

УДК 004:001.891:177.9-022.241

DOI: 10.31866/2617-796X.4.2.2021.247496

Бережна Оксана,

магістр,

завідувач сектору організації електронних документів наукової бібліотеки,

Київський національний університет культури і мистецтв,

Київ, Україна

oberezhna23@gmail.com

<https://orcid.org/0000-0003-4792-3382>

Дорошенко Євгеній,

завідувач сектору мережевої та технічної підтримки наукової бібліотеки,

Київський національний університет культури і мистецтв,

Київ, Україна

knukim.journals@gmail.com

<https://orcid.org/0000-0003-1050-5478>

Богущ Ірина,

магістр,

методист вищої категорії наукової бібліотеки,

Київський національний університет культури і мистецтв,

Київ, Україна

ira.bogush89@ukr.net

<https://orcid.org/0000-0003-4138-8534>

ДУМАЙ ЯК ШАХРАЙ, АБО ЯК ПІДВИЩИТИ УНІКАЛЬНІСТЬ ТЕКСТУ

Мета статті – дослідити й теоретично обґрунтувати способи підвищення відсотка унікальності наукової праці, визначивши основні методи модифікації тексту на основі вивчення принципів роботи сайтів, діяльність яких спрямована на штучне підвищення оригінальності поданого матеріалу.

Методи дослідження. Для досягнення мети використано такі загальнонаукові методи дослідження, як емпіричний, метод аналізу, логічний метод, порівняння, експеримент і метод візуалізації.

Наукова новизна. Уперше описано принцип роботи таких сайтів, як «Перший український сервіс підвищення унікальності та зменшення плагіату», «Підвищити-антиплагіат.рф», «Анти-антиплагіат.рф», «Антиплагіат кілер», механізм дії яких полягає в штучному підвищенні оригінальності тексту.

Висновки. У процесі дослідження встановлено, що яскраві заголовки сайтів, які обіцяють підняти відсоток унікальності роботи, не є гарантією того, що подібне шахрайство не помітять під час перевірки. Виявлення методів модифікації тексту наукового дослідження – одне із завдань фахівця, який займається перевіркою академічних робіт, а тому зростає потреба щоразу поглиблювати знання у сфері комп'ютерних технологій, розвивати

здатність логічно мислити, підвищувати рівень кваліфікації, ознайомлюючись з досвідом колег. Наявність антиплагіатних систем, програмне забезпечення яких постійно оновлюється, спрощує пошук збігів, дає змогу зекономити час та людські ресурси, спрямовані на аналіз надісланого файлу та виявлення ознак порушення академічної доброчесності.

Ключові слова: плагіат; модифікації тексту; XML; підвищення унікальності; Unicheck.

Вступ. Високий відсоток унікальності тексту на сьогодні є одним з критеріїв успішного оформлення результатів наукового дослідження. Проте з упровадженням контролю якості наукових робіт (статей, магістерських, дисертацій тощо) і з розробленням систем перевірки наявності плагіату в недобросовісних авторів виникла потреба штучного підвищення оригінальності тексту, що спонукає їх звертатися до пошуку відповідних сайтів чи приватних установ, діяльність яких пов'язана з наданням окреслених послуг. Це зумовило збільшення кількості модифікацій у текстах з метою маскування запозичень із чужих праць. Системи пошуку збігів не можуть виявити всі прояви академічного шахрайства, саме тому фахівцям, які перевіряють наукові дослідження, необхідно розвивати вміння та навички використання інформаційних технологій, зокрема розуміти властивості тексту й алгоритм пошуку плагіату, здійснювати не тільки технічні операції (наприклад, завантаження/перевірку документів), а й працювати безпосередньо з поданим матеріалом, а також уміти аналізувати звіт, в якому зазначено відсоток подібності. Результати роботи спеціальних сервісів з пошуку збігів (звіти) з високим показником оригінальності тексту не є свідченням дотримання академічної доброчесності, адже сама програма – це лише інструмент, за допомогою якого можемо швидко проаналізувати роботу й виявити потенційні ознаки плагіату. Водночас за умови ретельного аналізу документа, навіть не беручи до уваги звіт, маємо змогу виявити модифікований текст. Саме тому запорукою найкращого результату перевірки якості наукових праць є «синергія машини і людини» (Аналізуємо звіт про перевірку на плагіат» – Unicheck та Academic IQ).

Джерельну базу дослідження становлять праці, присвячені вивченню текстових модифікацій, а також авторський проєкт Олександра Павлова «Спосіб обходу антиплагіату “зі схованими словами”» (<https://apavlov.ru/>). У 2010 році Дж. Хізер (Heather, 2010) одним з перших дослідив й описав недоліки в програмному забезпеченні щодо виявлення збігів. Є. Чиркін (2013) у статті «Використання систем антиплагіату в освіті» наводить алгоритм роботи сервісів пошуку запозичень. О. Тугарев робить спробу класифікувати способи підвищення унікальності запозиченого тексту. Для розрізнення мов розмітки стандарту Open XML у пригоді стала книга Воутера Ван Вугта «Open XML Коротко та доступно» (2007) з документацією формату Open XML. Виявлення плагіату в наукових публікаціях, класифікацію, стан і проблеми боротьби з плагіатом в українській освіті вивчали такі дослідники, як М. Карпенко, І. Кременовська, О. Ляска, О. Прокопова, О. Рижко, К. Сопова, А. Штефан та інші, проте у вітчизняних працях не описано принцип роботи сайтів («Перший український сервіс підвищення унікальності та зменшення плагіату», «Підвищити-антиплагіат.рф», «Анти-антиплагіат.рф», «Антиплагіат кілер»), механізм дії яких

спрямований на штучне підвищення оригінальності тексту, в чому й виявляється *новизна дослідження*. Недостатня теоретична й методична розробленість проблеми виявлення збігів, її соціальна та наукова значущість, необхідність пошуків удосконалення методичних шляхів і підходів до досягнення та викриття всіх модифікацій у текстовому файлі зумовлюють *актуальність* обраної теми.

Мета статті – дослідити й теоретично обґрунтувати способи підвищення відсотка унікальності наукової праці, визначивши основні методи модифікації тексту на основі вивчення принципів роботи сайтів, діяльність яких спрямована на штучне підвищення оригінальності поданого матеріалу. Мета передбачає виконання таких завдань:

- схарактеризувати основні види модифікацій тексту, спрямованих на приховування запозичень із чужих праць від програмного забезпечення, за допомогою якого здійснюють перевірку наукових робіт на ознаки плагіату;
- висвітлити методи виявлення таких модифікацій;
- обґрунтувати можливості програмного забезпечення, зокрема онлайн-сервісу Unicheck;
- проаналізувати недоброчесні сайти, за допомогою яких підвищують відсоток оригінальності тексту.

Для досягнення мети використано такі загальнонаукові методи дослідження, як емпіричний, метод аналізу, логічний метод, порівняння, експеримент і метод візуалізації.

Результати дослідження. Виявлення плагіату здійснюють за допомогою спеціальних програм пошуку збігів, що можуть бути як хмарними сервісами, так і класичними застосунками, які встановлюють на комп'ютер. Більшість з них працює з файловими форматами doc, docx, rtf, txt, odt, html, zip, pdf тощо. У закладах освіти використовують файли для перевірок в основному у форматах doc та docx.

Для повного розуміння теми дослідження передусім потрібно вивчити теорію, зокрема знати, як взагалі влаштований будь-який сучасний текстовий документ на фундаментальному рівні. В електронних пристроях текст має зберігатись у вигляді послідовності бітів (біт – найменша одиниця зберігання інформації). Кожному символу має бути призначена своя унікальна послідовність бітів, це своєрідний ідентифікатор. Для цього розробили таблиці кодувань (скорочено їх називають кодування). Є багато різних стандартів, але на сьогодні найпопулярнішими вважаємо кодування Unicode, а саме UTF-8. Перевага Unicode в тому, що він містить цифрове представлення системи графічних знаків усіх народів світу, уживаних для писання в якій-небудь мові, а також спецсимволів, чого не було раніше у зв'язку з технічними обмеженнями попередніх стандартів кодування. Тепер можемо з упевненістю передавати текстову інформацію і знати, що отримувач побачить той самий текст, який йому надіслали, а не набір так званих «ієрогліфів», що неможливо прочитати (рис. 1).

Але такі потужні можливості стандартів Unicode мають і зворотний бік. Це відкриває широкий простір для модифікацій із вмістом текстових файлів (Представлені символи, таблиці кодировок). Важливим для розуміння принципу реалізації деяких модифікацій є знання про те, з чого складаються сучасні документи в най-

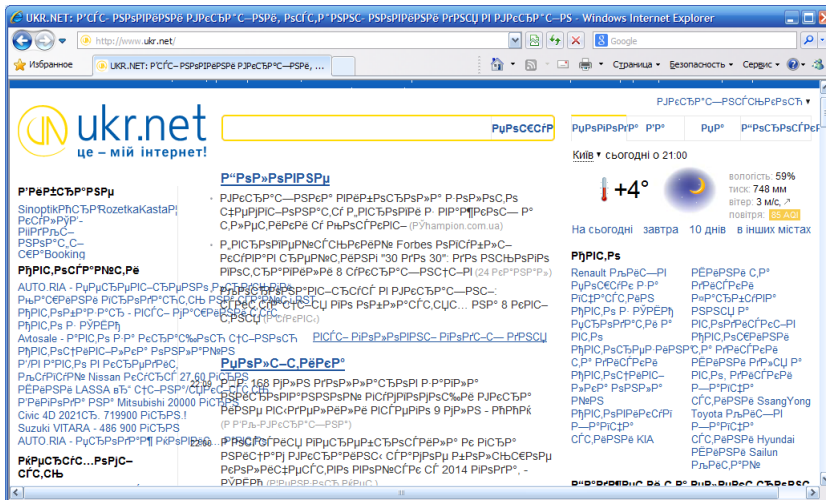


Рис. 1. Проблема з неправильним визначенням кодування наочно

поширенішому форматі, а саме Microsoft Office Open XML. Це сімейство відкритих форматів, заснованих на специфікації Open XML. XML – це гнучка мова розмітки інформації, яка застосовує теги. Завдяки останнім ми можемо впливати на форматування тексту: вирівнювати абзаци, змінювати розмір кегля, гарнітуру, міжрядковий інтервал тощо. Документи в цьому форматі можуть містити різні графічні файли та вбудовані шрифти. Усе це упаковується в контейнер-архів у форматі ZIP задля компактності та зменшення розміру кінцевого файлу і зберігається з розширенням docx (у такому разі з Microsoft Word) (Формати Open XML і розширення імен файлів). Тому маємо змогу розпакувати цей архів за допомогою будь-якої програми-архіватора й отримати у такий спосіб доступ до його вмісту для дослідження.

Системи перевірки текстів на збіги, як правило, працюють так: 1 етап – спеціальний алгоритм програми вилучає текст з наданого текстового документа (pdf, doc, docx, txt тощо); 2 етап – пошук вилученого тексту в базах даних з інтернету і/або в бібліотеці закладу вищої освіти; 3 етап – аналіз і кількісне визначення об’єму запозиченого тексту, формування звіту. Саме на другому етапі починаються найбільші труднощі, пов’язані з відображенням обсягу запозиченого тексту.

Поширеними є такі типи технічних модифікацій:

1. Заміна символів кирилиці на латиницю

Найчастіше трапляються випадки заміни символів кириличного тексту на символи латиниці (наприклад: «а» – «a», «о» – «o», «і» – «i» та ін.). Хоч ці символи на перший погляд є однаковими, можемо перевірити їхні коди. За допомогою текстового процесора Microsoft Word виділяємо символ, який нас цікавить, і натискаємо комбінацію клавіш Alt + X, унаслідок чого відобразиться код обраного символу. Українська буква «а» має код «0430», англійська «a» – «0061», українська «о» – «043E», англійська «o» – «006F», українська «і» – «0456», англійська «i» – «0069» і т. п. Таку підміну ще можна опосередковано помітити, якщо звер-

нути увагу на підкреслення текстовим процесором модифікованих слів з метою перевірки правильного написання.

Цей метод виявлення підміни символів є дуже довгим і трудомістким, адже потрібно вручну їх перевіряти та зіставляти з тими, що в таблиці юнікодів, проте сучасні системи перевірки тексту на збіги вже з легкістю виявляють такі підміни символів, що не робить цей метод підвищення відсотка унікальності тексту актуальним (рис. 2).

2

У сучасному світі музика позиціонується не просто як вид мистецтва. Розмаїття її образів виражається у контенті, джерелі інформації, мистецтві, способі комунікації, творчій інтелектуальній діяльності, комерційності продукту тощо. Але у журналістиці, музичний контент, в основному, сприймається як розважальний та рекреаційний, а музичні програми належать до ігор, шоу і концертів. Навіть крізь призму інформаційної та аналітичної складової журналістики, музичний контент переважно стосується зіркового музичного життя, аналізу і дослідження актуальних тем, що мають безпосереднє або опосередковане відношення до музики і світу шоу-бізнесу в цілому.

Рис. 2. Метод підміни символів (скріншот зі звіту системи виявлення збігів Unicheck)

2. Використання скріншотів з уривками запозиченого тексту

Досить простий спосіб схитрувати – вставити скріншоти з уривками тексту, проте не дуже ефективний, адже такі вставки легко виявити під час повного виділення абзацу. Фрагменти вставлених скріншотів розміщено в прямокутну рамку (рис. 3).



Рис. 3. Зображення, вставлене в документ замість тексту

3. Використання лігатур

Наступний метод маскуванню схожий з описаним у першому пункті. Але на відміну від нього заміна відбувається не на подібний символ з іншої мови, а на лігатуру, що являє собою злиття двох або більше літер. Їх можна підставляти замість кількох букв, проте сучасні системи пошуку плагіату в роботах можуть виявляти й такі підміни (рис. 4).

$AE \rightarrow \mathcal{A}E$	$ij \rightarrow ij$
$ae \rightarrow \mathcal{a}e$	$st \rightarrow \mathcal{s}t$
$OE \rightarrow \mathcal{O}E$	$ft \rightarrow \mathcal{f}t$
$oe \rightarrow \mathcal{o}e$	$et \rightarrow \mathcal{e}t$
$ff \rightarrow \mathcal{f}f$	$fs \rightarrow \mathcal{f}s$
$fi \rightarrow \mathcal{f}i$	$ffi \rightarrow \mathcal{f}fi$

Рис. 4. Порівняння лігатур з літерами

4. Метод модифікації шрифтів

Насамперед потрібно з'ясувати, що таке шрифт. Комп'ютерний шрифт – це набір растрових або векторних зображень символів алфавіту. Кожне з цих зображень має код, який буде вказувати на те, що саме цей символ і означає певну букву алфавіту або знак пунктуації. Для візуального сприйняття тексту людині достатньо лише зображення символу. Цю особливість використовують для описаного методу маніпуляції, послуговуючись яким застосовують різні програми (векторні графічні редактори, програми для редагування шрифтів тощо). Це дає змогу створювати подібний до оригінального шрифт, що тільки на вигляд видаватиметься справжнім. Коди символів не збігаються з їх зображеннями, і програма перевірки неправильно зіставляє значення тексту та його відображення. Тобто українська літера «і» може мати код «043B», що відповідає кириличній літері «л», а не «0456», який є правильним кодом символу «і» (рис. 5).

uni0436	uni0437	uni0438	uni0439	uni043A	uni043B	uni043C	uni043D
Ж	З	И	Й	К	і	М	Н
uni0442	uni0443	uni0444	uni0445	uni0446	uni0447	uni0448	uni0449
Т	У	Ф	Ц	Ч	Ш	Щ	
uni044E	uni044F	uni0451	uni0452	uni0453	uni0454	uni0455	uni0456
Ю	Я	ё	ђ	ѓ	є	ѕ	Л
uni045B	uni045C	uni045E	uni045F	uni0460	uni0461	uni0462	uni0463
ћ	ќ	ђ	џ	џ	џ	џ	џ

Рис. 5. Заміна символів у модифікованому шрифті

У тексті, надрукованому таким шрифтом, неозброєним оком неможливо побачити модифікації. Водночас алгоритм програми пошуку збігів сприймає його як набір букв, з яких складаються слова, що в такий спосіб втрачають своє значення, відповідно система не може знайти схожість з матеріалами, розміщеними в інтернеті чи бібліотеці, а тому визначає 100 % оригінальності. Для наочності на рисунку 6 зображено уривок тексту, надрукований модифікованим шрифтом. Тобто текст для програми перевірки здається нормальним, але через модифікований шрифт у словах ніби просто є помилки, які вбудована система перевірки орфографії не підкреслює.

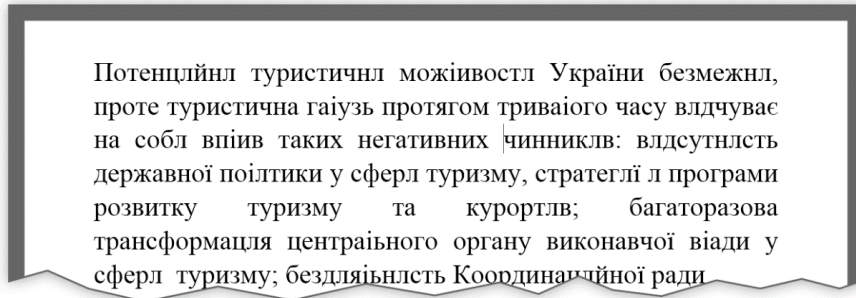


Рис. 6. Текст, надрукований модифікованим шрифтом

Бувають також випадки, коли в модифікованому шрифті видаляють зображення символів. Текст, набраний таким шрифтом, стає невидимим для людини, але розпізнається програмами. Такий спосіб дає змогу приховати посеред заповненого тексту його «сміттєву» частину, що штучно збільшує відсоток оригінальності. Для того щоб виявити такий вид модифікації, варто скористатися способом, описаним у наступному пункті.

5. Вміщення прихованого тексту білим кольором та нульовим кеглем

Ще один спосіб модифікувати текст полягає в тому, що між буквами або між самими словами вставляють символи чи слова, це можна зробити вручну або ж за допомогою спеціального скрипту, який за короткий проміжок часу автоматично вміщує символ/слово, і таким чином більшість антиплагіатних систем не може коректно «прочитати» текст і порівняти з подібним, унаслідок чого підвищується оригінальність. Проте просто вставити символи/слова між літерами замало, адже їх легко побачити, якщо бути уважним. Задля уникнення цього змінюють колір вставленого фрагмента на білий, зменшують розмір до мінімуму й ущільнюють інтервал. Унаслідок такого маскування вставлений фрагмент стає непомітним для ока людини, проте система перевірки збігів сприймає вставлені символи як звичайний текст. Якщо є підозри наявності таких модифікацій, виділяємо текст, збільшуємо інтервал між символами, збільшуємо шрифт і змінюємо колір тексту, після цього уважно переглядаємо. У Microsoft Word можна скористатися комбінацією клавіш Ctrl + пробіл – це налаштує форматування до параметрів за замовчуванням. Така операція дасть змогу побачити вставлені зайві символи (рис. 7, 8).

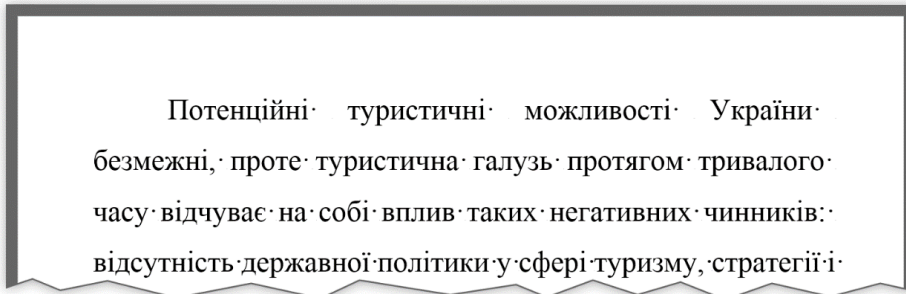


Рис. 7. Модифікований текст зі вставленими символами (початковий вигляд)

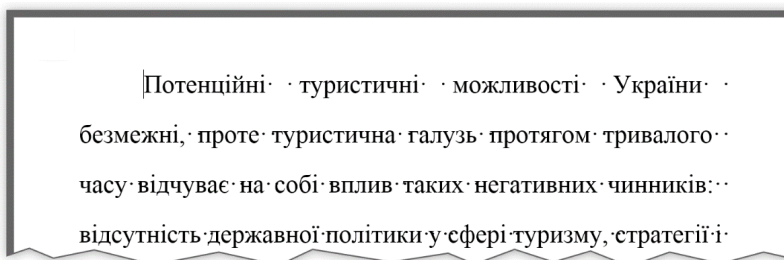


Рис. 8. Модифікований текст зі вставленими символами (після скидання параметрів форматування)

На рисунку наведено текст з увімкненими прихованими символами (крапки позначені пробіли). У нормальному тексті має бути один пробіл, а в модифікованому бачимо два.

На рисунку 9 (з 21 по 43 рядок) зображено витяг з xml-файлу цього текстового документа, на якому між словами видно велику кількість зайвих тегів з характеристиками вміщеного сміттевого форматування, що перешкоджає адекватно прочитати текст і порівняти зі збігами в інтернеті. Особливої уваги заслуговують атрибути тегів `<w:rFonts>`, `<w:spacing>` та `<w:sz>` (з 29 по 31 рядок), які мають нетипові значення. Завдяки їм розуміємо, що вставлений символ використовує шрифт «Gungsuh», міжзнаковий інтервал «-30000» та розмір кегля «2». Користуючись пошуком за кодом символу, визначаємо, що ключовий у цій модифікації символ, який розміщений поміж тегами `<w:t>` на 34 рядку, має назву «Hangul Choseong Filler» (Хангиль чхосон filler U+115F) і являє собою так званий символ-заповнювач, тому немає видимого зображення (його зараховують до фонематичної корейської абетки (Хангиль.)).

Для порівняння на наступному рисунку зображено частину xml-файлу документа без модифікацій (рис. 10), де простежуємо зменшення кількості тегів з характеристиками.


```

17     <w:lang w:val="uk-UA" />
18   </w:rPr>
19   <w:t>Потенційні</w:t>
20 </w:r>
21 <w:r w:rsidR="007E45B9">
22   <w:rPr>
23     <w:lang w:val="uk-UA" />
24   </w:rPr>
25   <w:t xml:space="preserve" />
26 </w:r>
27 <w:r w:rsidR="007E45B9" w:rsidRPr="007E45B9">
28   <w:rPr>
29     <w:rFonts w:ascii="Gungsuh" w:hAnsi="Gungsuh" w:cs="Gungsuh" />
30     <w:spacing w:val="-30000" />
31     <w:sz w:val="2" />
32     <w:lang w:val="uk-UA" />
33   </w:rPr>
34   <w:t> </w:t>
35 </w:r>
36 <w:r w:rsidR="007E45B9" w:rsidRPr="007E45B9">
37   <w:rPr>
38     <w:spacing w:val="-30000" />
39     <w:sz w:val="2" />
40     <w:lang w:val="uk-UA" />
41   </w:rPr>
42   <w:t xml:space="preserve" />
43 </w:r>
44 <w:r w:rsidRPr="00DC6F75">
45   <w:rPr>
46     <w:lang w:val="uk-UA" />
47   </w:rPr>
48   <w:t>туристичні</w:t>
49 </w:r>

```

Рис. 9. Вміст файлу document.xml з наявними модифікаціями

```

13   <w:rPr>
14     <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
15     <w:sz w:val="28"/>
16     <w:szCs w:val="28"/>
17   </w:rPr>
18   <w:t>Потенційні туристичні можливості України безмежні, проте туристична галузь протягом
    тривалого часу відчуває на собі вплив таких негативних чинників: відсутність державної політи
    </w:t>
19 </w:r>

```

Рис. 10. Вміст оригінального файлу document.xml

Ще одним доволі схожим методом модифікації тексту є вміщення прихованого тексту в документ. Це відбувається схожим чином, як у вищенаведеному варіанті, проте вміщують таку кількість «сміттевого» тексту, щоб він за розмірами перевищував видимий текст. У такому разі буде підвищення оригінальності через витіснення неоригінального тексту прихованим. Побічним ефектом є суттєве збільшення розміру файлу, що може привернути увагу того, хто перевіряє цей документ.

Нещодавно розробники Unichesk упровадили функцію виявлення прихованого тексту, тому тепер такі очевидні маніпуляції відображаються у звітному доку-

менті. А також на панелі з'явилася нова вкладка «Модифікації», де зазначаються можливі прояви академічного шахрайства (рис. 11).

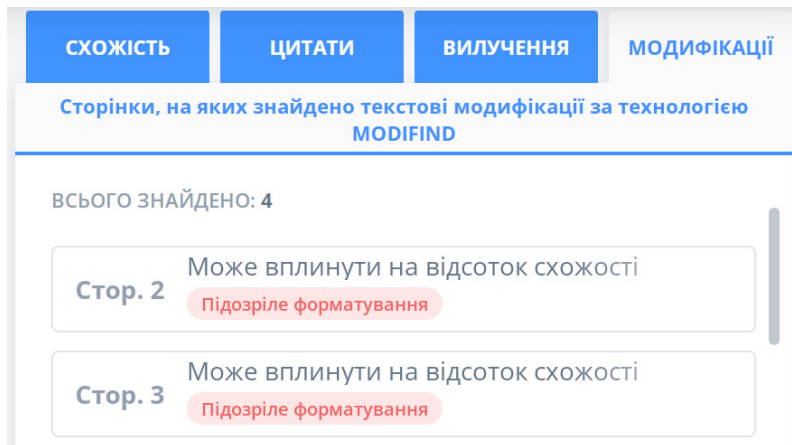


Рис. 11. Система Unicheck попереджає про модифікації тексту

Дослідження результатів штучного підвищення унікальності охоплює три етапи:

1. Початковий перегляд файлу формату docx, де звертали увагу на візуальні відмінності оригінального документа (без технічних модифікацій) і того, який завантажили після обробки на сайті.

2. Тестування наявності збігів через систему перевірки Unicheck. На цьому етапі аналізували можливості програми, спостерігали, чи може система виявити технічне втручання в текстовий файл.

3. Аналіз первинної структури текстового файлу у форматі XML.

Практичну частину роботи становить дослідження послуг сайтів з підвищення унікальності текстів. За основу взято чотири сайти, на яких можна модифікувати фрагмент роботи безплатно.

1. Сайт «Перший український сервіс підвищення унікальності та зменшення плагіату» (<https://antiplagiat.in.ua/>), домен якого зареєстрований у 2014 році, надає послуги з технічного підвищення унікальності текстів. Текстовий файл відправляють на вказану електронну адресу. У листі потрібно вказати назву програми, в якій буде перевірятися робота на збіги, та бажаний відсоток оригінальності. До того ж на сайті зазначено, що система не приймає заявки на оптимізацію файлів, що будуть перевірятися у програмі Unicheck, для якої адміністрація сайту рекомендує рерайт. Цей факт підтверджує надійність системи Unicheck та вбудованих у ній модулів з виявлення академічного шахрайства.

Отже, у процесі дослідження з'ясовано, що компанія надає послугу ознайомлення з модифікованим файлом, частина якого надходить на пошту, до оплати. Прояви шахраювання бачимо відразу (на першому етапі), навіть не вдаючись до глибокого аналізу. У разі ввімкненої функції відображення прихованих символів у тексті бачимо вставки нерозривних пробілів, які розривають слова (рис. 12). Під час завантаження файлу до програми Unicheck текст має нормальний вигляд,

розривів у словах не видно, оригінальність тексту після перевірки 100 %, але в системі підсвічується розділ «Модифікації», де можна виявити розриви у словах, які спричинені нерозривними пробілами.

Як	лікувальне	читання
бібліо°		терапія·
спочатк°		у·
використовувалась·		
°		у· шпитальних·
°		і· санаторних·

Рис. 12. Відображення нерозривних пробілів у Microsoft Word

2. Сайт «Підвищити-антиплагіат.pdf» (<https://xn----8sbaagj5acc1auhmyrdh0npxa.xn--p1ai/>), домен якого зареєстрований у 2018 році. Тут вказано, що підвищення унікальності тексту здійснюється за допомогою нової програмної розробки 2021 року, яка виконує приховану модифікацію коду внутрішніх параметрів файлу, причому текст візуально залишається незмінним, а відсоток оригінальності підвищується до рівня 100 %. Є функція безплатної модифікації до двох сторінок; експорт і завантаження файлу здійснюють на сайті через форму, а модифікований файл готовий до завантаження миттєво.

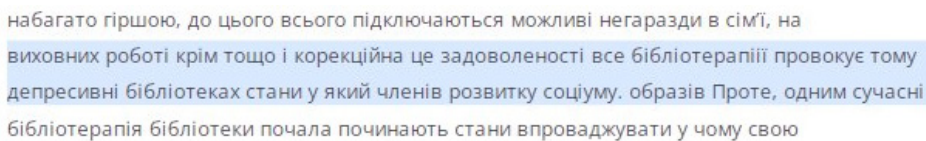
Переглянувши модифікований документ (docx), відразу бачимо, що в середині слів вставлені крапки, вони зменшені у розмірі, до того ж навіть не замасковані білим кольором. Проте погоджуємося з тим, що можна не помітити вставок, якщо переглядати файл нашвидкуруч (рис. 13). Програма Unichesk видала результат з оригінальністю 100 %, але текст якимось чином зменшився в обсязі. Наприклад, з двох сторінок залишився лише абзац, а деякі слова відображаються без пробілів. У вкладці «Модифікації» справді немає ніяких відміток, як і заявлено на сайті. Через те що вміщено приховані крапки, кількість слів залишилася незмінною, а кількість символів збільшилася.

, · а · згодом · з · методами · наукового · кни
 ерапія · зба002Eга002Eчува002Eла002Eсь · методами · і
 вуватись · у · громадських · і · шкільних ·
 вана · педагогам-практиками, завдяки ·

Рис. 13. Вставки крапок, підкреслені рисками; 002E – їх код

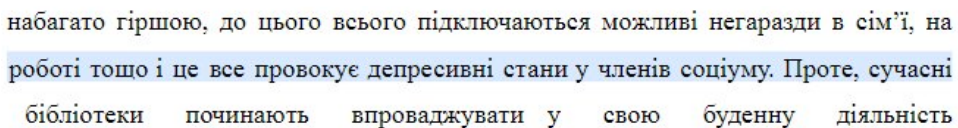
3. Домен сайту «*Анти-антиплагиат.рф*» (<https://xn----7sbbaar5acc1ard1a0-beh.xn--p1ai/>) зареєстрований у 2014 році, хоча вказано, що їхня діяльність з підвищення унікальності розпочалася у 2005 році. Щоб запобігти виявленню модифікованих файлів усіма відомими системами пошуку збігів, в основу роботи сайту закладено метод кодування. Є можливість безплатно спробувати кодування фрагмента тексту (до 8 сторінок). Необхідний текст надсилають на вказану електронну адресу або завантажують через форму на сайті, проте функція замовлення необхідного відсотка оригінальності відсутня. Тестовий модифікований результат надсилають на пошту через 10 хвилин. У листі з пробним модифікованим файлом вказують реквізити для оплати роботи, а також зазначають, що для гарантії компанія надсилає лише половину тексту, тобто тільки 4 сторінки з 8 заявлених, а за умови відсутності оплати впродовж 24 годин файл відправлять до баз безплатних рефератів. Якщо після перегляду тестового фрагмента тексту замовника все влаштовує, то, оплативши замовлення за банківськими реквізитами, можна завантажити всю роботу.

Уже на етапі візуального порівняння двох файлів бачимо відмінності, зокрема у модифікованому тексті перед або після слова наявна вставка літер «є» та «ї», видимість яких спричинена відсутністю цих букв у російській абетці, адже сайт направлений на модифікацію російськомовного тексту. У процесі глибокого аналізу виявлено, що в роботу вмістили додатковий «сміттєвий» текст, а також використали модифікований шрифт, що має назву «Times New Roman», в якому відсутні зображення символів. Під час перевірки модифікованого фрагмента файлу в системі Unichек отримали 0 % схожості з іншими джерелами, але у вкладці «Модифікації» бачимо два пункти «підозрілого форматування», що вказують на ознаки програмного втручання в текст (рис. 14, 15).



набагато гіршою, до цього всього підключаються можливі негаразди в сім'ї, на виховних роботі крім тощо і корекційна це задоволеності все бібліотерапії провокує тому депресивні бібліотеках стани у який членів розвитку соціуму. образів Проте, одним сучасні бібліотерапія бібліотеки почала починають стани впроваджувати у чому свою

Рис. 14. Текст із форматуванням



набагато гіршою, до цього всього підключаються можливі негаразди в сім'ї, на роботі тощо і це все провокує депресивні стани у членів соціуму. Проте, сучасні бібліотеки починають впроваджувати у свою буденну діяльність

Рис. 15. Фрагмент без форматування, де видно вміщений текст

4. Домен сайту «*Антиплагиат кілер*» (<https://killer-antiplagiat.ru/>) зареєстрований у 2018 році, але вказано, що компанія займається підвищенням унікальності текстів з 2006 року. Принцип роботи фірми ідентичний до попередньої, схоже навіть оформлення сайтів. Є можливість безплатно спробувати підвищити оригінальність тексту (до 2 сторінок) без його візуальних змін. Під час замовлен-

ня можна обрати потрібний відсоток оригінальності. Роботи для шифрування відправляють через форму на сайті, звідки приблизно через хвилину можемо завантажити готовий модифікований файл.

У модифікованому документі виявлено вставлені літери «є» та «ї» меншого розміру, а також використання штучного шрифту «Times New Roman». Під час перевірки цього файлу в системі Unichек отримали 0% збігів, але програма звертає увагу на вкладку «Модифікації», де поміж звичайним відображено велику кількість «сміттевого» тексту, наявність якого можемо виявити за допомогою декількох дій, описаних у п'ятому пункті статті – вміщення прихованого тексту білим кольором та нульовим кеглем. Потрібно зауважити, що відбулося збільшення розміру самого файлу (із 17 кБ до 1234 кБ) і кількості слів (з 377 до 678). Варто зазначити, що методи модифікації двох останніх сайтів є ідентичними як на візуальному, так і на програмному рівні, тому можемо припустити, що ці фірми якимось чином пов'язані між собою.

Висновки. У процесі дослідження виявлено три різні методи модифікацій, але всі вони зводяться до одного – технічного втручання у внутрішню структуру документа, а також видозміни шрифтів. Отож найефективнішим і найскладнішим щодо реалізації та виявлення є метод технічного підвищення оригінальності двох останніх сайтів («Анти-антиплагіат.рф», «Антиплагіат кілер»), які схарактеризовано вище.

Підсумовуючи, є підстави твердити, що яскраві заголовки сайтів, які обіцяють підняти відсоток унікальності тексту, не є гарантією того, що подібне шахрайство не помітять під час перевірки. Автори, які постійно звертаються до таких послуг, – завзяті порушники академічної доброчесності. Подібна поведінка є підставою для притягнення до академічної відповідальності здобувачів освіти, що передбачає повторне складання іспитів або заліків, відрахування із закладу, позбавлення стипендії тощо. Науково-педагогічним працівникам у такому разі можуть відмовити в присудженні наукового ступеня / присвоєнні вченого звання, а також позбавити права обіймати визначені законом посади чи брати участь у роботі органів, визначених законом.

Отже, виявлення методів модифікації тексту наукового дослідження – одне із завдань фахівця, який займається перевіркою академічних робіт, а тому зростає потреба щоразу поглиблювати знання у сфері комп'ютерних технологій, розвивати здатність логічно мислити, підвищувати рівень кваліфікації, ознайомлюючись з досвідом колег. Наявність антиплагіатних систем, програмне забезпечення яких постійно оновлюється, спрощує пошук збігів, дає змогу зекономити час та людські ресурси, спрямовані на аналіз надісланого файлу та виявлення ознак порушення академічної доброчесності.

СПИСОК ПОСИЛАНЬ

- Аналізуємо звіт про перевірку на плагіат» – Unicheck та Academic IQ. *Academic Integrity and Quality Initiative – Academic IQ*. [online] Доступно: <<https://www.youtube.com/watch?v=RqGFSYRM6gY>> (матеріали вебінару) [Дата звернення 18 вересня 2021].
- Вугт, В.В., 2007. *OpenXML. Кратко и доступно*. [online] Доступно: <https://mihailromanov.files.wordpress.com/2014/01/oxml_book_ru.pdf> [Дата обращения 07 октября 2021].
- Представление символов, таблицы кодировок. *Викиконспекты*. [online] Доступно: <<https://cutt.ly/ET874bX>> [Дата обращения 28 сентября 2021].
- Способ обхода антиплагиата «со спрятанными словами», 2018. *Авторский проект Александра Павлова*, [online] 20 февраля 2018. Доступно: <<https://apavlov.ru/sposob-obhoda-antiplagiata-so-spryatannymi-slovami/>> [Дата обращения 24 мая 2021].
- Форматы Open XML и расширения имен файлов. *Microsoft*. [online] Доступно: <<https://cutt.ly/dT85rOF>> [Дата обращения 30 сентября 2021].
- Хангиль. *Вікіпедія*. [online] Доступно <<https://cutt.ly/zT85o7w>> [Дата обращения 30 сентября 2021].
- Хангиль чхосон filler U+115F. *Unicode*. [online] Доступно: <<https://unicode-table.com/ru/115F/>> [Дата обращения 30 сентября 2021].
- Чиркин, Е.С., 2013. Использование систем антиплагиата в образовании. *Вестник Томского государственного университета*, 18 (6), с.3380-3387. Доступно: <<https://cyberleninka.ru/article/n/ispolzovanie-sistem-antiplagiata-v-obrazovanii>> [Дата обращения 07 октября 2021].
- Heather, J., 2010. Turnitoff: identifying and fixing a hole in current plagiarism detection software. *Assessment & Evaluation in Higher Education*, [online] 35 (6), pp.647-660. Available at: <<https://doi.org/10.1080/02602938.2010.486471>> [Accessed 09 February 2021].

REFERENCES

- Analizuiemo zvit pro perevirku na plahiat – Unicheck ta Academic IQ. [We analyze the plagiarism check report – Unicheck and Academic IQ]. *Academic Integrity and Quality Initiative – Academic IQ*. [online] Available at: <<https://www.youtube.com/watch?v=RqGFSYRM6gY>> (матеріали вебінару) [Accessed 18 October 2021].
- Chirkin, E.S., 2013. Ispolzovanie sistem antiplagiata v obrazovanii [The use of anti-plagiarism systems in education]. *Vestnik Tomskogo gosudarstvennogo universiteta*, 18 (6), pp.3380-3387. Available at: <<https://cyberleninka.ru/article/n/ispolzovanie-sistem-antiplagiata-v-obrazovanii>> [Accessed 07 October 2021].
- Formaty Open XML i rasshireniia imen failov [Open XML Formats and Filename Extensions]. *Microsoft*. [online] Available at: <<https://cutt.ly/dT85rOF>> [Accessed 30 September 2021].
- Heather, J., 2010. Turnitoff: identifying and fixing a hole in current plagiarism detection software. *Assessment & Evaluation in Higher Education*, [online] 35 (6), pp.647-660. Available at: <<https://doi.org/10.1080/02602938.2010.486471>> [Accessed 09 February 2021].
- Khangil [Hangil]. *Vikipediya*. [online] Available at: <<https://cutt.ly/zT85o7w>> [Accessed 30 September 2021].
- Khangyl chkhoson filler U+115F [Hangul Choson filler U+115F]. *Unicode*. [online] Available at: <<https://unicode-table.com/ru/115F/>> [Accessed 30 September 2021].

Predstavlenie simvolov, tablitsy kodirovok [Representation of characters, tables of encodings]. *Wikiconspect*. [online] Available at: <<https://cutt.ly/ET874bX>> [Accessed 28 September 2021].

Sposob obkhoda antiplagiata «so spriatannymi slovami» [A way to bypass anti-plagiarism “with hidden words”], 2018. *Author’s project by Alexander Pavlov*, [online] 20 February 2018. Available at: <<https://apavlov.ru/sposob-obhoda-antiplagiata-so-spriatannymi-slovami/>> [Accessed 24 May 2021].

Vugt, V.V., 2007. OpenXML. Kratko i dostupno [OpenXML. Brief and accessible]. [online] Available at: <https://mihailromanov.files.wordpress.com/2014/01/oxml_book_ru.pdf> [Accessed 07 October 2021].

UDC 004:001.891:177.9-022.241

Berezhna Oksana,

Master,

E-Document Sector Manager of KNUCA Scientific Library,

Kyiv National University of Culture and Arts,

Kyiv, Ukraine

oberezhna23@gmail.com

<https://orcid.org/0000-0003-4792-3382>

Doroshenko Evhenii,

Network and Service Desk Sector Manager of KNUCA Scientific Library,

Kyiv National University of Culture and Arts,

Kyiv, Ukraine

knukim.journals@gmail.com

<https://orcid.org/0000-0003-1050-5478>

Bohush Iryna,

Master,

Supervisor of the Highest Category of the KNUCA Scientific Library,

Kyiv National University of Culture and Arts,

Kyiv, Ukraine

ira.bogush89@ukr.net

<https://orcid.org/0000-0003-4138-8534>

THINK LIKE A CHEATER OR HOW TO INCREASE THE UNIQUENESS OF THE TEXT

The purpose of the article is to investigate and theoretically justify ways to increase the percentage of the scientific paper uniqueness, defining the basic procedure of text modification based on studying the principles sites work designed to increase the originality of the research material artificially.

Research methodology. The authors have applied general scientific research methods as empirical, analysis, logic, comparison, experiment, and visualisation methods to achieve the purpose.

Scientific novelty. The article first describes the working principle of the following sites ‘The First Ukrainian Service to Increase Uniqueness and Reduce Plagiarism’, ‘Pidvyschyty-

antiplagiat.rf', 'Anti-antiplagiat.rf', 'Anti-plagiat killer', the work's mechanism of which is to increase artificially the originality of the text.

Conclusions. The study has established that eye-catching site titles that promise to increase the percentage of the work uniqueness are not a guarantee that such cheat will not be noticed during verification. Identifying methods for modifying the text of the scientific paper is one of a specialist's tasks who checks academic papers. Therefore, there is an increasing need to deepen knowledge in the field of computer technology every time, to develop the ability to think logically, to improve the level of qualifications, getting acquainted with the experience of colleagues. The presence of anti-plagiarism systems, the software of which is constantly updated, simplifies the search for matches, saves time and human resources aimed at analysing the submitted file and identifying signs of academic integrity violation.

Keywords: plagiarism; text modifications; XML; increasing uniqueness; Unicheck.

УДК 004:001.891:177.9-022.241

Бережная Оксана,

магистр,

заведующая сектором организации электронных документов научной библиотеки,

Киевский национальный университет культуры и искусств,

Киев, Украина

oberezhna23@gmail.com

<https://orcid.org/0000-0003-4792-3382>

Дорошенко Евгений,

заведующий сектором сетевой и технической поддержки научной библиотеки,

Киевский национальный университет культуры и искусств,

Киев, Украина

knukim.journals@gmail.com

<https://orcid.org/0000-0003-1050-5478>

Богущ Ирина,

магистр,

методист высшей категории научной библиотеки,

Киевский национальный университет культуры и искусств,

Киев, Украина

ira.bogush89@ukr.net

<https://orcid.org/0000-0003-4138-8534>

ДУМАЙ КАК МОШЕННИК, ИЛИ КАК ПОВЫСИТЬ УНИКАЛЬНОСТЬ ТЕКСТА

Цель статьи – исследовать и теоретически обосновать способы повышения процента уникальности научных трудов, определив основные методы модификации текста на основе изучения принципов работы сайтов, деятельность которых направлена на искусственное повышение оригинальности представленного материала.

Методы исследования. Для достижения цели использованы такие общенаучные методы исследования, как эмпирический, метод анализа, логический метод, сравнение, эксперимент и метод визуализации.

Научная новизна. Впервые описан принцип работы таких сайтов, как «Первый украинский сервис повышения уникальности и уменьшения плагиата», «Повысить-антиплагиат.рф», «Анти-антиплагиат.рф», «Антиплагиат киллер», механизм действия которых заключается в искусственном повышении оригинальности текста.

Выводы. В процессе исследования установлено, что яркие заголовки сайтов, обещающих поднять процент уникальности работы, не гарантируют, что подобное мошенничество останется незамеченным во время проверки. Выявление методов модификации текста научного исследования – одно из задач специалиста, занимающегося проверкой академических работ, поэтому возрастает потребность каждый раз углублять знания в сфере компьютерных технологий, развивать способность логически мыслить, повышать уровень квалификации, знакомясь с опытом коллег. Наличие антиплагиатных систем, программное обеспечение которых постоянно обновляется, упрощает поиск совпадений, позволяет сэкономить время и человеческие ресурсы, направленные на анализ отправленного файла и выявление признаков нарушения академической честности.

Ключевые слова: плагиат; модификации текста; XML; повышение уникальности; Unicheck.