

ДИРІВ АНДРІЙ

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0009-0615-1161>e-mail: andrii.i.dyriv@lpnu.ua

ЛОЗИНСЬКА ОЛЬГА

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-5079-0544>e-mail: olha.v.lozynska@lpnu.ua

АНАЛІЗ ПРОБЛЕМАТИКИ ПЛАГІАТУ ФОРМУЛ В НАУКОВИХ ПУБЛІКАЦІЯХ ТА ДОСЛІДЖЕННЯ МЕТОДУ ПОРІВНЯННЯ НА ОСНОВІ СХОЖОСТІ ЗМІННИХ

У сучасному науковому середовищі, де математичні формули широко використовуються, виявлення плагіату в наукових публікаціях стає надзвичайно важливим завданням. Хоча багато існуючих систем виявлення плагіату здатні ефективно ідентифікувати копіювання та незначне перефразування тексту, вони часто не в змозі виявити прихований плагіат, особливо у математичних формулах, де акцент робиться на схожості змінних. Це особливо важливо у таких дисциплінах, як математика, фізика, інженерія та інші галузі, де формули і рівняння є невід'ємною частиною досліджень.

Ця публікація пропонує детальний аналіз проблеми плагіату формул, зосереджуючись конкретно на схожості змінних. Авторами досліджено тонкощі структури математичних виразів і того, як змінні в цих виразах можуть бути маніпульовані для маскування плагіату. Досліджуючи різні підходи та методи аналізу плагіату, особливо ті, що зосереджені на виявленні схожості змінних, ми оцінюємо їх ефективність у цьому контексті. Це включає ретельний огляд сучасних технологій та алгоритмів, які використовуються в системах виявлення плагіату, оцінюючи їхні сильні та слабкі сторони при застосуванні до математичного контенту.

Наша дослідницька методологія включає систематичне вивчення широкого спектру наукових публікацій. Ми ідентифікуємо та класифікуємо різноманітні випадки плагіату формул, враховуючи змінні та їхню схожість. Ця класифікація допомагає зрозуміти різні способи, якими плагіат може проявлятися у математичних текстах, такі як пряме копіювання, перефразування назв змінних або більш тонкі зміни, що зберігають основну структуру формули, змінюючи її зовнішній вигляд.

Результати цього дослідження відкривають нові перспективи для подальшого розвитку систем виявлення плагіату в науковому середовищі. Вони підкреслюють необхідність більш досконалих інструментів, які можуть точно аналізувати та порівнювати математичний контент, враховуючи не лише текстові елементи, але й математичну логіку та структуру. Ці вдосконалення можуть бути застосовані для підвищення якості та достовірності наукових публікацій, забезпечуючи належне визнання та оцінку оригінальних досліджень.

Ключові слова: плагіат формул, наукові публікації, схожість змінних, аналіз формул, академічна етика, порівняння змінних, інтелектуальна власність, методи виявлення плагіату, математичний контент

DYRIV ANDRII, LOZYNKA OLGA

Lviv Polytechnic National University

ANALYSIS OF THE PLAGIARISM ISSUES IN SCIENTIFIC PUBLICATIONS AND RESEARCH OF COMPARISON METHOD BASED ON VARIABLE SIMILARITY.

In the contemporary scientific environment, where mathematical formulas are extensively used, detecting plagiarism in scientific publications becomes an exceptionally crucial task. While many existing plagiarism detection systems are capable of efficiently identifying copying and minor text paraphrasing, they often fail to detect concealed plagiarism, particularly in mathematical formulas, where the focus lies on the similarity of variables. This is especially important in disciplines such as mathematics, physics, engineering, and other fields where formulas and equations are integral components of the research.

This publication offers an in-depth analysis of the issue of formula plagiarism, focusing specifically on variable similarity. Authors discovered the intricacies of how mathematical expressions are structured and how variables within these expressions can be manipulated to disguise plagiarism. By examining various approaches and methods for plagiarism analysis, particularly those centered on variable similarity detection, we assess their effectiveness in this context. This involves a thorough review of current technologies and algorithms used in plagiarism detection systems, evaluating their strengths and weaknesses when applied to mathematical content.

Our research methodology includes the systematic examination of a wide range of scientific publications. We identify and classify diverse cases of formula plagiarism, taking into account variables and their similarity. This classification helps in understanding the different ways in which plagiarism can manifest in mathematical texts, such as through direct copying, paraphrasing of variable names, or more subtle alterations that maintain the underlying structure of the formula while changing its appearance.

The findings of this research open up new perspectives for the further development of plagiarism detection systems in the scientific environment. They highlight the need for more sophisticated tools that can accurately analyze and compare mathematical content, taking into consideration not just the textual elements but also the mathematical logic and structure. These advancements can be applied to enhance the quality and credibility of scientific publications, ensuring that original research is properly recognized and credited. Keywords: formula plagiarism, scientific publications, variable similarity, plagiarism detection, formula analysis, academic ethics, variable comparison, intellectual property, plagiarism detection methods, mathematical content

Keywords: plagiarism of formulas, scientific publications, similarity of variables, analysis of formulas, academic ethics, comparison of variables, intellectual property, methods of detecting plagiarism, mathematical content

Вступ

Поняття плагіату в наукових публікаціях – це неправомірне використання чужих ідей, тексту, даних або результатів досліджень без належного визначення джерела. В науковому середовищі це вважається серйозним порушенням етики та норм академічного знання. Проблема плагіату є порушення авторських прав. Плагіат порушує права авторів на їхню інтелектуальну власність. Це не лише порушення етичних норм, але і юридичне правопорушення.

Публікації є основним засобом розповсюдження нових знань та ідей. Плагіат підриває довіру до наукових результатів та авторитету науковців, що може призвести до загального знецінення довіри в науковому середовищі.

Плагіат є загрозою якості наукових досліджень, він впливає на їхню якість, оскільки може призвести до поширення неправдивої або недостовірної інформації, стати перешкодою на шляху розвитку науки та прогресу у різних галузях знань.

Забезпечення рівності умов конкуренції, як от чесність та прозорість у представленні результатів досліджень важливі для забезпечення рівних умов конкуренції між вченими та науковими установами. Плагіат створює нерівність умов та перешкоджає об'єктивному оцінюванню наукових досягнень.

Плагіат впливає на науковий процес наступним чином:

-Втрати довіри до результатів досліджень. Плагіат може призвести до втрати довіри до результатів наукових досліджень та публікацій, що може мати вплив на прийняття рішень у наукових галузях.

-Зниження авторитету наукових видань та установ. Якщо виявляться факти плагіату у наукових виданнях чи відомих наукових установах, це може призвести до зниження авторитету цих видань та установ у науковому співтоваристві.

-Послаблення ефективності наукового рецензування. Плагіат може ускладнити процес наукового рецензування, оскільки рецензентам доведеться більше часу й уваги приділяти перевірці оригінальності публікацій.

Сучасна наукова спільнота стикається зі значним зростанням обсягу наукових публікацій у різних галузях знань. За останні роки кількість наукових робіт і публікацій значно збільшилася, що створює більше можливостей для плагіату та недопущення оригінальності наукових досліджень.

Необхідність збереження етичних стандартів у наукових дослідженнях є вкрай важливим для забезпечення чесності та довіри у науковому середовищі. Плагіат порушує ці стандарти та може поставити під сумнів авторитет наукової спільноти.

Вплив на розвиток науки та технологій, чесність та об'єктивність наукових публікацій є основою для розвитку науки та технологій. Плагіат може спричинити поширення недостовірної або некоректної інформації, що може сповільнити науковий прогрес.

Підвищення якості рецензування, розробка та впровадження нових методів виявлення плагіату може сприяти підвищенню якості процесу наукового рецензування. Це, в свою чергу, допоможе у забезпеченні якості та об'єктивності наукових публікацій.

Вивчення проблеми плагіату та розробка методів його виявлення також мають важливе значення для освітнього процесу. Ефективні методи протидії плагіату можуть допомогти у вихованні етичних стандартів серед молодих дослідників та студентів.

Таким чином, проблема плагіату в наукових публікаціях має серйозний вплив на науковий процес, порушуючи етичні, юридичні та наукові норми, а також загрожуючи якості та об'єктивності наукових досліджень.

Отже, аналіз проблематики плагіату формул в наукових публікаціях та дослідження методу порівняння на основі схожості змінних для наукової спільноти відображає потребу у збереженні етичних стандартів, підтримці чесності та довіри у науковому середовищі, а також важливість для подальшого розвитку науки та освіти.

Постановка проблеми

Мета дослідження: метою цієї роботи є аналіз наявних наукових досліджень у сфері ідентифікації плагіату формул у публікаціях та роботах, розробка методу аналізу плагіату в наукових публікаціях, зокрема за допомогою порівняння формул та визначення схожості змінних між ними. Головною метою є виявлення та усунення випадків плагіату у наукових текстах, що допоможе зберегти чесність та довіру у науковому середовищі.

Формули є ключовим елементом наукових робіт, особливо в областях, де використання математичних концепцій і виразів є стандартом, таких як фізика, математика, інженерія та комп'ютерні науки. Відтак, крадіжка формул може серйозно підірвати довіру до авторів і робити шкоду науковому процесу.

З урахуванням цих факторів, метою цього дослідження є розробка програмного забезпечення, яке здатне ефективно виявляти плагіат у наукових формулах. Це дозволить забезпечити підвищення якостей наукових публікацій та сприятиме відновленню довіри до авторів і наукових джерел.

Завдання дослідження:

а) Розроблення методики аналізу плагіату в формулах наукових публікацій на основі схожості змінних.

б) Проведення аналізу наукових робіт з метою виявлення можливих випадків плагіату в формулах.

в) Розроблення програмного забезпечення для автоматизованого аналізу та виявлення плагіату в формулах наукових публікацій.

г) Визначення ефективності розробленого методу та програмного забезпечення на прикладі реальних наукових даних.

д) Формулювання рекомендацій щодо підвищення ефективності виявлення та запобігання випадків плагіату в наукових публікаціях.

е) Аналіз отриманих результатів та висновки щодо можливостей застосування розробленого методу в практиці та подальших перспектив дослідження у цій області.

Таким чином, ці завдання дослідження спрямовані на розроблення та впровадження методу аналізу плагіату в наукових публікаціях на основі схожості змінних, що сприятиме підтримці етичних стандартів та довіри у науковому середовищі.

Аналіз останніх публікацій та розвитку досліджуваної тематики

Якщо коротко спробувати описати розвиток тематики досліджень проблеми плагіату формул, то це буде виглядати наступним чином:

- Аналіз ранніх досліджень (до 2000-х років):

У період до 2000-х років було обмежено дослідження щодо плагіату в наукових текстах та формулах. Більшість робіт у цей період фокусувалися на загальних аспектах плагіату в літературних та гуманітарних науках.

- Поява програмного забезпечення (2000-2010-і роки):

У 2000-2010-ті роки почав активно розвиватися ринок програмного забезпечення для виявлення плагіату в наукових текстах. Було створено перші онлайн сервіси та програми для перевірки оригінальності текстів.

- Розвиток методів аналізу формул (2010-2020-і роки):

У цей період з'явилися перші дослідження, присвячені аналізу плагіату в наукових формулах. Деякі з них розробляли методи аналізу схожості математичних виразів та змінних. Були запропоновані методи порівняння формул та визначення їхньої унікальності, використовуючи алгоритми комп'ютерного зору та обробки зображень.

- Актуальні дослідження (після 2020 року):

У останні роки велика увага приділяється розробці нових методів виявлення плагіату в наукових формулах з використанням штучного інтелекту та машинного навчання. Дослідники також досліджують можливості використання алгоритмів комп'ютерного зору та обробки зображень для виявлення плагіату у графічних представленнях формул.

З цієї ретроспективи можна зробити наступні висновки:

Розроблено різноманітні методи та програмне забезпечення для виявлення плагіату в наукових текстах. Розвиваються методи аналізу плагіату в наукових формулах, але ця область ще потребує подальшого дослідження та розвитку. Ця ретроспектива вказує на еволюцію досліджень щодо виявлення плагіату в наукових текстах та формулах, а також на актуальні напрямки розвитку цієї області.

Відомі підходи для аналізу плагіату формул у наукових публікаціях:

1. Текстові методи аналізу:

Ці методи базуються на використанні текстового представлення формул. Формули конвертуються у текстовий вигляд, і потім використовуються алгоритми текстового порівняння для визначення схожості.

Переваги:

- Простота реалізації.
- Швидкість обробки текстових даних.

Недоліки:

- Не забезпечують достатньої точності для виявлення схожості між формулами, особливо при складних математичних виразах.
- Не враховують семантичних аспектів формул.

2. Методи порівняння математичних виразів:

Ці методи використовують алгоритми порівняння математичних виразів для аналізу схожості між формулами. Вони враховують математичні властивості та структуру формул.

Переваги:

- Забезпечують більш точну оцінку схожості між формулами, оскільки враховують їхню математичну структуру.
- Можуть виявляти схожість навіть у випадку різних текстових представлень формул.

Недоліки:

- Вимагають складних алгоритмів порівняння та обробки математичних виразів.
- Можуть бути вимогливими до обчислювальних ресурсів.

3. Аналіз змінних та операторів:

Цей підхід базується на аналізі змінних та операторів, що використовуються у формулах. Він порівнює структурні елементи формул, такі як змінні, функції та операції.

Переваги:

- Дозволяє виявляти схожість навіть у випадку різних формул з однаковим математичним значенням.

- Може бути ефективним для виявлення плагіату, коли формули перероблені або змінені місцями.

Недоліки:

- Вимагає складних алгоритмів та великої кількості обчислювальних ресурсів.
- Може бути вразливим до змін у синтаксисі або представленні формул.

Детальніше розглянемо різні методи визначення схожості змінних у формулах:

1. Статистичний аналіз використання змінних:

Цей метод полягає у визначенні статистичних характеристик використання змінних у формулах. Він може включати підрахунок частоти використання кожної змінної, а також визначення спільних змінних між різними формулами.

Переваги:

- Простота в реалізації.
- Можливість виявлення частоти використання та спільних змінних.

Недоліки:

- Не враховує семантичний зміст змінних та їхніх взаємозв'язків у формулах.
- Не забезпечує загальну оцінку схожості між формулами.

2. Аналіз структури формул:

Цей підхід визначає схожість змінних у формулах, враховуючи їхню структуру та взаємозв'язки. Методи можуть використовувати алгоритми порівняння дерев або графів, щоб визначити структурні схожості між формулами.

Переваги:

- Дозволяє виявляти складні структурні схожості між формулами.
- Враховує структурні властивості змінних та їхні взаємозв'язки.

Недоліки:

- Вимагає складних алгоритмів обробки та порівняння структур дерев або графів.
- Може бути вимогливим до обчислювальних ресурсів.

3. Аналіз контексту використання змінних:

Цей підхід враховує контекст використання змінних у формулах, такі як їхні функціональні призначення та взаємозв'язки з іншими змінними. Методи можуть використовувати високорівневі алгоритми обробки природної мови або машинного навчання для аналізу контексту.

Переваги:

- Дозволяє виявляти схожість між змінними з урахуванням їхнього призначення та контексту використання.
- Може бути ефективним для виявлення плагіату, коли формули переформульовані або змінені.

Недоліки:

- Вимагає великої кількості навчальних даних та обчислювальних ресурсів для реалізації алгоритмів машинного навчання.
- Може бути вразливим до невірного аналізу контексту та неправильностей в даних.

Кожен з цих методів має свої переваги та недоліки, і вибір конкретного методу залежить від конкретних вимог дослідження та обмежень ресурсів.

Для того щоб глибше зрозуміти проблематику питання було досліджено різні актуальні роботи з цієї тематики.

У науковій роботі [1] розглянуто актуальні проблеми плагіату, особливо в контексті математичних наук, де традиційні системи виявлення плагіату часто не впораються через унікальні виклики, які виникають внаслідок формальної математичної нотації. Автори роблять два значущі внески. По-перше, вони встановлюють таксономію повторного використання математичного контенту шляхом анотування значної кількості пар наукових документів, підозрюваних у плагіаті. По-друге, вони аналізують ефективність різних підходів до виявлення плагіату та схожості математичного контенту на основі цієї новоствореної таксономії.

Результати показують, що існуючі методи виявлення плагіату та схожості математичного контенту мають проблеми з ефективністю, з найкращими методами, які досягають відносно низьких загальних показників виявлення. Зокрема, показники виявлення (PlagDet) для плагіату складають 0,06, а для схожості математичного контенту - 0,16. Найкращі методи не вдалося виявити більшість випадків у всіх семи встановлених типах схожості математичного контенту.

Описані внески цього дослідження мають значний вплив на галузь систем виявлення плагіату, систем рекомендацій, систем питань та відповідей та пошукових систем. Представлення коду експерименту та анотованого набору даних для спільноти надає автори важливі ресурси для подальших досліджень і розвитку в цій області.

У статті [2] досліджено потреби в ідентифікації академічного плагіату, що є значущою проблемою для освітніх установ, видавців та фінансових установ. Хоча поточні системи виявлення плагіату ефективні у виявленні копійованого та незначно переформульованого тексту, вони часто мають проблеми з прихованим плагіатом, таким як сильні перефразування, переклади та використання нетекстового контенту та ідей.

Автори продовжують свої попередні дослідження щодо аналізу математичного контенту та академічних цитат, обидва з яких вказують на перспективи у покращенні виявлення прихованого

академічного плагіату, особливо в галузі науки, технологій, інженерії та математики (STEM). Вони роблять кілька внесків:

- пропонують двоетапний процес виявлення, який поєднує оцінки схожості математичного контенту, академічних цитат та тексту.
- вводять нові міри схожості, які враховують порядок математичних ознак, які перевершують попередні міри.
- порівнюють ефективність підходів на основі математики, цитат та тексту, використовуючи підтвержені випадки академічного плагіату.
- демонструють, що комбінований аналіз математичних та цитатних ознак дає змогу виявляти потенційно підозрілі випадки у великій колекції документів у галузі STEM.

У цілому, дослідження показує, що аналіз схожості математичного контенту та академічних цитат доповнює традиційні підходи до виявлення плагіату за текстом в академічній літературі в галузі STEM. Автори надають свої дані та код відкрито доступними, що надає цінні ресурси для подальших досліджень у цій області.

Авторами [3] описано проблеми оцінки індексу подібності математичних та інших наукових публікацій, що містять рівняння та формули. Присутність рівнянь та формул значно ускладнює вивчення таких текстів. Досліджено можливості найпопулярнішого антиплагіатного програмного забезпечення, системи iThenticate, яка на даний момент використовується в наукових журналах, для виявлення плагіату та самоплагіату. Представлені результати обробки цією системою конкретних тестових завдань, що містять багато рівнянь та формул. Виявлено, що система iThenticate часто значно переоцінює індекс подібності і, отже, не може встановити відмінність між самоплагіатом та псевдосамоплагіатом (фальшивим самоплагіатом). Ця стаття буде корисною для дослідників та викладачів університетів у галузі математики, фізики та інженерних наук, програмістів-розробників програмного забезпечення, а також широкого кола читачів, які цікавляться питаннями плагіату та самоплагіату.

У науковій роботі [4] досліджено складність аналізу та розпізнавання математичних формул, що містяться в текстових документах. Для виявлення часткових дублікатів потрібно не лише аналізувати графічне зображення, проводячи фільтрацію, виділення контурів і використовуючи специфічні методи порівняння, але і текстову інтерпретацію формули, щоб мати можливість ідентифікувати частковий дублікат, за умови, що у формулі було змінено позначення літер, символи математичних операцій, форми дужок тощо. Тому для виявлення часткових дублікатів математичних формул пропонується гібридний підхід, що передбачає використання шаблонів, створених відповідно до особливостей графічних редакторів, та спеціальних конверторів формул з різних форматів до канонічного формату.

Формулювання цілей статті

Розглянувши розвиток теми оцінки формульного плагіату, виявлено, що це надзвичайно складне і мінливе питання, яке вимагає глибокого та комплексного дослідження. Щоб зрозуміти всі його аспекти, необхідно розглянути кожен модуль окремо та вивчити його взаємодію з іншими, враховуючи контекст та особливості конкретного дослідження.

Одним з найбільш фундаментальних аспектів є аналіз схожості змінних у формулах. Виявлення подібності між змінними може бути ключовим показником плагіату, тому саме на цьому аспекті ми зосереджуємо нашу увагу у даній роботі. Але це лише перший крок у дослідженні цієї проблеми.

В подальшому, планується розширити дослідження, аналізуючи інші аспекти формульного плагіату. Наприклад, порівняння структури формул, використання математичних операторів, оцінка співвідношення між числовими та символічними змінними, а також врахування контексту та зв'язків між різними частинами формул. Це допоможе нам отримати більш повне розуміння плагіату у формулах та розробити більш ефективні методи його виявлення.

Загалом, кожен аспект нашої роботи відкриває нові можливості для подальшого дослідження та розвитку систем виявлення плагіату. Наші знання та методики стануть фундаментом для подальшого пошуку шляхів запобігання плагіату та забезпечення чесності та надійності наукових досліджень у майбутньому.

Системний аналіз

Здійснимо простий аналіз системи спроектувавши діаграму варіантів використання (рис. 1), для опису внутрішніх станів у системі, а також діаграму послідовності, для представлення основного сценарію роботи в системі.

На діаграмі зазначено варіанти використання, які існують в системі, вони допомагають нам зрозуміти, які актори і в якій мірі взаємодіють із системою, та які стани взаємодіють між собою в системі.

Існує кілька ключових елементів, що використовуються у діаграмі варіантів використання:

- **Актори:** Представляють користувачів системи або інші системи, які взаємодіють з основною системою, на діаграмі зображено двох, це загальний Actor, що уособлює будь-якого користувача, та БД – база даних.
- **Варіанти використання (Use Cases):** Представляють конкретні можливості використання системи або функціональність, яку може викликати актор, всі ці варіанти відповідно знаходяться в системі.
- **Взаємодії (Associations):** Відображають, як актори взаємодіють з варіантами використання.
- **Включення (Include):** Показує, що один варіант використання включає в себе інший.

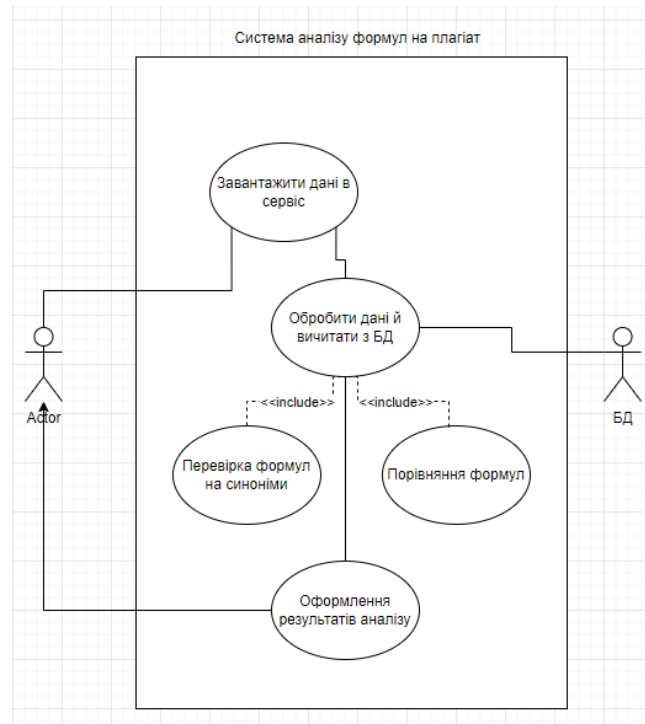


Рис. 1. Діаграма варіантів використання

Відповідно на представленій діаграмі варіантів використання бачимо, як саме користувач вносить дані, та яким чином відбувається процес їхньої обробки та видачі результату назад користувачеві.

Тепер розглянемо діаграму послідовності (рис. 2)

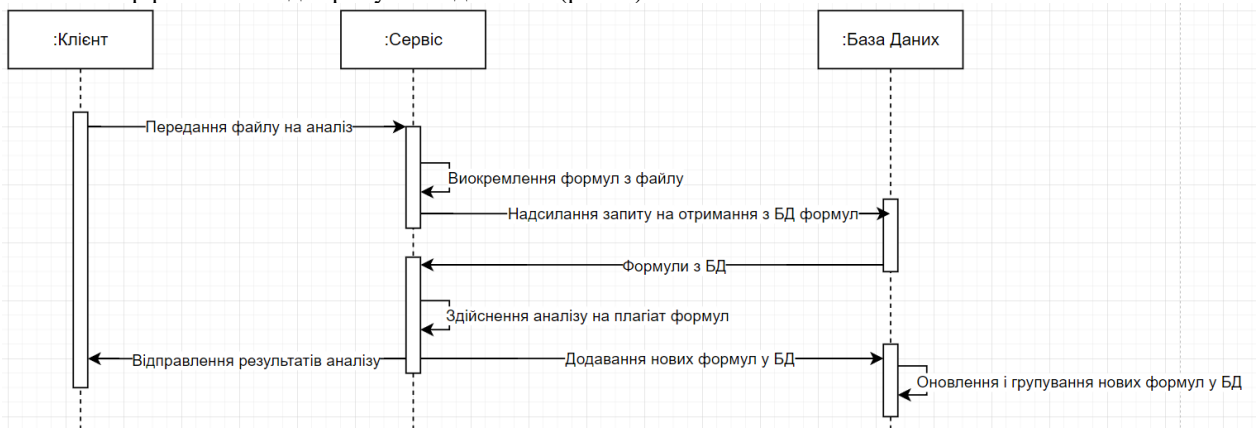


Рис. 2. Діаграма послідовності

Діаграма послідовності зображає, як саме в хронологічному порядку очікується виконання операцій в основному сценарії використання системи. Для розроблюваної системи послідовність виглядає так: спочатку клієнт передає у сервіс файл, який в собі містить формули, які потрібно перевірити на предмет плагіату, або схожості з іншими формулами, які вже містяться у базі даних. Далі відбувається процес виокремлення формул, для цього потрібно буде впровадити алгоритм виявлення і виокремлення формул із загального тексту наданого файлу. Після опрацювання сервіс подає запит у базу даних для того, щоб отримати формули, які там містяться і з якими нам потрібно порівняти формули надалі у файлі клієнта. Цю ситуацію можна покращити додавши додатковий алгоритм, щоб не витягувати всі формули, а тільки ті, які потенційно можуть бути схожими. Після того як інформацію про формули надійде у сервіс, почнеться обробка основного алгоритму, який має порівняти формули користувача і з бази даних і скласти прогнозоване значення плагіату, після чого нові дані про формули занесуть у базу даних, а результати обробки алгоритму будуть представлені клієнту.

Виклад основного матеріалу

Проблема плагіату в наукових публікаціях є серйозним завданням, що стає все більш актуальним у сучасному науковому середовищі. Плагіат може виявлятися не лише в копіюванні текстового матеріалу, але й у використанні формул та математичних виразів, що робить його виявлення ще складнішим. У зв'язку з цим, розроблення програмного забезпечення для ефективного виявлення плагіату в наукових формулах є важливою задачею.

Для початку визначимо за допомогою чого будемо проводити реалізацію. Мова Python є дуже зручною для реалізації різних алгоритмів, вона багато різних модулів і бібліотек, які ми зможемо в майбутньому використати щоб розширити наше рішення. А для початкового рішення ця мова програмування також дуже добре підходить за рахунок того, що вона є не типізованою, а в нас не використовуватимуться складні класи то ми зможемо отримати дуже компактне рішення з дуже хорошою читабельністю коду, що в майбутньому допоможе легко інтегрувати наші напрацювання і в сторонні системи.

Для зручності оцінки того як працює алгоритм будемо спостерігати як він працює із фізичними формулами, адже там дуже широко застосовуються математичні вирази і зручно на них пояснювати роботу алгоритму.

Звісно перевірку плагіату потрібно застосовувати на комплексних формулах, які представляють нові рішення, адже саме для цього і потрібна перевірка, але враховуючи, що алгоритм, який ми реалізували, це лише частина великого модуля, то його відпрацювання краще спостерігати на простіших випадках.

На рис. 3 подано представлення формул для обрахунку переміщення і для обрахунку прискорення тіла. В двох цих випадках ключова змінна обраховується через математичні операції над змінними швидкості та часу. І хоч у відсотковому складі ми будемо бачити високу повторюваність змінних, очевидно що найважливішу роль тут відіграє саме знак операції. Цим прикладом ми показуємо, що хоч метод порівняння змінних у формулі є базовим, але без інших методів порівняння він не може використовуватись для серйозного аналізу.

```

Enter the first form: S=v*t
Enter the second form: a=Δv/t

Common variables in both formulas:
Variable: v, Count: 1
Variable: t, Count: 1

Unique variables in the first formula:
Variable: S, Count: 1
Variable: v, Count: 0
Variable: t, Count: 0

Unique variables in the second formula:
Variable: a, Count: 1
Variable: Δ, Count: 1
Variable: v, Count: 0
Variable: t, Count: 0

=== Code Execution Successful ===

```

Рис. 3. Формули для обрахунку переміщення і прискорення

Тепер розглянемо інший випадок. Припустимо хтось винайшов якусь унікальну формулу для якогось процесу, в нашому випадку нехай це буде формула обрахунку координати вісі ординат для тіла кинутого під кутом до горизонту. І, припустимо, з'явилося «нове» рішення даної задачі, представлене іншою формулою (рис. 4). Очевидно, що друга формула це просто переписана перша з перестановкою доданків і з вилученою частиною, де була змінна яка зберігала початкову координату тіла, тобто, за замовчуванням, розглядається це значення рівне 0. У даному випадку ми бачимо практично відсутню унікальність у формулах і велике число повторів змінних, що при аналізі розділових знаків покаже нам, що ці формули є однакового сенсового навантаження.

Є ще важливий аспект який стосується тематики аналізу змінних, а саме – це їхня заміна. Для цього розглянемо випадок з формулою яка розраховує роботу для переміщення тіла (рис.5). Бачимо що формули майже ідентичні, але не на 100%, хоча насправді змінні l і S – це змінні що під собою мають на увазі одну й ту ж величини – переміщення або ж шлях тіла. Так переміщення і шлях це не завжди одне й те ж саме, але у таких задачах ці величини майже все однакові. Отже, іншим важливим фактором може стати заміна змінної просто іншою змінною хоча і з тим самим контекстом для розуміння.

Для цієї формули було представлено приклад коли поміняно лише одну змінну, але так можна зробити і з всіма змінними, просто використати іншу букву для їх представлення. Тому важливим є також контекст кожної змінної у формулі. Програма, яка б могла хоча б на примітивному рівні розрізняти такі контексти, є доволі складною і на даному етапі дослідження її реалізація не вважається доцільною.

```

Enter the first form: y=g*t*t/2+x*t
Enter the second form: y=y+x*t+g*t*t/2

Common variables in both formulas:
Variable: t, Count: 3
Variable: x, Count: 1
Variable: g, Count: 1
Variable: y, Count: 1

Unique variables in the first formula:
Variable: y, Count: 0
Variable: g, Count: 0
Variable: t, Count: 0
Variable: x, Count: 0

Unique variables in the second formula:
Variable: y, Count: 1
Variable: x, Count: 0
Variable: t, Count: 0
Variable: g, Count: 0

=== Code Execution Successful ===

```

Рис. 4. Формули для обрахунку координати тіла кинутого під кутом до горизонту

Для цього ж прикладу, аналіз математичних знаків спокійно показав, що тут можливо наявний плагіат, адже дерево, яке ми можемо побудувати для математичних операцій, буде мати однаковий вигляд вершин, але це також питання яке варто розглянути як окрему детальну працю.

```

Enter the first form: A=F*l*cos(α)
Enter the second form: A=F*S*cos(α)

Common variables in both formulas:
Variable: α, Count: 1
Variable: o, Count: 1
Variable: s, Count: 1
Variable: c, Count: 1
Variable: A, Count: 1
Variable: F, Count: 1

Unique variables in the first formula:
Variable: A, Count: 0
Variable: F, Count: 0
Variable: l, Count: 1
Variable: c, Count: 0
Variable: o, Count: 0
Variable: s, Count: 0
Variable: α, Count: 0

Unique variables in the second formula:
Variable: A, Count: 0
Variable: F, Count: 0
Variable: S, Count: 1
Variable: c, Count: 0
Variable: o, Count: 0
Variable: s, Count: 0
Variable: α, Count: 0

=== Code Execution Successful ===

```

Рис. 5. Формули для обрахунку роботи по переміщенню тіла

Отже, розроблений алгоритм працює так, як це і було задумано, наступним кроком має стати впровадження і поєднання з модулем, який відповідає за математичні операції, даний варіант вже буде прийнятним для роботи.

Висновки

Перш за все дана робота є тільки маленькою частиною складного комплексу, який варто використовувати для оцінювання наукових публікацій на предмет формульного плагіату, вона закладає основи методики аналізу, дослідження якого допоможе спрямувати подальший розвиток у цій сфері.

У цій роботі ми провели аналіз плагіату формул у наукових публікаціях з використанням підходу, який зосереджується на схожості змінних. Наші дослідження показують, що виявлення плагіату у математичних формулах може бути складним завданням через неоднорідність формулювань і труднощі в інтерпретації символів.

Проведений аналіз також підтверджує ефективність підходу, що базується на порівнянні змінних, у виявленні плагіату. Ми виявили, що цей метод може ідентифікувати певні форми плагіату.

Результати досліджень свідчать про те, що підхід, спрямований на аналіз схожості змінних у формулах, може бути корисним доповненням до існуючих систем виявлення плагіату. Враховуючи значення математичних формул у наукових дослідженнях, розвиток ефективних методів виявлення плагіату в цих формулах є важливим кроком у підвищенні якості та достовірності наукових публікацій.

Загалом, наша робота надає цінний внесок у сферу виявлення плагіату у наукових публікаціях, а її результати можуть бути використані для подальших досліджень та розвитку систем виявлення плагіату в науковій сфері.

У дослідженні ми зосередилися на виявленні плагіату у наукових публікаціях через аналіз схожості змінних у математичних формулах.

Завдяки проведеному дослідженню, ми досягли поставлених цілей, виявивши ефективність методу аналізу схожості змінних у формулах для виявлення плагіату. Це відкриває нові перспективи для подальших досліджень у сфері виявлення плагіату в наукових публікаціях та дозволяє покращити якість та достовірність наукових досліджень.

У роботі зроблено кілька ключових висновків, які мають велике значення для наукової спільноти.

По-перше, ми показали, що аналіз схожості змінних у математичних формулах є ефективним методом для виявлення плагіату.

По-друге, наші результати підтверджують важливість уваги до математичних аспектів при аналізі наукових публікацій. Виявлення плагіату у формулах може виявитися критичним для забезпечення інтелектуальної чесності та високої якості наукових досліджень.

Нарешті, наші висновки підкреслюють необхідність подальших досліджень у цій області. Шляхом вдосконалення методів виявлення плагіату в математичних формулах ми можемо сприяти створенню більш надійних та довірених наукових джерел.

У цілому, наші основні висновки демонструють важливість розвитку методів виявлення плагіату в математичних формулах для забезпечення чесності та якості наукових досліджень.

Література

1. Ankit Satpute, André Greiner-Petter, Noah Gießing, Isabel Beckenbach, Moritz Schubotz, Olaf Teschke, Akiko Aizawa & Bela Gipp. Taxonomy of Mathematical Plagiarism (2024). https://link.springer.com/chapter/10.1007/978-3-031-56066-8_2

2. Norman Meuschke; Vincent Stange; Moritz Schubotz; Michael Kramer; Bela Gipp. Improving Academic Plagiarism Detection for STEM Documents by Analyzing Mathematical Content and Citations (2017). <https://ieeexplore.ieee.org/abstract/document/8791126/metrics#metrics>

3. Andrei D. Polyaniin & Inna K. Shingareva. The Similarity Index of Scientific Publications with Mathematical Equations and Formulas (2022). <https://link.springer.com/article/10.1007/s12109-022-09869-2#Abs1>

4. Petro Lizunov, Andrii Biloschchytskyi, Larysa Chala, Svitlana Biloschchytska, Oleksandr Kuchanskyi, Sergii Udovenko. Hybrid approach to the analysis and recognition of mathematical formulas for detecting similarities (2017). <https://dSPACE.uzhnu.edu.ua/jspui/handle/lib/19203>