути кнопку «Перевірити на унікальність». Програма працює добре, хоча трохи повільно (одна наукова робота на 34 сторінки формату \*.doc. перевірялася більше години). Інтерфейс зрозумілий і зручний, не унікальний текст розпізнаний, результат виглядає правдоподібно. За допомогою цієї програми можна перевіряти наукові тексти на унікальність.

Програма *Advego Plagiatus* – програма пошуку в Інтернеті часткових або повних копій текстового документа з достатньо простим інтерфейсом. *Advego Plagiatus* показує ступінь унікальності тексту, джерела тексту, відсоток збігу тексту. Також програма перевіряє унікальність зазначеного URL. За принципом роботи майже не відрізняється від програми *Etxt Антиплагіат*, хоча *Etxt Антиплагіат* має більше можливостей. Текст, який перевіряється у *AdvegoPlagiatus*, ділиться на фрази, а потім відбувається пошук подібних фраз в Інтернеті через пошукові машини. Якщо фраза знайдена в Інтернеті, то вона вважається неоригінальною і підсвічується жовтим виділенням. Схожих фраз в Інтернеті можна знайти багато, але це не означає, що був факт плагіату. Дана програма неспроможна правильно оцінити унікальність, тому її не варто використовувати.

Програма *DupeFreePro* виявляє ступінь схожості текстів та дає змогу перевіряти ступінь унікальності тексту. Функція перевірки текстів на унікальність реалізована не досить добре і не завжди показує точний результат. За допомогою *DupeFreePro* із легкістю можна порівнювати декілька текстів між собою на схожість. Ця особливість програми дає можливість використовувати її викладачами у ВНЗ для перевірки своєї бази студентських робіт по роках і дізнаватися, чи використовують студенти наукові роботи своїх попередників. Для нашого дослідження використання програми буде недоцільним, оскільки нам потрібно визначити рівень унікальності текстів.

*Viper* – 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 водночас два документи і виявити їх рівень схожості. Однаковий текст, який є в обох документах, підсвічується іншим кольором, тобто показуються скопійовані частини. *Viper* є платною програмою.

Система автоматичного виявлення плагіату в текстах *CognitiveTextAnalyzer*, яку представила компанія Cognitive Technologies. Принцип роботи *CognitiveTextAnalyzer* полягає в такому. У систему вводиться досліджуваний текст. Далі *CognitiveTextAnalyzer* 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 об'єму тексту і різноманітності конструкцій, що зустрічаються в тексті. Швидкість розбору тексту на сучасному офісному комп'ютері становить 450–500 сторінок тексту на хвилину. Порівняно з іншими програмами ця програма працює дуже швидко.

Для вирішення проблеми плагіату була створена власна програма на мові програмування C#.

Розроблена комп'ютерна програма AntyCopyPaste для перевірки текстів на плагіат реалізована засобами мови програмування C#. Програма дає змогу здійснювати пошук ідентичних масивів текстових даних шляхом порівняння досліджуваного файлу з наявними текстовими файлами у попередньо створених базах даних. У програмі реалізовані два методи пошуку плагіату, а саме: метод ідентифікації повного співпадіння речень та метод аналізу кількості співпадіння слів. Інтерфейс програми забезпечує користувачу можливість вибору методу пошуку та введення його параметрів (рис. 1). Метод ідентифікації повного співпадіння речень дає змогу швидко виявити факти наявності плагіату в текстовому документі. Метод аналізу кількості співпадіння слів реалізує пошук послідовностей заданої кількості слів (5–10), які співпадають в обох порівнюваних файлах. Окрім реалізації алгоритму символного порівняння слів, метод забезпечує можливість ідентифікації слів за їх коренем.

Після завершення процесу перевірки текстового файлу на плагіат користувачу виводиться результат роботи програми у вигляді прогрес-бару з визначеного відсотку наявності плагіату в роботі.

Крім того, програма надає можливість переглядати речення, в яких був виявлений плагіат.

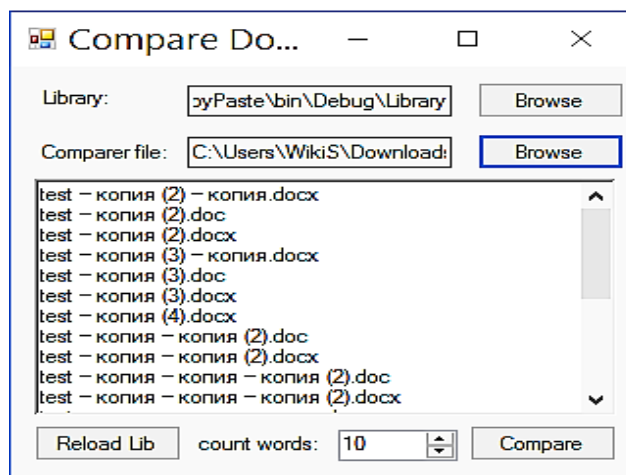


Рис. 1. Інтерфейс програми AntyCopyPaste

**Висновки.** Проведений аналіз програм для виявлення плагіату показав, що сучасні засоби перевірки тексту на плагіат недосконалі, існує багато програм для виявлення плагіату, але вони не є універсальними та легкодоступними, часто платними у використанні. Розроблений програмний продукт може бути використаний навчальними закладами як інструмент для перевірки на наявність плагіату рефератів, курсових та дипломних робіт, наукових статей тощо.

#### БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Гилинский Я. Тотальный плагиат как норма российской научной жизни / Я. Гилинский [Электронный ресурс]. – Режим доступа: [http://sartraccc.ru/i.php?oper=read\\_file&filename=Pub/gilinsky\(17-10-07\).htm](http://sartraccc.ru/i.php?oper=read_file&filename=Pub/gilinsky(17-10-07).htm).
2. Заимствования и плагиат. Накипь и ржавчина в науке [Электронный ресурс]. – Режим доступа: <http://trv-science.ru/2009/12/22/sonin-tabl/>.
3. Пархомов В.А. Плагиат как угроза модернизации экономики / В.А. Пархомов. – Иркутск, 2010. – 141 с.
4. Порало І.В. Інтелектуальний капітал України в полоні плагіату / І.В. Порало // Теорія і практика управління соціальними системами: філософія, психологія, педагогіка, соціологія. – 2009. – № 4. – 194 с.
5. Тест на плагіат студентських робіт – престижно і непрактично [Електронний ресурс]. – Режим доступу: [http://vidido.ua/index.php/pogliad/article/test\\_na\\_plagiat\\_students\\_kih\\_robit\\_-\\_prestizhno\\_i\\_nepraktichno](http://vidido.ua/index.php/pogliad/article/test_na_plagiat_students_kih_robit_-_prestizhno_i_nepraktichno).
6. Antiplagiat [Електронний ресурс]. – Режим доступу: [www.antiplagiat.ru/](http://www.antiplagiat.ru/)
7. Copyscape [Електронний ресурс]. – Режим доступу: <http://www.copyscape.com>.