

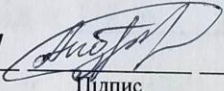
КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

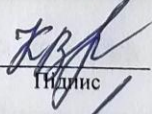
на тему Метод адаптивної фільтрації наукових статей для
персоніфікованих рекомендацій засобами машинного навчання

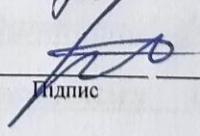
Галузь знань 12 – Інформаційні технології
Шифр і назва галузі знань

Спеціальність 122 – Комп'ютерні науки
Шифр і назва спеціальності

Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студент групи КНс-22-1  Дмитро АНДРУЩЕНКО
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Керівник: асистент каф. КН  Валерія КЛІМЕНКО
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор

Підпис

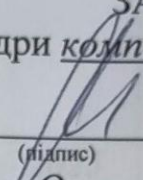
Олександр БАРМАК

Ім'я, ПРІЗВИЩЕ

18 06 2025 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
Факультет інформаційних технологій
Кафедра комп'ютерних наук
Освітній ступінь бакалавр
Галузь знань 12 – Інформаційні технології
Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри комп'ютерних наук

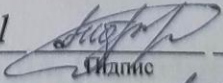

(підпис)
д.т.н., професор Олександр БАРМАК
« 10 » 02 2025 року

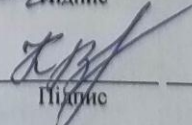
ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: «Метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання»
2. Завдання видано студенту Дмитро Андрущенко
(Ім'я, прізвище)
3. Керівник роботи асистент каф. КН Валерія КЛІМЕНКО
(посада, ім'я, прізвище)
4. Затверджено наказом університету від « 07 » 02 2025 р. № 23
5. Дата видачі завдання студенту: « 10 » 02 2025 р.
6. Зміст пояснювальної записки (перелік задач) та вихідні дані:
Мета роботи – покращення процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання. Для досягнення мети слід виконати такі задачі: виконати аналіз інформаційних моделей області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій; створити метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання; виконати прикладну програмну реалізацію інтелектуальної системи та провести дослідження ефективності розробленого методу з використанням розробленої програмної реалізації.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напрямку дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником, складання календарного графіка виконання	січень 2025	Виконано
2	Ознайомлення з предметною областю, формулювання мети і задач дослідження, визначення об'єкта та предмета дослідження	лютий 2025	Виконано
3	Проектування та розроблення методу вирішення завдання, загальної архітектури програмного забезпечення, інтерфейсу користувача, вибір засобів реалізації програмного забезпечення	березень 2025	Виконано
4	Створення та тестування програмного забезпечення, дослідження ефективності, висновки з виконаної роботи	квітень 2025	Виконано
5	Написання пояснювальної записки, урахування зауважень керівника, оформлення згідно з вимогами	травень 2025	Виконано
6	Розробка презентаційних матеріалів та попередній захист кваліфікаційної роботи	травень 2025	Виконано
7	Отримання відгуку керівника, рецензії, перевірка на плагіат, нормоконтроль	червень 2025	Виконано
8	Підготовка до захисту та захист кваліфікаційної роботи	червень 2025	Виконано

Виконавець: студент групи КНс-22-1  Дмитро АНДРУЩЕНКО
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Керівник: асистент каф. КН  Валерія КЛІМЕНКО
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: «Метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання»

Виконавець кваліфікаційної роботи бакалавра: студент групи КНс-22-1 Дмитро Андрущенко

Керівник кваліфікаційної роботи бакалавра: асистент каф. КН Валерія Кліменко

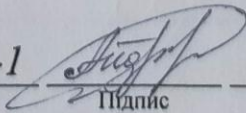
Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
50	18	2	43	5

Метою кваліфікаційної роботи бакалавра є покращення процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання. У процесі розробки методу адаптивної фільтрації було використано мову програмування Python, а також засоби машинного навчання. Розроблений метод виконує аналіз наукових статей на основі запиту користувача для формування персоніфікованих рекомендацій. Метод імплементовано у вигляді веб-застосунку, що забезпечує зручний доступ до функціоналу системи для кінцевих користувачів. Розроблений веб-застосунок призначений для використання дослідниками, науковцями та освітніми платформами.

Напрямами практичного використання розробленої системи є покращення якості наукових пошуків, підвищення релевантності рекомендованих матеріалів.

Ключові слова: адаптивна фільтрація, релевантність, all-MiniLM-L6-v2.

Виконавець: студент групи КНс-22-1  Дмитро АНДРУЩЕНКО
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень	4
Вступ.....	5
Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій.....	7
1.1 Аналіз інформаційних моделей в області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій	7
1.2 Засоби автоматизованої адаптивної фільтрації наукових статей	10
1.3 Аналіз існуючих програмних рішень з адаптивної фільтрації наукових статей для персоніфікованих рекомендацій.....	13
1.4 Аналіз наукових досліджень в області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій	18
1.5 Мета та задачі кваліфікаційної роботи бакалавра	19
Розділ 2 Метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання.....	20
2.1 Схема та кроки методу адаптивної фільтрації наукових статей засобами машинного навчання.....	20
2.2 Проектна архітектура та взаємозв'язок компонентів інтелектуальної системи адаптивної фільтрації наукових статей.....	22
2.3 Підготовка датасету для адаптивної фільтрації наукових статей.....	25
2.4 Нейромережева архітектура для отримання векторних представлень.....	27
2.5 Особливості використання програмних розширень.....	30
2.6 Висновки до розділу 2	31
Розділ 3 Експериментальне дослідження методу	33
3.1 Методика дослідження	33
3.2 Програмна структура інтелектуальної системи.....	34
3.3 Вибір засобів розробки інтелектуальної системи адаптивної фільтрації наукових статей	36

3.4 Особливості реалізації програмних складових інтелектуальної системи адаптивної фільтрації наукових статей.....	38
3.5 Аналіз функціональності створеної інтелектуальної системи адаптивної фільтрації наукових статей	40
3.6 Результати досліджень	43
3.7 Висновки до розділу 3	45
Загальні висновки.....	46
Перелік посилань.....	47
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
БД	База даних
КРБ	Кваліфікаційна робота бакалавра
BERT	Bidirectional Encoder Representations from Transformers
GloVe	Global Vectors for Word Representation
Doc2Vec	Document to Vector
TF-IDF	Term Frequency – Inverse Document Frequency
ВБД	Векторна база даних
DAIN	Deep Adaptive Interest Network
FAISS	Facebook AI Similarity Search
CLS	Classification
SEP	Separator
JSON	JavaScript Object Notation
СКБД	Система керування базами даних
API	Application Programming Interface
AI	Artificial Intelligence
СКБД	Система керування базами даних
CI/CD	Continuous Integration / Continuous Delivery

Вступ

Кваліфікаційна робота бакалавра присвячена покращення процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання.

Актуальність. Актуальність дослідження, присвяченого методам адаптивної фільтрації наукових статей для персоніфікованих рекомендацій, зумовлена як безпрецедентним зростанням обсягів наукової інформації, так і зміною способів її споживання в умовах цифровізації наукового простору. У сучасному інформаційному середовищі науковці стикаються з проблемою інформаційного перевантаження, коли кількість доступних публікацій значно перевищує можливості їх осмислення та аналізу. Ця ситуація створює суттєві виклики для формування релевантного знання, ускладнюючи процес виявлення цінних і змістовних джерел, необхідних для проведення досліджень.

У відповідь на ці виклики зростає потреба у таких інструментах, які здатні не лише автоматизувати процес доступу до наукової інформації, а й робити його максимально індивідуалізованим, ураховуючи унікальний профіль інтересів і дослідницьких пріоритетів кожного користувача. У цьому контексті адаптивні системи фільтрації, побудовані на базі методів машинного навчання, демонструють високий потенціал для розв'язання означеної проблеми. Їхня здатність виявляти приховані закономірності у поведінці користувачів, аналізувати семантичний зміст текстів і динамічно оновлювати модель переваг дозволяє створювати рекомендаційні механізми, що забезпечують точне і своєчасне надання релевантних матеріалів.

Дослідження у сфері адаптивної фільтрації наукових статей є не лише відповіддю на потреби сучасної наукової спільноти, але й важливим внеском у розвиток інтелектуальних інформаційних систем, здатних підтримувати ефективну наукову комунікацію, сприяти формуванню міждисциплінарних зв'язків та пришвидшувати генерацію нових знань.

Таким чином, вдосконалення алгоритмів персоніфікованої рекомендації наукових статей має суттєве практичне значення для підвищення ефективності наукової діяльності та доступності знань у сучасному інформаційному середовищі.

Об'єкт дослідження – процес адаптивної фільтрації наукових статей для персоніфікованих рекомендацій.

Предмет дослідження – методи та технології машинного навчання адаптивної фільтрації наукових статей для персоніфікованих рекомендацій.

Мета кваліфікаційної роботи бакалавра – покращення процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання.

Завдання кваліфікаційної роботи бакалавра – виконати аналіз інформаційних моделей області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій; створити метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання; виконати прикладну програмну реалізацію інтелектуальної системи та провести дослідження ефективності розробленого методу з використанням розробленої програмної реалізації.

Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій

1.1 Аналіз інформаційних моделей в області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій

У сучасному світі наукова інформація стає деталі більш доступною, водночас її кількість стрімко зростає. Щоденно в наукових журналах, базах даних, електронних журналах, бібліотеках та репозиторіях публікуються тисячі нових статей. Відповідно до цього це створює проблему інформаційного перевантаження. Дослідники витрачають чимало зусиль на пошук релевантної інформації, що стає складним та потребує значного часу. У зв'язку з цим зростає потреба в інтелектуальних системах, здатних автоматизовано аналізувати, фільтрувати та рекомендувати науковий контент залежно від інтересів користувача.

Адаптивна фільтрація наукових статей – це процес автоматичного відбору публікацій, які найкраще відповідають інтересам або поточним дослідницьким потребам користувача [1]. Також використовується в системах персоналізованих рекомендацій, що враховують інтереси, уподобання, історію переглядів та інші фактори щоб надати користувачеві найбільш релевантні матеріали.

Публікація наукової статті у фаховому виданні засвідчує рівень наукової зрілості автора, дозволяє йому долучитися до професійної спільноти, а також сприяє загальному розвитку науки [2]. Незалежно від напряму чи то технічні науки, гуманітарні або прикладні дослідження. Наукова стаття залишається універсальною формою представлення результатів інтелектуальної діяльності.

Наукова стаття – це один із основних жанрів наукової комунікації, який використовується для представлення результатів досліджень, аналітичної роботи або наукових розвідок [3]. Її мета – це донесення до наукової спільноти нових знань, обґрунтування певних гіпотез, узагальнення результатів експериментів або огляд актуальних наукових проблем.

Назва наукової статті є першим елементом, з яким зустрічається читач, тому вона повинна бути максимально точною, інформативною та лаконічною [4]. Саме назва дозволяє зрозуміти основний зміст роботи та визначати, чи буде стаття корисною для конкретного дослідника.

Наукові статті можуть містити різні типи назв [5]:

- тематичні, що відображають тему дослідження;
- проблемні назви акцентують увагу на проблемі;
- питальні назви формулюють дослідницьке питання;
- комбіновані, що об'єднують кілька підходів.

Автори в науковій статті – це особи, які безпосередньо брали участь у створенні наукової публікації, провели дослідження, обробили результати, сформулювали висновки та оформили текст статті відповідно до наукових вимог.

Автором вважаються той, хто зробив суттєвий інтелектуальний внесок хоча б один із етапів роботи:

- формулювання ідеї дослідження;
- проведення експериментів або аналізу;
- обробка даних;
- написання тексту або редагування;
- наукове керівництво.

Ключові слова у науковій статті – це основні терміни або стійкі словосполучення, які стисло відображають зміст наукової публікації. Вони визначають головні поняття, що розглядаються у статті, і допомагають орієнтуватися в тематиці наукової статті [6].

Ключові слова використовуються для індексації та пошуку статті в електронних БД. Зазвичай зазначаються після анотації у кількості від чотирьох до восьми одиниць у називному відмінку та перелічуються через кому.

Ключові слова допомагають швидко зрозуміти зміст статті, а також полегшити пошук публікації в наукових БД. Вони відображають найважливіші поняття, що розкривають зміст дослідження.

Анотація – це короткий, стислий виклад змісту наукової статті, що містить основну інформацію про дослідження [7]. Анотація надає читачеві уявлення про наукову новизну та значущість публікації без необхідності читати повний текст наукової статті. Її обсяг становить 100-250 слів. Анотація виконує важливу функцію в популяризації наукової статті та є невід’ємним елементом академічного тексту.

Список використаних джерел є невід’ємною частиною наукової статті, що містить певні описи джерел, які були використані під час написання наукової роботи [8]. Джерела відіграють важливу роль у забезпеченні наукової достовірності дослідження та полегшує зв’язок з іншими науковими роботами в БД.

Як підсумок, в умовах стрімкого зростання обсягів наукової інформації, стаття є основним засобом представлення результатів досліджень. Також має чітко визначену структуру, яка складається з наступних елементів:

- назва;
- відомості про авторів;
- анотація;
- ключові слова;
- основний текст;
- список використаних джерел.

Така побудова сприяє ефективному і логічному викладенню наукового матеріалу.

Отже, в рамках дослідження буде автоматизовано адаптивний пошук наукових статей за такими атрибутами: назва, анотація та комбінація назви з анотацією. Така автоматизація дозволить підвищити релевантність результатів, забезпечити персоніфіковане подання наукового контенту відповідно до індивідуальних запитів користувача.

1.2 Засоби автоматизованої адаптивної фільтрації наукових статей

Збільшення обсягів наукової інформації призвело до необхідності створення інтелектуальних систем фільтрації, здатних автоматично адаптуватися до потреб конкретного користувача. Традиційні методи фільтрації, засновані на ключових словах або ручному сортуванні, виявляються недостатньо ефективними. У зв'язку з цим виникає потреба в автоматизованих системах адаптивної фільтрації, які здатні враховувати індивідуальні інтереси користувачів та контекст наукових статей.

Основу таких систем складають методи машинного навчання, зокрема штучні нейронні мережі та векторні бази даних.

Штучні нейронні мережі є потужним інструментом для обробки та аналізу текстової інформації [9]. Вони забезпечують можливість автоматичного виявлення закономірностей у великих обсягах даних без явного програмування.

У контексті фільтрації наукових публікацій особливо важливими є методи обробки природної мови, класифікацій, кластеризації та рекомендаційних моделей. Фільтрація реалізується, як задача класифікації або регресії, де алгоритм навчається на основі попередніх уподобань користувача. Рекомендаційна система може оцінювати релевантність нової статті до профілю користувача та адаптувати свої результати в режимі реального часу.

Особливо ефективними в задачах обробки природної мови є трансформерні моделі, такі як ALBERT, BERT, RoBERTa та їх спеціалізовані версії для наукових текстів [10]. Ці моделі здатні формувати глибокі семантичні представлення тексту, враховуючи контекст та взаємозв'язки між словами в реченнях, що критично важливо для поточного розуміння наукового контенту.

Моделі, спеціально натреновані на наукових корпусах демонструють кращі результати при роботі з вузькопрофільною термінологією. Вони використовуються для задач тематичної класифікації, витягування сутностей, побудови резюме, а також генерації векторних подань документів.

Нейроні моделі виконують векторизацію тексту, перетворюючи кожен статтю на вектор фіксованого розміру, що зберігає її семантичний зміст. На основі цих векторів здійснюється пошук подібності, формуються рекомендації та оновлюються профілі користувачів.

Векторне представлення тексту є ключовим етапом у розробці системи адаптивної фільтрації. Завдяки цьому процесу текстові дані перетворюються на числовий формат, що дозволяє алгоритмам машинного навчання ефективно аналізувати, класифікувати та фільтрувати інформацію відповідно до індивідуальних уподобань користувача [11].

Найпоширеніші методи векторного представлення тексту варіюються за складністю, точністю та здатністю відображати семантичний зміст. Вибір конкретного підходу залежить від завдань, які потрібно вирішити, а також від характеру тексту, що обробляється.

Серед найбільш вживаних методів можна виділити такі: TF-IDF, Word2Vec та GloVe, Doc2Vec, Sentence-BERT, SciBERT та BioBERT.

Метод TF-IDF [12] є класичним статистичним методом, який оцінює важливість слова в конкретному документі в порівнянні з усією колекцією документів. Він базується на частоті появи терміну, однак не враховує значення слів у контексті. Метод TF-IDF є простим та інтерпретованим, його обмеження полягають у слабкій здатності відображати семантичні зв'язки між словами.

Важливе місце займають у побудові векторних представлень Word2Vec та GloVe [13]. Ці моделі будують векторні представлення окремих слів, враховуючи контексти, в яких ці слова з'являються. Зазвичай для представлення всього документа використовують середнє значення векторів слів або інші методи об'єднання. Вони краще передають семантичну інформацію, ніж TF-IDF.

Подальшим розвитком Word2Vec стала модель Doc2Vec [14], яка дає змогу створювати вектори не тільки для слів, а ще й для цілих документів. Такий підхід є корисним для порівняння або класифікації великих тестових об'єктів, із збереженням їх змістовного контексту.

Окрему нішу займає Sentence-BERT [15], що є модифікацією моделі BERT. Вона спеціально розроблена для порівняння речень або коротких тестових фрагментів. Модель генерує високоякісні векторні представлення речень, що дозволяє ефективно оцінювати їхню схожість або виконувати класифікацію на основі змісту.

Серед моделей, орієнтованих на роботу з науковими текстами, варто виділити SciBERT та BioBERT [16]. Ці трансформерні моделі, спеціально натреновані на наукових текстах. Вони забезпечують вищу точність при роботі з термінами та виразами, характерними для вузькоспеціалізованих галузей таких як медицина, біологія та технічні науки.

Завдяки таким підходам можна обрати найбільш релевантний інструмент для побудови ефективного методу адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання.

Створення ВБД знань, ґрунтуючись на зазначених підходах, відкриває нові перспективи для впровадження активної фільтрації [17]. Кожна наукова стаття може бути представлена як багатовимірний вектор у спільному векторному просторі. У цьому просторі пошук схожих наукових статей перетворюється на задачу обчислення відстані або подібності між векторами, що суттєво підвищує швидкість і точність пошуку.

Основними перевагами застосування ВБД у задачах адаптивної фільтрації полягає [18]:

- маштабованість, що дає можливість ефективну роботу з великими масивами наукових публікацій без втрати якості результатів;
- швидкий пошук, який використовується для індексації векторів що дозволяє реалізувати пошук найближчих сусідів з високою продуктивністю;
- семантична релевантність, що вказує на відміну від класичних ключових слів на векторні представлення, що дозволяють виявляти наукові статті, які є релевантними за змістом, навіть якщо терміни в них різні;

– можливість персоналізації, що дає змогу подавати у векторному вигляді профілі інтересів користувача, а також будувати персоніфіковані рекомендації шляхом порівняння з базою наукових статей.

Отже, використання векторних баз у процесі адаптивної фільтрації наукових статей для персоніфікованих рекомендацій із застосуванням методів машинного навчання сприяє підвищенню ефективності інформаційного пошуку, за рахунок покращення семантичної релевантності, оперативності обробки даних та здатності до персоналізації результатів. Завдяки властивості векторних подань відображати глибинні смислові залежності між текстовими об'єктами, формується концептуальна основа для побудови рекомендаційних систем, які здатні адаптуватися до індивідуальних інформаційних потреб користувача в умовах динамічного наукового середовища.

1.3 Аналіз існуючих програмних рішень з адаптивної фільтрації наукових статей для персоніфікованих рекомендацій

Сучасні технології машинного навчання дедалі активніше інтегруються у наукове середовище, зокрема у процес пошуку та аналізу публікації. Адаптивні системи фільтрації, побудовані на основі алгоритмів машинного навчання, відіграють ключову роль у створенні рекомендацій, орієнтованих на індивідуальні потреби користувачів.

Яскравим прикладом застосування таких систем є наукові платформи Google Scholar [19], Mendeley [20] та ResearchGate [21], які реалізують різні підходи до персоналізованих рекомендацій. Їх порівняльний аналіз дає змогу зрозуміти, як саме алгоритмічні рішення впливають на якість інформаційної підтримки дослідницької діяльності, а також виявити сильні та слабкі сторони кожного підходу.

Google Scholar безкоштовна пошукова система, створена компанією Google, яка спеціалізується на індексації пошукового тексту або метаданих наукових публікацій з різних галузей знань. Вперше була запущена у 2004 році, і

відтоді стала одним із найпопулярніших інструментів для пошуку академічної літератури серед дослідників, студентів, викладачів і науковців у всьому світі (рисунок 1.1).

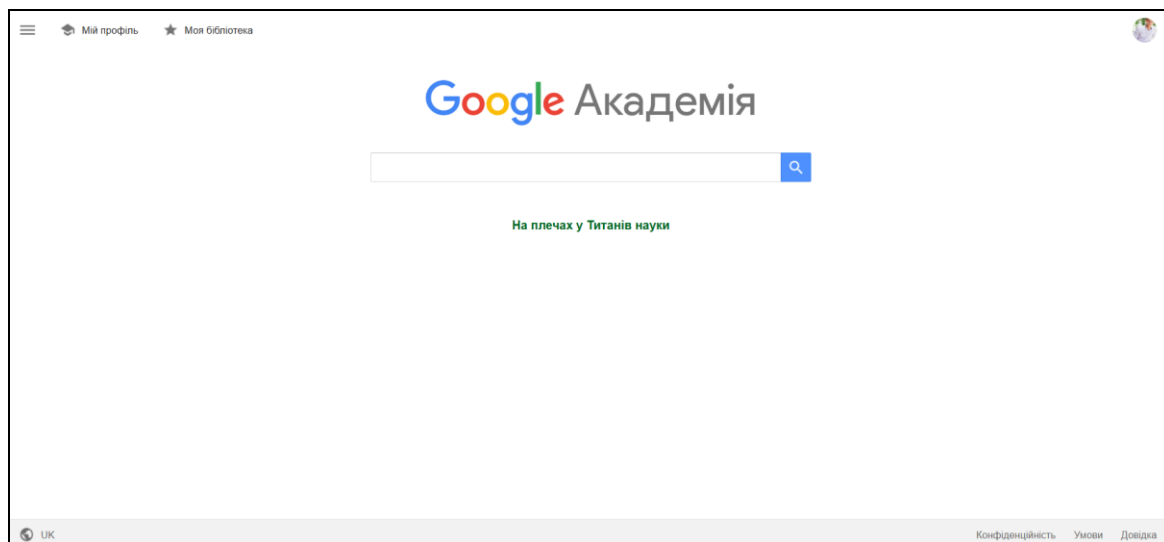


Рисунок 1.1 – Головна пошукова сторінка Google Scholar [19]

Сервіс охоплює широкий спектр джерел, включаючи рецензовані журнали, дисертації, книги та монографії, тези конференцій, передруки, патенти та цифрування інших робіт [22]. Головна перевага – це масштабованість і міждисциплінарність охоплення, що дозволяє знайти матеріали практично з будь-якої теми (рисунок 1.2).

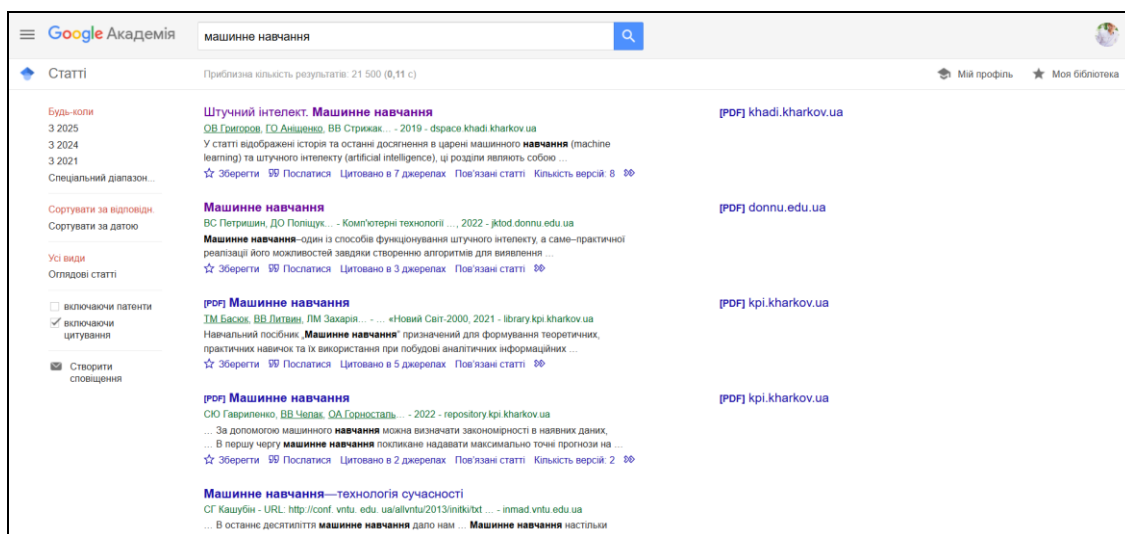


Рисунок 1.2 – Результати запити в Google Scholar [19]

Можна виділити наступні функціональні можливості [23]:

- результати пошуку ранжуються за складними критеріями, що враховують повний текст, кількість цитувань та дату публікації;
- веб-застосунок, автоматично генерує бібліографічні посилання у різних стилях;
- користувачі можуть створювати особисті академічні профілі з автоматичним відстеженням цитувань;
- можливість підписатися на сповіщення про нові публікації за ключовими словами, а також авторами;

Google Scholar не має повноцінної адаптивної системи рекомендацій на основі штучного інтелекту у відкритому доступі, як це реалізовано у Mendeley чи у ResearchGate. Однак веб-застосунок Google Scholar використовує власні алгоритми для покращення релевантності результатів, беручи до уваги історію пошукових запитів, особисті уподобання, а також дані з Google Scholar Profile.

Серед подібних рішень виділяється також потужний менеджер посилань Mendeley. Він призначений для допомоги дослідникам у організації їхніх наукових матеріалів, встановленні співпраці з колегами та пошуку актуальних публікацій [24] (рисунок 1.3).

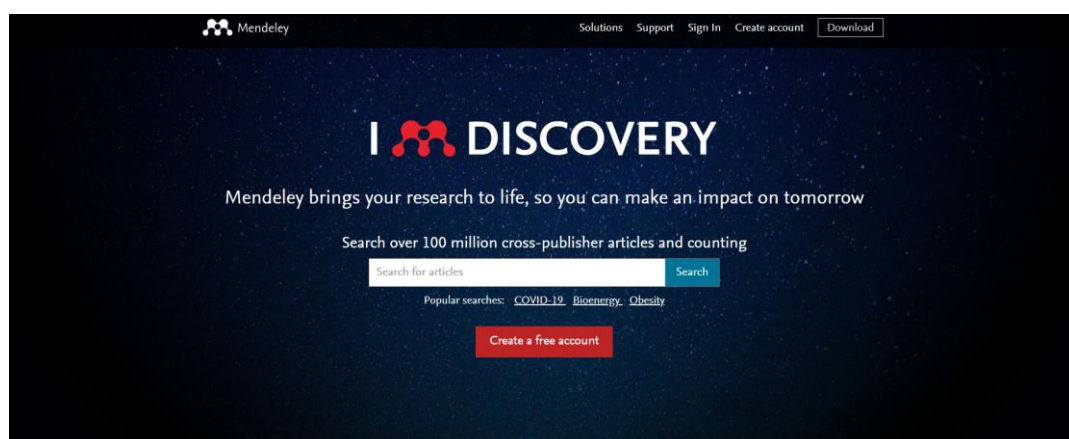


Рисунок 1.3 – Головна сторінка Mendeley [20]

Однією з ключових функцій Mendeley є персоналізовані рекомендації на основі вмісту бібліотеки користувача та раніше прочитаних статей (рисунок 1.4).

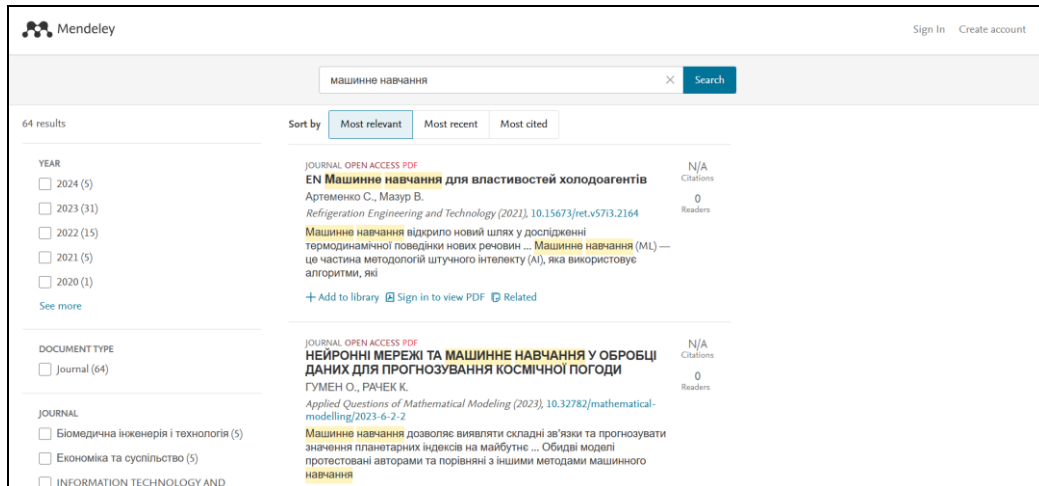


Рисунок 1.4 – Користування Mendeley [20]

Також сервіс надає можливість для спільної роботи, а це означає що користувачі можуть створювати групи, обмінюватися статтями та працювати над науковими проектами разом, незалежно від місця перебування.

Наступним прикладом є ResearchGate, що представляє онлайн-платформу та об'єднує вчених і дослідників для обміну науковими публікаціями, а також використовується для обговорення професійних питань і розвитку співпраці [25] (рисунок 1.5).

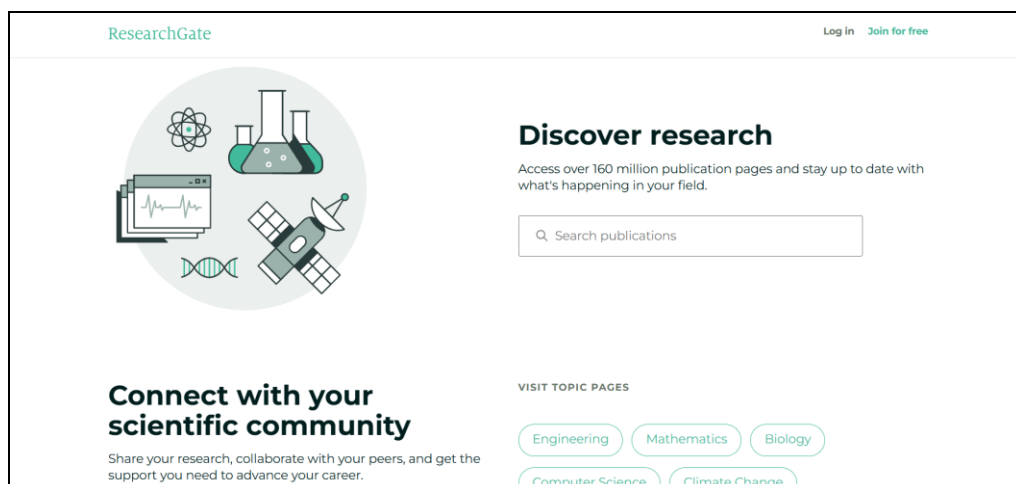


Рисунок 1.5 – Онлайн платформа ResearchGate [21]

Однією з ключових функцій є персоналізована стрічка новин, яка формується на основі профілю користувача та його публікацій і активності, що

дозволяє отримувати лише релевантні дослідження. Також платформа надає аналітичні інструменти для відстеження переглядів, завантажень і цитувань власних робіт, що допомагає краще оцінити їх вплив.

Інтерфейс пошуку наукових матеріалів за ключовими словами, зображений на рисунку 1.6, ілюструє зручність доступу до релевантних публікацій.

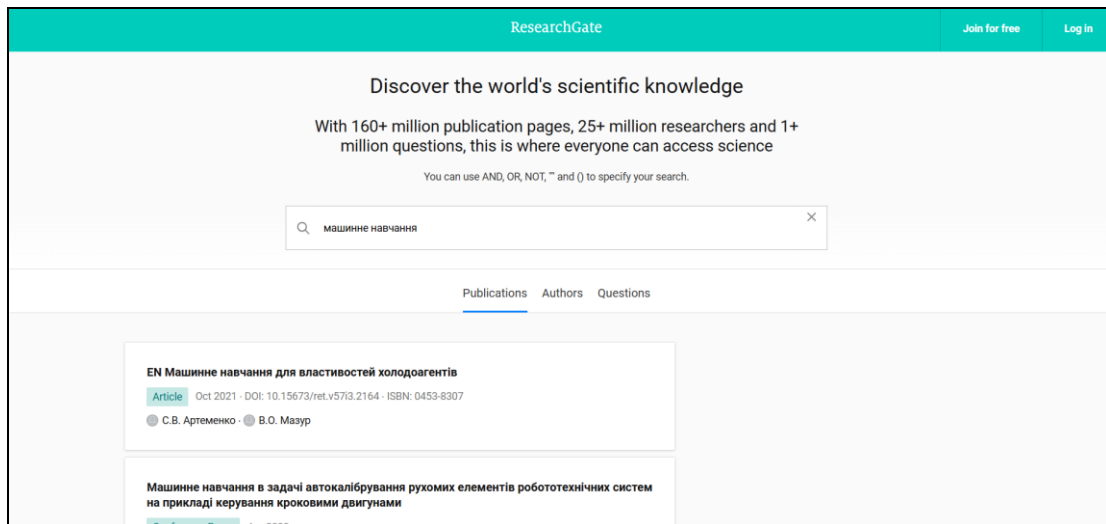


Рисунок 1.6 – Пошук наукових публікацій на платформі ResearchGate за запитом «машинне навчання»

Також ResearchGate сприяє науковій співпраці, дозволяючи користувачам ставити запитання спільноті та отримувати відповіді від фахівців у відповідній галузі.

Отже, такі платформи, як Google Scholar, Mendeley та ResearchGate, успішно впроваджують адаптивні методи фільтрації для надання персоналізованих рекомендацій наукових статей. Завдяки використанню машинного навчання ці системи можуть аналізувати поведінку користувачів і пропонувати відповідні матеріали, що значно спрощує дослідникам доступ до необхідної інформації. Проте, незважаючи на досягнуті успіхи, існує потреба в подальшому вдосконаленні цих систем для забезпечення більш точних і контекстуально релевантних рекомендацій.

1.4 Аналіз наукових досліджень в області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій

Проблема адаптивної фільтрації наукових статей для персоналізованих рекомендацій привертає широку увагу вчених. Зі стрімким зростанням кількості наукових публікацій дослідники стикаються з труднощами у пошуку релевантної інформації, яка відповідає їхнім поточним запитам, науковим інтересам.

У дослідженні [26] описується модель DAIN, яка зосереджується на динамічному моделюванні інтересів користувача. Ця модель інтегрує контекстну обробку з глибокими нейронними мережами, що дозволяє ефективно адаптувати рекомендації відповідно до актуальних потреб. Результати експериментів на публічних наборах даних демонструють значне покращення якості персоналізованих рекомендацій у порівнянні з традиційними підходами.

Наступне рішення описано в дослідженні [27], що організовує елементи в семантично узгоджені кластери та використовує механізм дослідження для підвищення різноманітності рекомендацій. Система трансформує елементи у векторну форму та кластеризує їх відповідно до семантичної схожості через онлайн-алгоритм з адаптивним порогом. Механізм, яким може керувати користувач, вибірково показує елементи з менш вивчених груп, завдяки чому рекомендації стають більш різноманітними. Експерименти на наборі даних MovieLens [28] показують, що включення дослідження зменшує схожість елементів у списку з 0.34 до 0.26 та збільшує несподіваність до 0.73. Це дає кращі результати, ніж традиційні методи колаборативної фільтрації та рекомендацій за популярністю.

У роботі [29] дослідники зосередилися на створенні інтелектуальної системи рекомендацій на основі механізмів самоуваги. Вони підкреслюють важливість глибокого аналізу поведінки користувача, зокрема взаємодії з попередніми рекомендаціями. Запропонована модель продемонструвала високі показники точності до 85,2% та F1-міри 84,3%, що свідчить про її перспективність для впровадження в наукових системах.

Окремо варто відзначити дослідження [30], яке розглядає адаптивний алгоритм колаборативної фільтрації з урахуванням згасання інтересу з часом. Такий підхід дає змогу ефективно вирішувати проблему «холодного старту» та адаптувати рекомендації відповідно до змін у науковій діяльності користувача.

Також у роботі [31] розглянуто гібридну систему, що об'єднує переваги контентного та колаборативного підходів. Основна увага зосереджена на розробці моделі, яка враховує багатовимірні характеристики статей, зокрема метадані, семантичні вбудовування та історію переглядів.

Отже, сучасні дослідження в сфері адаптивної фільтрації наукових статей підтверджують ефективність моделей машинного навчання у створенні персоналізованих систем рекомендацій. Запропоновані рішення відрізняються гнучкістю, здатністю адаптуватися до змін у поведінці користувачів та високою точністю. Це створює міцну основу для розробки інтелектуальних інструментів та наукового пошуку.

1.5 Мета та задачі кваліфікаційної роботи бакалавра

Метою кваліфікаційної роботи бакалавра є покращення процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання.

Для досягнення мети, ставляться до виконання такі задачі:

- виконати аналіз інформаційних моделей області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій;
- створити метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання;
- виконати прикладну програмну реалізацію інтелектуальної системи адаптивної фільтрації наукових статей для персоніфікованих рекомендацій;
- виконати дослідження ефективності розробленого методу з використанням прикладної програмної реалізації.

Розділ 2 Метод адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання

2.1 Схема та кроки методу адаптивної фільтрації наукових статей засобами машинного навчання

Метод адаптивної фільтрації наукових статей засобами машинного навчання призначений для автоматизованого відбору релевантних наукових публікацій відповідно до індивідуальних інформаційних потреб користувача. Метод сприяє покращенню процесу адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання. Схема та кроки методу наведено на рисунку 2.1.

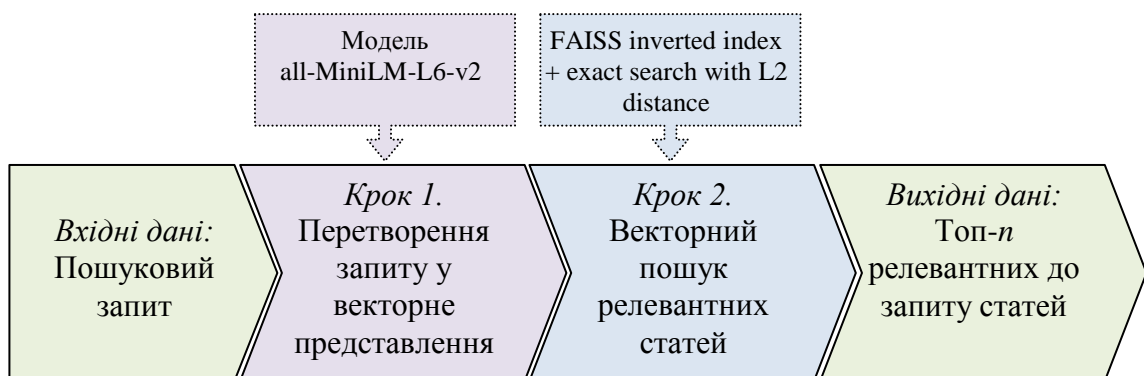


Рисунок 2.1 – Схема та кроки методу адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання

Вхідними даними методу є пошуковий запит користувача, який формується на основі його інформаційних потреб.

Пошуковий запит є агрегованим та складається з:

- авторів;
- ключових слів;
- назви наукової публікації;
- анотації до публікацій;
- поєднання назви та анотації.

На першому кроці здійснюється перетворення запиту у векторне представлення за допомогою моделі «all-MiniLM-L6-v2», яка є результатом попереднього тренування з використанням знань більших трансформерних архітектур у межах підходу «knowledge distillation», що дозволяє зберегти семантичну інформативність при значно меншій кількості параметрів. Вибір цієї моделі обумовлений її здатністю формувати щільні векторні представлення зі збереженням контекстної семантики, що є критично важливим у задачах інтелектуального пошуку та кластеризації наукових публікацій.

На другому кроці здійснюється векторний пошук релевантних наукових статей. Для цього використовується бібліотека FAISS, яка дозволяє ефективно виконувати пошук найближчих сусідів «nearest neighbors» у великих векторних просторах.

У реалізованому методі застосовується поєднання інвертованого індексу з точною метрикою «L2-distance», що забезпечує баланс між швидкістю пошуку та точністю результатів. Пошук виконується у попередньо згенерованому векторному індексі, який містить векторні представлення наукових публікацій з бази даних. На цьому етапі система знаходить ті документи, яких векторне представлення є найближчим до векторизованого запиту користувача, що дає змогу виявити публікації зі схожою семантикою.

Вихідними даними є топ-*n* найбільш релевантних до запиту наукових статей, упорядкованих за ступенем схожості. Результати можуть бути використані для подальшого аналізу, побудови рекомендацій або персоналізованої фільтрації наукової інформації.

Такий підхід дозволяє автоматизувати процес інформаційного пошуку, зменшуючи навантаження на користувача та підвищуючи ефективність доступу до релевантних знань.

Отже, метод адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання дозволяє на основі проведеного аналізу наукових публікацій формувати персоналізовані рекомендації для користувачів. Цей ефект досягається шляхом перетворення вхідних даних у

вигляді текстових метаданих і вмісту наукових статей через етапи попередньої обробки тексту, векторизації за допомогою алгоритмів машинного навчання. Покращення процесу адаптивної фільтрації наукових статей полягає у підвищенні релевантності та персоналізації рекомендацій завдяки застосуванню моделі all-MiniLM-L6-v2 для створення семантично точних векторних представлень тексту, що враховують контекст і змістовну схожість статей.

2.2 Проектна архітектура та взаємозв'язок компонентів інтелектуальної системи адаптивної фільтрації наукових статей

Проектна архітектура інтелектуальної системи адаптивної фільтрації наукових статей подана на рисунку 2.2.



Рисунок 2.2 – Архітектура інтелектуальної системи адаптивної фільтрації наукових статей

Архітектура системи базується на модульному підході, де кожен компонент виконує власну функціональну роль, але всі вони взаємопов'язані.

Центральним елементом архітектури є база даних, яка служить сховищем для усіх текстових представлень наукових статей та їх вектори. Особливістю розробленої системи є використання векторного представлення текстів замість традиційного підходу, що базується на зберіганні сирого тексту.

У класичних системах пошуку кожен запит вимагає повторного опрацювання текстових даних, що призводить до високого навантаження на систему та обмежує можливості глибинного аналізу змісту. Натомість у запропонованій архітектурі тексти перетворюються у вектори та математичні об'єкти, які зберігають основні семантичні характеристики текстового матеріалу у вигляді координат у багатовимірному просторі.

Завдяки такому підходу система може порівнювати не самі тексти, а їхні векторні образи, що значно пришвидшує процес пошуку та дозволяє оцінювати змістовну подібність між науковими статтями. Це означає, що користувач отримує персоналізовані рекомендації не лише за збігом конкретних слів чи фраз, а й за глибинною схожістю ідей, тематики та контексту різних наукових статей.

Одним із перших етапів функціонування системи є робота підсистеми збору та управління даними. Її основна роль полягає в організації регулярного отримання нових наукових статей із різноманітних джерел, таких як відкриті наукові репозитарії, електронні бібліотеки та спеціалізовані наукові платформи. Функціонал підсистеми включає кілька важливих операцій.

Перш за все, вона забезпечує можливість додавання нових статей у базу даних. Додавання передбачає автоматичну обробку тексту статті. Крім того, підсистема реалізує функцію видалення статей із бази даних. Якщо певні записи втрачають актуальність, містять помилки або не відповідають вимогам релевантності, вони можуть бути видалені для підтримання цілісності та якості інформаційного середовища системи.

На наступному етапі кожна наукова стаття обробляється попередньо навченою моделлю перетворення тексту у числові вектори, в результаті чого текст конвертується у багатовимірний вектор, що компактно відображає його зміст та є придатним для подальшої обробки й аналізу.

Отримані вектори зберігаються у базі даних, структура якої оптимізована для забезпечення швидкого пошуку за принципом векторної близькості. Кожен вектор супроводжується відповідними метаданими такими як унікальний ідентифікатор статті, її назва, імена авторів, дата публікації та коротка анотація. Наявність метаданих дозволяє ефективно ідентифікувати результати після виконання семантичного пошуку.

Одночасно з процесами збору та обробки тексту працює підсистема, яка відповідає за навчання моделі машинного навчання. Підсистема створює модель, здатну ефективно прогнозувати, наскільки нові статті відповідають запиту користувача. Спочатку модель навчається на основі текстів, які були попередньо оброблені і представлені у вигляді векторів, а також на прикладах, що показують, які статті були релевантними для користувача.

Система виконує пошук по векторах, знаходячи найближчі за семантичною відстанню вектори. Для цього використовуються метрики відстані, зокрема косинусна подібність, евклідова відстань або більш складні диференційовані метрики.

Результатом роботи підсистеми фільтрації є формування персоналізованого списку статей, ранжованого за ступенем релевантності. Користувач отримує не тільки безпосередні рекомендації, а й можливість переглянути пов'язану інформацію. Наприклад, інші статті на ту саму тему або роботи, написані тим самим автором.

Отже, сформовано проєктну архітектуру інтелектуальної системи, що складається із трьох взаємопов'язаних підсистем та бази даних. Після збору даних матеріали потрапляють до підсистеми машинного навчання. Тут відбувається перетворення тексту наукової статті на вектори. Після цього вектори зберігаються у БД.

2.3 Підготовка датасету для адаптивної фільтрації наукових статей

У контексті розробки систем персоніфікованих рекомендацій наукових статей, ключовим етапом є підготовка якісного та структурованого датасету. Для реалізації методу адаптивної фільтрації було обрано датасет Research Papers Dataset [32] (рисунок 2.3).

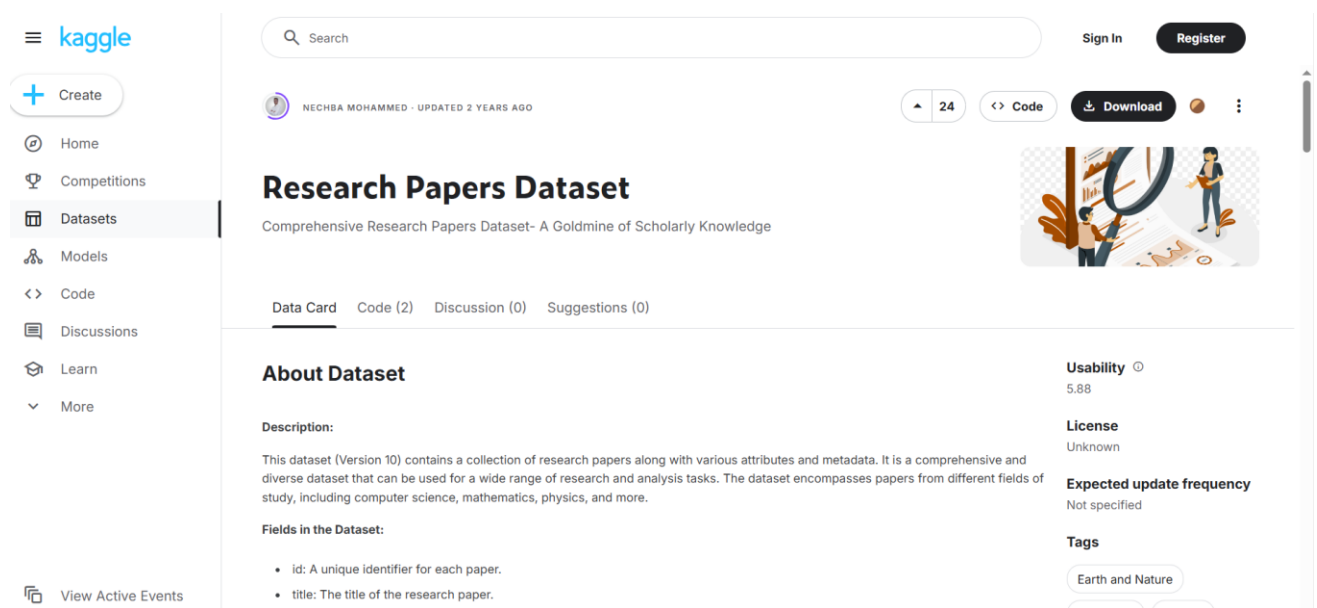


Рисунок 2.3 – Головна сторінка датасету Research Papers Dataset [32]

Датасет складається з багатьох записів, кожен з яких представляє окрему наукову статтю. Кожен запис містить поля які зображено на рисунку 2.4.

Поле «id» містить унікальний ідентифікатор статті, що є унікальним та ключовим для структурованої роботи з даними. Цей ідентифікатор служить основою для зв'язків між статтями, зокрема в полі «references», де зберігаються посилання на інші роботи.

Поле «title» містить у собі назву статті, що є першим елементом з яким користувач вступає в контакт. Поле «title» має важливе значення не лише для попередньої класифікації чи ранжування, але й для побудови векторних подань тексту. Саме «title» містить ключові слова, які відображають основну тематику

публікації. У рамках адаптивної фільтрації «title» було використано як індикатор релевантності.

Δ abstract	Δ authors	# n_citation	Δ references	Δ title	Δ venue	# year	∞ id
[null] 17% Quality and Reliabi... 0% Other (827395) 83%	865804 unique values		[null] 12% [] 6% Other (819188) 82%	999064 unique values	[null] 18% international confe... 1% Other (810413) 81%		1000000 unique values
We studied an autoassociative neural network with dynamic synapses which include a facilitating mech...	['Joaquin J. Torres', 'Jesus M. Cortés', 'Joaquin Marro', 'Hilbert J. Kappen']	50	4996... ['4017c9d2-9845-4ad2-ad5b-ba6552372c5', 'b1187381-cc24-471e-b288-5da09c3d4ca9', 'ecf8e7f0-cac6-47e4...']	Attractor neural networks with activity-dependent synapses: The role of synaptic facilitation	Neurocomputing	2007	4ab39729-af77-46f7-a662-16984fb9c1db
It is well-known that Sturmian sequences are the non ultimately periodic sequences that are balanced...	['Genevi eve Paquin', 'Laurent Vuillon']	50	['1c655ee2-067d-4bc4-b8cc-bc779e9a7f10', '2e4e57ca-d190-4bc7-b7ad-2e33aca2f574', '508ea83a-835b-4542...']	A characterization of balanced episturmian sequences	Electronic Journal of Combinatorics	2007	4ab3a4cf-1d96-4ce5-ab6f-b3e19fc260de
One of the fundamental challenges of recognizing actions is accounting for the variability that aris...	['Yaser Sheikh', 'Mumtaz Sheikh', 'Mubarak Shah']	221	['056116c1-9e7a-4f9b-a918-44eb199e67d6', '05ac52a1-dfa2-4c32-9a54-6c1e58cfcca5', '0b2f71cc-6797-445a...']	Exploring the space of a human action	international conference on computer vision	2005	4ab3a98c-3620-47ec-b578-884ecf4a6206
This paper generalizes previous optimal upper bounds on the minimum Euclidean distance for phase shi...	['Efraim Laksman', 'Håkan Lennerstad', 'Magnus Nilsson']	0	['01a765b8-0cb3-495c-996f-29c36756b435', '5dbc8ccb-0014-4cb8-b72b-e2a5b44b671d', '71ea2d72-eb37-409e...']	Generalized upper bounds on the minimum distance of PSK block codes	Ins Journal of Mathematical Control and Information	2015	4ab3b585-82b4-4207-91dd-b6bce7e27c4e
Queueing networks with multiple classes of customers play a fundamental role for evaluating the perf...	['Simonetta Balsamo', 'Gian-Luca Dei Rossi', 'Andrea Marin']	6	['1c26e228-57d2-4b2c-b8c9-8d5851c17fac', '75399207-1568-4ee7-ba52-ef10f26add66', '85244323-5e23-...']	Applying BCMP multi-class queueing networks for the performance evaluation of hierarchical and	International Journal of Computer Aided Engineering and Technology	2015	4ab3e768-78c9-4497-8b8e-9e934cb5f2e4

Рисунок 2.4 – Структура полів, які містить датасет Research Papers Dataset [32]

Поле «authors» містить імена всіх дослідників, що брали участь у написанні роботи. Це поле дозволяє відстежувати наукову активність окремих осіб або груп, будувати мережі співпраці, а також визначати експертність в конкретних галузях.

Поле «venue» вказує, де саме було опубліковано дослідження. Важливість цього поля полягає у можливості фільтрувати статті за престижністю чи тематикою видання.

Поле «year» є надзвичайно важливим у контексті оновлюваності інформації. Для багатьох галузей знань актуальність дослідження є критичним фактором зокрема в комп'ютерних науках, машинному навчанні. Поле «year» дає змогу фільтрувати або ранжувати результати за хронологією, будувати часові лінії інтересів користувача.

«N_citation» поле, яке показує наскільки робота була процитована іншими дослідженнями. Кількість цитувань традиційно розглядається як показник якості.

Поле «reference» означає список ідентифікаторів інших робіт, які цитуються у даній статті.

Поле «abstract» є одне з найважливіших полів у контексті NLP і рекомендаційних систем. Анотація дає стислий, але інформативний виклад суті дослідження. Аналіз цього тексту дозволяє застосовувати векторні подання, тематичне моделювання, трансформерні архітектури для глибшого розуміння змісту.

Отже, датасет містить дані для роботи та розробки системи адаптивної фільтрації маючи вже наявні статті, що належать до кількох категорій, що дозволяє моделювати складні взаємозв'язки між тематиками та інтересами користувачів. Також поля «title» та «abstract» надають текстовий матеріал для застосування методів обробки природної мови.

2.4 Нейромережева архітектура для отримання векторних представлень

Метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання використовує «all-MiniLM-L6-v2» [33] компактну і потужну модель із родини «MiniLM», спеціально оптимізовану для завдань семантичного пошуку та створення «sentence embeddings». Архітектурні особливості цієї моделі, які забезпечують її продуктивність при збереженні точності, зображено на рисунку 2.5.

Початковим етапом обробки є токенизація вхідного тексту. Модель використовує «WordPiece» токенизатор, який є частиною BERT-архітектури. Метою токенизатора є перетворення вхідного речення на послідовність ідентифікаторів токенів, які зрозумілі моделі.

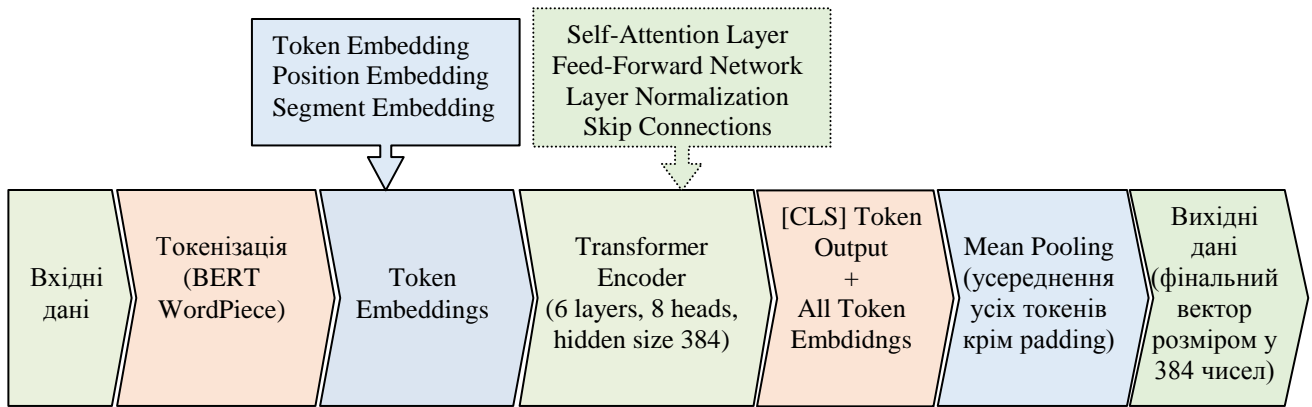


Рисунок 2.5 – Архітектура моделі all-MiniLM-L6-v2

Перед початком і після завершення послідовності додаються спеціальні токени «CLS» і «SEP». «CLS» зазвичай використовується для узагальнення всього речення, а «SEP» для розділення речень у парах або завершення послідовності.

Після етапу токенізації кожен токен перетворюється на вектор фіксованої розмірності за допомогою спеціального шару ембедінгів. Цей шар формує векторне представлення, поєднуючи три типи ембедінгів «Token Embedding», «Position Embedding» та «Segment Embedding».

Token «Embedding» відповідають унікальним векторним значенням, закладеним у словник моделі, і відображають семантичне значення кожного токена.

«Position Embedding» додаються з метою передачі інформації про порядок слів у реченні, оскільки трансформерна архітектура за своєю природою не має вбудованого механізму обробки послідовності.

«Segment Embedding» використовуються для розрізнення частин тексту у випадках, коли модель працює з кількома реченнями водночас.

Ці три типи ембедінгів підсумовуються, утворюючи вхідні вектори, які подаються на перший шар трансформера.

У моделі «all-MiniLM-L6-v2» таких шарів є шість. Кожен шар трансформера складається з двох основних компонентів «Self-Attention» і «Feed-Forward Network».

«Self-Attention» забезпечує можливість кожному токенові аналізувати інші токени в межах тієї ж послідовності, визначаючи їхню значущість для поточного контексту. У даній моделі використовується вісім голів уваги, кожна з яких працює у власному проєкційному просторі. Завдяки цьому моделі можуть паралельно враховувати різні аспекти контексту, виявляючи як локальні, так і віддалені залежності між словами.

Після проходження через шар уваги, кожен токен передається до повнозв'язного шару, який застосовує нелінійну функцію активації зазвичай «GELU» або «ReLU». Таким чином допомагає моделі краще обробляти дані та виділяти важливі ознаки з вхідних векторів. Обидві частини трансформерного блоку, шар уваги, і повнозв'язна мережа доповнюються залишковими зв'язками та «Layer Normalization». Робиться для того, щоб модель навчалася стабільніше та швидше.

У моделі «all-MiniLM-L6-v2» є шість таких блоків, які повторюються один за одним. Кожен наступний блок уточнює представлення токенів, враховуючи контекст усіх інших токенів у реченні. Тобто, на виході кожен токен має вже не лише своє значення, а й інформацію про оточуючі слова.

На завершальному етапі модель об'єднує всі ці токен-вектори в одне загальне представлення для всього речення. Для цього застосовується метод «mean pooling», що обчислює середнє значення всіх векторів, крім тих, що відповідають службовим токенам. У результаті формується «sentence embedding» фіксованої довжини 384 елементи.

Отже, модель «all-MiniLM-L6-v2» використовує архітектуру трансформера, яка складається з послідовних шарів уваги та повнозв'язних мереж. Поєднання різних типів ембедінгів, механізмів самоуваги та глибокої обробки векторів дозволяє моделі формувати контекстуалізовані представлення токенів. Завдяки цьому вона здатна точно відображати зміст речень у вигляді компактних векторів «sentence embeddings».

2.5 Особливості використання програмних розширень

У процесі розробки методу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання було використано низку програмних бібліотек, які забезпечили ефективну реалізацію всіх етапів методу від попередньої обробки даних до побудови та навчання моделей, а також формування рекомендацій.

Однією з бібліотек, що використовується для програмної реалізації методу, є «psycopg2» [34], яка дозволяє ефективно зчитувати, фільтрувати, з'єднувати та трансформувати табличні дані, що є необхідним етапом перед подальшим аналізом і векторизацією.

Для підвищення якості векторного подання тексту в роботі було застосовано модель «sentence_transformers» [35], яка базується на трансформерній архітектурі «BERT» та її варіаціях. На відміну від класичного підходу «TF-IDF», що враховує лише статистичні частоти слів, «sentence_transformers» створює контекстуалізовані вектори речень, які передають глибше семантичне значення тексту. Моделі можуть захоплювати не лише окремі терміни, але й їхні взаємозв'язки в контексті, що особливо корисно при роботі з науковими текстами, де значення слів залежить від теми та оточення. Використання «sentence_transformers» забезпечує більш точне порівняння та класифікацію наукових статей за змістом, підвищуючи ефективність семантичного аналізу.

Для зберігання та пошуку векторних представлень тексту було використано бібліотеку FAISS [36] у поєднанні з базою даних PostgreSQL [37].

FAISS – це бібліотека від Facebook, яка забезпечує ефективний пошук найближчих сусідів у великих векторних просторах. Вона широко застосовується для систем рекомендацій, класифікації, кластеризації та пошуку схожих об'єктів, підтримуючи як точний, так і приблизний пошук, що значно знижує витрати ресурсів при роботі з великими обсягами даних. Завдяки різноманіттю індексів і можливості масштабування на «GPU», FAISS

залишається одним із провідних інструментів для завдань інформаційного пошуку.

Для організації зберігання векторів та пов'язаних метаданих використовується PostgreSQL.

PostgreSQL надійна реляційна база даних із підтримкою розширень для роботи з векторними типами даних і індексацією. БД дозволяє керувати колекціями векторів, зберігати метадані а також організовувати гнучкий доступ до інформації, що є ключовим для побудови методу.

Завдяки поєднанню FAISS та PostgreSQL можна досягти високої продуктивності як у зберіганні, так і в пошуку даних. FAISS використовується безпосередньо для побудови індексів та швидкого пошуку, а PostgreSQL слугує надійним сховищем для ембеддингів і метаданих із можливістю масштабування й складної організації даних. Після формування ембеддингів вони зберігаються в PostgreSQL, а FAISS відповідає за швидкий пошук найближчих сусідів. У результаті користувачі можуть отримувати максимально точні відповіді навіть у випадку великих колекцій документів.

Отже, застосування зазначених бібліотек дозволяє реалізувати ефективну систему адаптивної фільтрації наукових статей, яка поєднує класичні методи векторизації, глибокий аналіз тексту та ефективний векторний пошук. Забезпечує високий рівень персоналізації рекомендацій та дозволяє користувачам отримувати релевантні наукові статті відповідно до їхніх інтересів.

2.6 Висновки до розділу 2

У ході виконання другого розділу було створено метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання. Розроблений підхід поєднує в собі сучасні методи обробки природної мови, алгоритми семантичного аналізу та технології векторного

пошуку, що дозволяє забезпечити високоточну й адаптивну систему інтелектуальних рекомендацій для користувача.

Було розроблено схему методу фільтрації, яка включає послідовні кроки обробки вхідних даних, формування векторних уявлень тексту, класифікацію за інтересами користувача та подальшу видачу релевантних результатів. Запропонований метод забезпечує не лише фільтрацію за ключовими словами, а й глибоке семантичне порівняння змісту статей, що сприяє покращенню процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання. Такий підхід можливий завдяки інтеграції, як класичних методів на кшталт TF-IDF, так і глибоких нейронних моделей, таких як BERT.

Наведено проектну архітектуру інформаційної системи, яка включає кілька взаємопов'язаних компонентів. Також було визначено датасет Research Papers Dataset, що є важливим практичним етапом.

Окрему увагу було приділено програмним розширенням, які використовувалися для реалізації кожного етапу. Основу системи склали бібліотеки «psycopg2» для обробки запитів, «sentence-transformers» для роботи з мовними моделями, FAISS для векторного пошуку і PostgreSQL. Завдяки сумісності цих інструментів вдалося реалізувати гнучку систему, здатну до масштабування та розширення.

Розділ 3 Експериментальне дослідження методу

3.1 Методика дослідження

З метою оцінювання ефективності методу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій із застосуванням технологій машинного навчання запропоновано проведення експертного дослідження. У межах цього дослідження два незалежні експерти з досвідом у галузі комп'ютерних наук здійснювали порівняльну оцінку якості сформованих системою рекомендацій. Оцінювання здійснювалося за п'ятибальною шкалою, де значення «1» відповідало повній нерелевантності, а «5» – повній релевантності рекомендованих статей щодо інтересів користувача.

Для забезпечення об'єктивності результатів метод пропонується порівняти з системою рекомендацій Google Scholar, що виступає еталонним прикладом у сфері наукової інформаційної аналітики. Ефективність пропонується оцінювати за кількома ключовими критеріями: відповідність рекомендованих матеріалів науковим інтересам користувача (релевантність), здатність системи адаптуватися до індивідуальних потреб (персоналізація), швидкість отримання змістовно значущої інформації, зручність взаємодії з інтерфейсом та загальний рівень задоволеності користувача результатами взаємодії із системою. Отримані оцінки лягли в основу аналізу точності, практичної доцільності та потенціалу впровадження розробленого методу в реальні інформаційно-наукові середовища.

Методика проведення дослідження наведено на рисунку 3.1. Схематично представлено логічну структуру проведення дослідження, яка охоплює всі ключові етапи від вхідних даних про наукові інтереси користувача, до формування рекомендацій із застосуванням методів машинного навчання, а також порівняння результатів за допомогою Google Scholar.

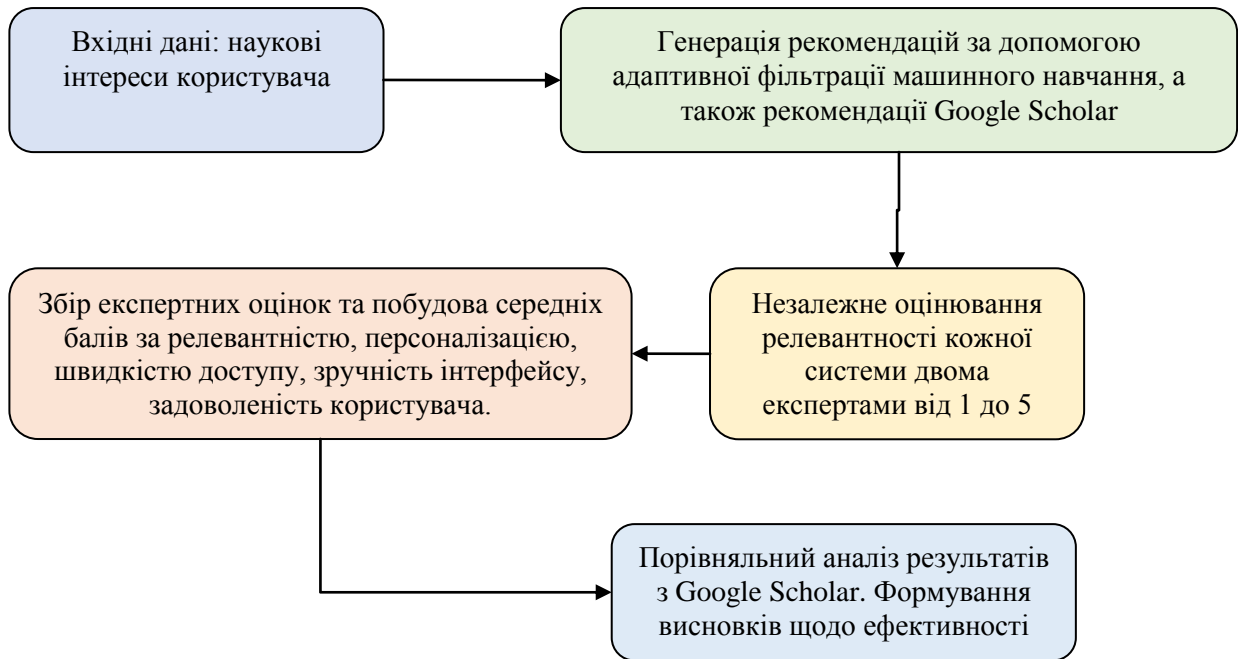


Рисунок 3.1 – Методика проведення дослідження

Отже, розроблена методика дослідження передбачає залучення експертного оцінювання за рядом ключових категорій, що дозволяє здійснити всебічний аналіз якості рекомендацій. Порівняння результатів з системою Google Scholar забезпечує об'єктивність висновків та демонструє практичну доцільність і переваги впровадженого підходу. Схематично представлений план дослідження дає змогу послідовно реалізувати етапи аналізу та забезпечити надійність отриманих результатів.

3.2 Програмна структура інтелектуальної системи

Інтелектуальна система розумової фільтрації наукових статей побудована у вигляді веб-застосунку з клієнтською та серверною частинами.

Вона складається з наступних сторінок:

- головна сторінка;
- сторінка статті.

Головна сторінка виконує основну функцію взаємодії користувача із системою. На ній розташована форма пошуку та селектор вибору за яким

критерієм робити пошук. Також форма дозволяє вводити ключові слова або ж використати голосовий пошук.

Після виконання запиту на сторінці динамічно відображається список результатів наукових статей, що відповідає критеріям пошуку. Кожен елемент результатів пошуку містить кнопку, яка дозволяє переходити до перегляду детальної інформації про вибрану статтю.

Сторінка статті надає детальну інформацію про наукову роботу включаючи авторів, рік публікації, анотацію та повний текст.

Функціонування інформаційної системи використовує мікросервісну архітектуру. Кожен сервіс відповідає за певну логіку, що дозволяє в майбутньому масштабувати систему. Схематично представлено архітектуру веб-застосунку на рисунку 3.2.

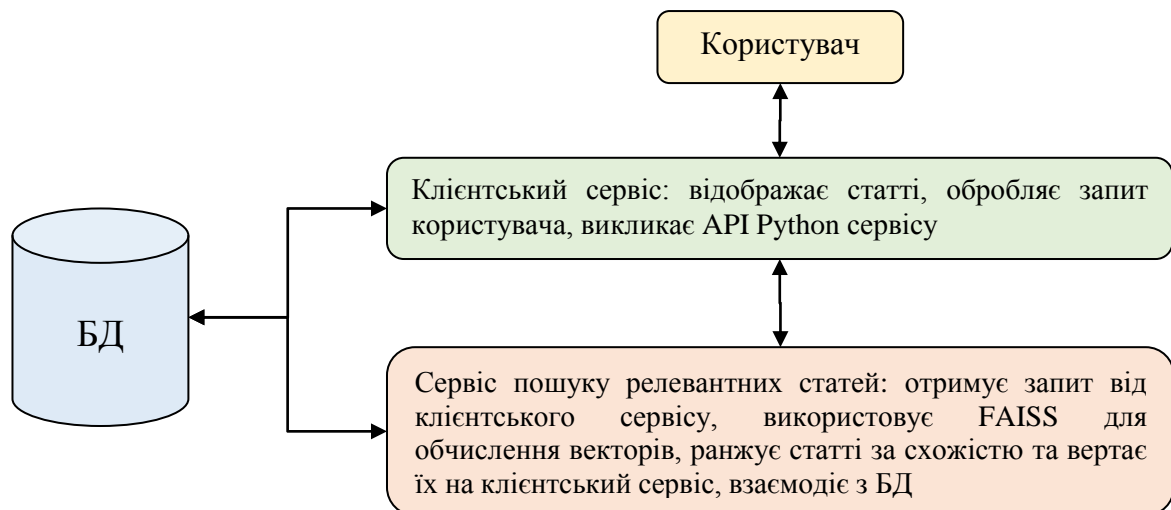


Рисунок 3.2 – Мікросервісна архітектура веб-застосунку

Користувач взаємодіє з клієнтським сервісом, тобто це є точка входу у веб-застосунок. Цей сервіс відповідає за відображення інтерфейсу користувача, обробку запитів та взаємодіє із сервісом пошуку релевантних статей через API. Користувач може здійснювати пошук за ключовими словами або за допомогою голосового вводу, а також переглядати отримані результати.

Сервіс пошуку релевантних статей отримує запит від клієнтського сервісу, використовує векторне перетворення ключових слів користувача за

допомогою FAISS, далі звертається у БД витягуючи всі вектори до кожної статті та обчислює схожість між ключовими словами та науковими статтями. На основі обчисленої релевантності ранжує список статей та повертає результат назад до клієнтського сервісу у форматі JSON [38].

Усі дані, включно з метаданими статей і векторними представленнями, зберігаються в єдиній базі даних PostgreSQL. Обидва сервіси взаємодіють із базою даних клієнтський для отримання загальної інформації, а серверний для доступу до векторних даних та їх обробки.

Таким чином, система забезпечує гнучкий, персоніфікований підхід до пошуку наукових статей з використанням методів машинного навчання, водночас залишаючись масштабованою завдяки розділенню функціоналу між окремими сервісами.

3.3 Вибір засобів розробки інтелектуальної системи адаптивної фільтрації наукових статей

Під час розробки інтелектуальної системи адаптивної фільтрації наукових статей із використанням машинного навчання важливим етапом є обґрунтований вибір програмних засобів, платформ, мов програмування, інструментів розробки та БД.

Для реалізації ядра системи персоніфікованих рекомендацій, зокрема компонентів, пов'язаних із машинним навчанням, було обрано мову програмування Python [39]. Такий вибір зумовлений широкими можливостями Python у сфері обробки природної мови, глибокого навчання та роботи з векторними представленнями тексту. Мова має потужну екосистему бібліотек, зокрема «scikit-learn», «transformers», що дозволяють швидко реалізовувати і тестувати різноманітні алгоритми фільтрації, класифікації, кластеризації, а також оцінювати їхню ефективність. Крім того, Python тісно інтегрується з такими системами векторного пошуку, як FAISS, що є одним із ключових компонентів проєкту.

Для реалізації веб-застосунку та серверної частини було використано фреймворк Laravel [40]. Він надає зручний та елегантний синтаксис для створення REST API, який слугує мостом між клієнтською частиною, системою рекомендацій та базою даних. Його вбудовані механізми маршрутизації, аутентифікації, контролю доступу, а також «ORM Eloquent» [41] значно спрощують розробку, тестування та масштабування серверної логіки.

Для зберігання наукових статей, метаданих, а також інформації про користувачів і їхні дії була обрана система керування базами даних PostgreSQL. PostgreSQL – це об'єктно-реляційна СКБД з відкритим кодом, яка підтримує складні запити, індексацію, транзакції та роботу з JSON-структурами. Такий вибір обумовлений її надійністю, продуктивністю при роботі з великими обсягами даних, а також сумісністю з багатьма бібліотеками машинного навчання. Крім того, PostgreSQL добре масштабується, що дозволяє у перспективі розширювати обсяг збережених даних без істотних змін в архітектурі.

FAISS є високопродуктивною бібліотекою, розробленою для швидкого пошуку схожих векторів у великих масивах даних. В інтелектуальній системі адаптивної фільтрації статей, FAISS використовується для зберігання векторних представлень наукових публікацій, які формуються за допомогою моделей перетворення тексту в ембедінги. Використання FAISS дозволяє ефективно реалізовувати пошук за семантичним змістом, тобто не лише за ключовими словами, а й за глибинною подібністю між текстами. Таким чином суттєво покращить якість персоналізованих рекомендацій, наближаючи результати до реальних потреб користувача.

Для розгортання всієї інформаційної системи, включно з сервісами машинного навчання, базою даних, веб-інтерфейсом та службами API, було використано Docker [42].

Docker дозволяє створювати ізольовані контейнери для кожного компонента системи, що забезпечує стабільність, повторюваність оточення, а також спрощує розгортання як у середовищі розробки, так і в продакшн

середовищі. Контейнеризація дозволяє уникнути проблем сумісності між бібліотеками, полегшує CI/CD процеси, а також дає можливість швидко масштабувати систему при збільшенні навантаження.

В межах проєкту було створено окремі контейнери для Laravel додатку, сервісу на Python із FAISS, бази даних PostgreSQL та фонових обробників.

Усі вищеописані компоненти було об'єднано в єдину систему, яка функціонує відповідно до концепції мікросервісної архітектури.

Сервіс Python відповідає за побудову векторних моделей і фільтрацію, Laravel виконує роль координаційного та комунікаційного шару з клієнтом, а PostgreSQL слугує надійним сховищем усіх даних.

Отже обрані технології дозволяють побудувати гнучку, модульну та надійну систему, здатну до подальшого розвитку в напрямі глибшої персоналізації та адаптації до змін користувацьких запитів.

3.4 Особливості реалізації програмних складових інтелектуальної системи адаптивної фільтрації наукових статей

Інтелектуальна система адаптивної фільтрації наукових статей реалізована у вигляді веб-застосунку. Процес фільтрації розпочинається з генерації ембедінгів для тексту. Нижче наведено алгоритм генерації.

```

Підключитись до бази даних
Отримати всі тексти у яких немає ембедінгів, але є title та abstract
Для кожної статті:
    Отримати title та abstract
    Побудувати ембедінг для abstract
    Побудувати ембедінг для title
    Побудувати ембедінг для title + abstract
Зберегти всі три ембедінги у відповідні поля в базі даних
Закрити з'єднання з базою
  
```

Цей алгоритм забезпечує формування векторних представлень тестових даних для кожної наукової статті. На початку підключається до БД та знаходить записи без ембедінгів. Наступні дії алгоритму полягають в обчисленні ембедінгів

за допомогою машинного навчання і зберігає їх назад у БД. Ці дії допомагають готувати дані до швидкого пошуку за змістом.

Наступний алгоритм описує отримання ембедінгів з БД.

```

підключитися до бази даних
отримати всі наукові статті з повними ембедінгами
ініціалізувати списки: ids, metadata, embeddings
для кожної наукової статті:
    зчитати id, title, abstract, тощо
    перетворити json-ембедінги у числові масиви
    додати до відповідних списків
закрити з'єднання
повернути списки

```

За допомогою цього алгоритму зчитуємо всі ембедінги, які вже були згенеровані. Кожен запис обробляється і дані зберігаються у відповідних структурах для подальшого використання FAISS індексації. Зокрема ембедінги перетворюються з формату JSON у числовий вигляд.

Після форматування потрібно ініціалізувати FAISS індекси, що робить наступний алгоритм.

```

для кожного набору ембедінгів (abstract, title, titleabstract):
    якщо список не порожній:
        створити FAISS індекс з L2 відстанню
        додати вектори до індексу
        повідомити про успішне створення
    інакше:
        повідомити, що індекс не створено

```

Будує індекси на основі попередньо з генерованих ембедінгів та використовує FAISS індекси для ефективного пошуку схожих елементів. Якщо набір ембедінгів є індекс створюється і заповнюється, а якщо немає то виводиться попередження.

Наступний алгоритм продемонструє роботу пошуку за схожістю тексту через FAISS.

```

якщо індекс не існує або порожній:
    повернути помилку
зчитати запит та кількість результатів
створити embedding з тексту запиту
знайти найближчі вектори через FAISS

```

для кожного результату:
 отримати id і відстань
 сформувати відповідь з метаінформацією
 повернути список результатів

Запит користувача перетворюється в ембедінг, а потім виконує пошук схожих ембедінгів у FAISS індексі. Як результат повернуться результати у форматі JSON із метаінформацією, яка передається на користувацький інтерфейс (рисунок 3.3).

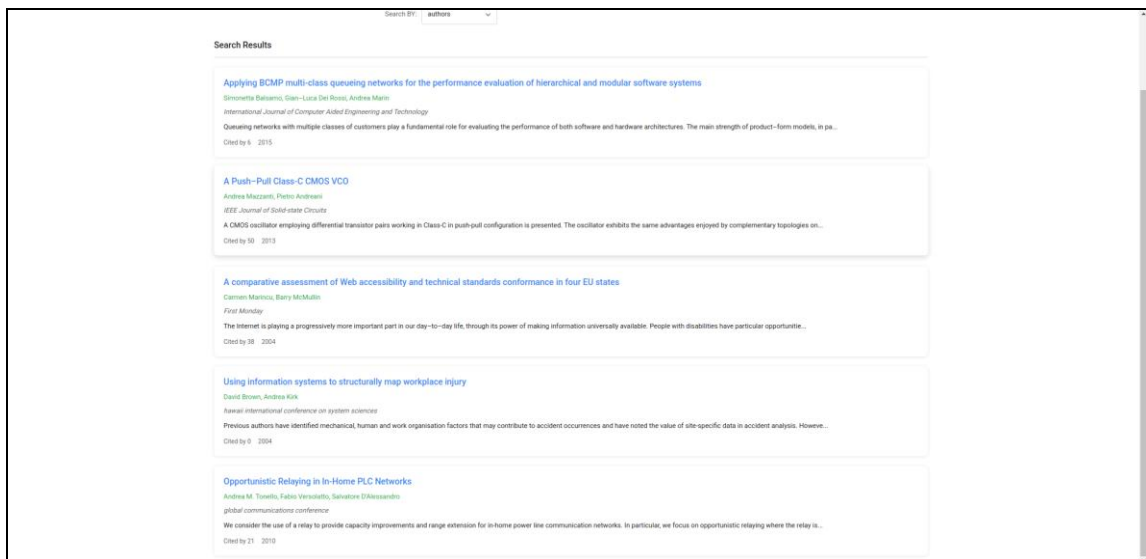


Рисунок 3.3 – Результат ранжованого списку наукових статей

Отже, було реалізовано повний цикл обробки наукових текстів від зчитування з БД до виконання семантичного пошуку за допомогою FAISS індексів.

3.5 Аналіз функціональності створеної інтелектуальної системи адаптивної фільтрації наукових статей

Для створення інтелектуальної системи адаптивної фільтрації наукових статей було прийнято рішення обрати веб-розробку. Перейшовши на головну сторінку сайту можна побачити у лівому куті логотип, а також саму пошукову систему (рисунок 3.4).

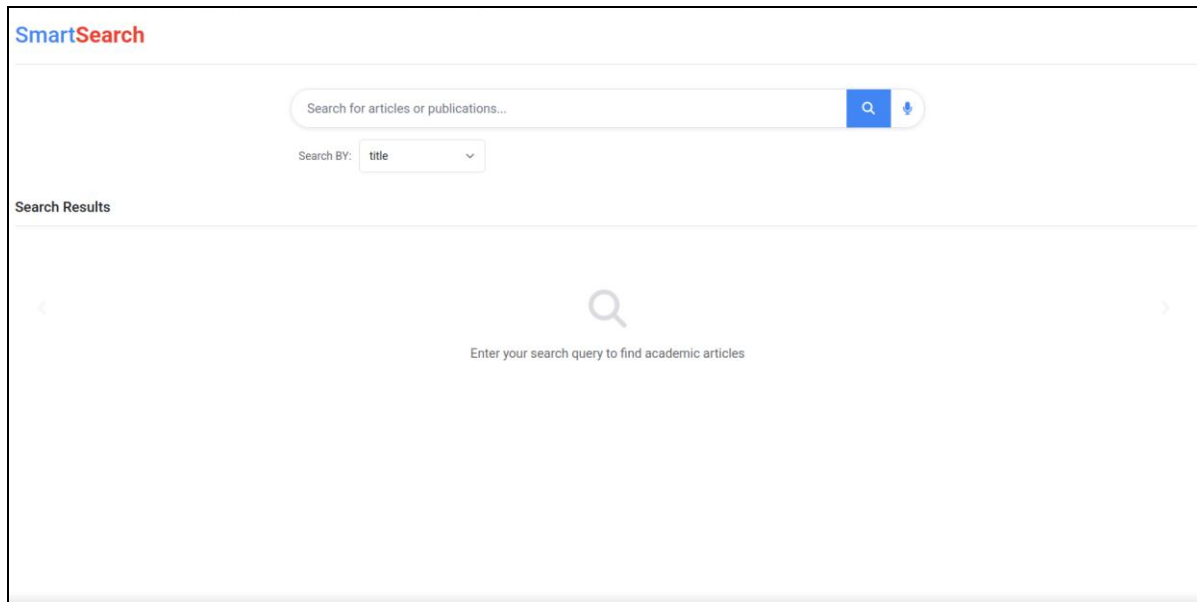


Рисунок 3.4 – Головна сторінка веб-сайту інтелектуальної системи адаптивної фільтрації наукових статей

Наступний етап це користувач може здійснити пошук за певними критеріями, тобто обравши за «title», «abstract» чи «title and abstract» (рисунок 3.5).

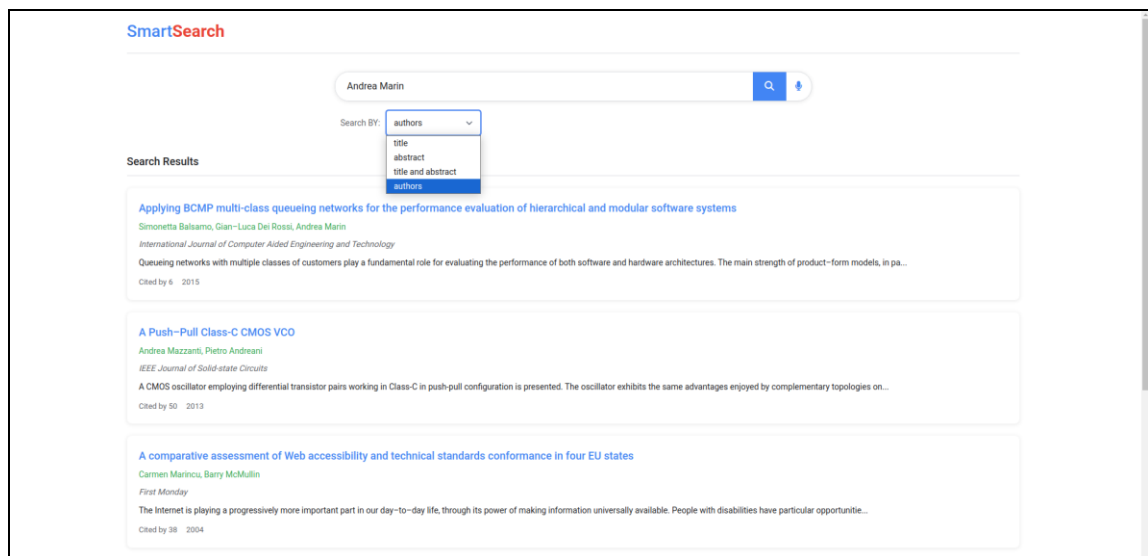


Рисунок 3.5 – Результат пошуку за критерієм пошуку «title»

Також користувач може переходити на сторінку яка відображає повну інформацію про наукову статтю (рисунок 3.6).

SmartSearch

Home /Limits of homology detection by pairwise sequence comparison

Limits of homology detection by pairwise sequence comparison

Authors: Rainer Spang, Martin Vingron,
Limits of homology detection by pairwise sequence comparison 2001 Venue: Bioinformatics

29 Citations

Abstract

Motivation: Noise in database searches resulting from random sequence similarities increases as the databases expand rapidly. The noise problems are not a technical shortcoming of the database search programs, but a logical consequence of the idea of homology searches. The effect can be observed in simulation experiments. Results: We have investigated noise levels in pairwise alignment based database searches. The noise levels of 38 releases of the SwissProt database, display perfect logarithmic growth with the total length of the databases. Clustering of real biological sequences reduces noise levels, but the effect is marginal.

Рисунок 3.6 – Інформаційна сторінка наукової статті

Перебуваючи на цій сторінці користувач зможе побачити список ідентифікаторів інших робіт, які цитуються у даній статті (рисунок 3.7).

Abstract

Motivation: Noise in database searches resulting from random sequence similarities increases as the databases expand rapidly. The noise problems are not a technical shortcoming of the database search programs, but a logical consequence of the idea of homology searches. The effect can be observed in simulation experiments. Results: We have investigated noise levels in pairwise alignment based database searches. The noise levels of 38 releases of the SwissProt database, display perfect logarithmic growth with the total length of the databases. Clustering of real biological sequences reduces noise levels, but the effect is marginal.

Related Articles

Towards a universal client for grid monitoring systems: design and implementation of the Ovid browser

Marios D. Dikaiakos, Artemakis Artemiou, George Tsouloupas

International parallel and distributed processing symposium 2006

In this paper, we present the design and implementation of Ovid, a browser for grid-related informat...

Multiple ant tracking with global foreground maximization and variable target proposal distribution

Mary Fletcher, Anna Dornhaus, Min C. Shin

Workshop on applications of computer vision 2011

Motion and behavior analysis of social insects such as ants requires tracking many ants over time. T...

THE SRI NIST 2008 speaker recognition evaluation system

Sachin S. Kajarekar, Nicolas Scheffer, Martin Graclarena, Elizabeth Shriberg, Andreas Stolcke, Luciana Ferrer, Tobias Bocklet

International conference on acoustics, speech, and signal processing 2009

The SRI speaker recognition system for the 2010 NIST speaker recognition evaluation (SRE) incorporat...

Рисунок 3.7 – Список ідентифікаторів інших робіт, які цитуються у даній статті

Отож, було продемонстровано функціональні особливості розробленої інтелектуальної системи адаптивної фільтрації наукових статей.

3.6 Результати досліджень

Відповідно до визначених шляхів дослідження методу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій із застосуванням технологій машинного навчання було проведено дослідження, де здійснено експертне оцінювання системи за такими ключовими критеріями: релевантність результатів, тобто відповідність рекомендованих статей науковим інтересам користувача; рівень персоналізації, що відображає здатність системи адаптуватися до індивідуальних потреб; швидкість доступу до змістовно значущого контенту; зручність користування інтерфейсом; а також загальний рівень задоволення користувачів результатами взаємодії з системою.

Оцінювання здійснювалося двома незалежними експертами за п'ятибальною шкалою, що дозволило врахувати суб'єктивну думку кожного фахівця та усереднити результати для об'єктивного порівняння. Дані експерименту наведені у таблиці 3.1 та 3.2.

Таблиця 3.1 – Експертне оцінювання методу адаптивної фільтрації наукових статей

Критерії	Експерт 1	Експерт 2	Усереднена оцінка
Релевантність результатів	4	5	4,5
Персоналізація	4	4	4
Швидкість доступу до релевантного контенту	3	4	3,5
Зручність користування інтерфейсом	4	5	4,5
Задоволення користувачів	4	4	4

Відповідно до таблиць 3.1 та 3.2, було побудовано діаграму порівняння розробленого та існуючого методів на основі рішень 2-х експертів (рисунок 3.8).

Таблиця 3.2 – Експертне оцінювання використання Google Scholar

Критерії	Експерт 1	Експерт 2	Усереднена оцінка
Релевантність результатів	3	4	3,5
Персоналізація	4	3	3,5
Швидкість доступу до релевантного контенту	4	4	4
Зручність користування інтерфейсом	4	3	3,5
Задоволення користувачів	4	5	4,5

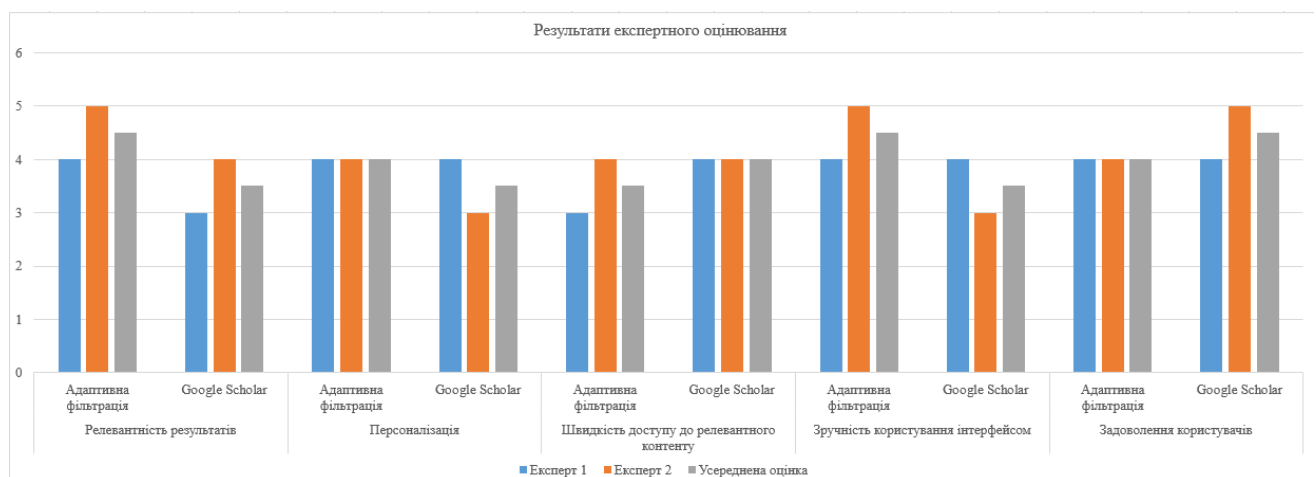


Рисунок 3.8 – Діаграма порівняння розробленого та існуючого методів на основі рішень 2-х експертів

Згідно з проведеним експертним оцінюванням, метод адаптивної фільтрації демонструє кращі результати у критеріях релевантності, персоналізації та зручності користування. Водночас Google Scholar показав вищі оцінки за критеріями швидкості доступу до релевантного контенту та рівня задоволення користувачів, що пояснюється його стабільністю, широкою базою даних та відомістю серед науковців.

Отже, результати дослідження підтверджують, що метод адаптивної фільтрації наукових статей із використанням машинного навчання є ефективним

для створення персональних рекомендацій. Це показує, що систему можна успішно вдосконалювати й використовувати в навчанні та науці.

3.7 Висновки до розділу 3

Було проведено практичне дослідження методу адаптивної фільтрації наукових статей, реалізованого у вигляді інформаційної системи з використанням сучасних веб-технологій та інструментів машинного навчання.

Розроблена методика дослідження передбачала залучення експертного оцінювання за ключовими критеріями таких як релевантність, персоналізація, швидкість доступу до контенту, зручність інтерфейсу та загальна задоволеність користувачів. Порівняльний аналіз з системою Google Scholar дозволив отримати об'єктивну оцінку якості рекомендацій та підтвердив практичну доцільність запропонованого підходу.

Було створено мікросервісну архітектуру, яка забезпечує гнучкість і масштабованість системи. Розділення функціоналу між клієнтською частиною і серверною дозволяє легко оновлювати або доповнювати компоненти без шкоди для стабільності всієї системи.

Реалізований цикл обробки охоплює всі етапи зчитування наукових статей з бази даних PostgreSQL, перетворення їх у векторне представлення за допомогою FAISS та здійснення семантичного пошуку відповідно до запиту користувача. Це забезпечує точну відповідність інтересам користувача на основі його запиту.

Також було реалізовано веб-інтерфейс системи, який дозволяє користувачам зручно здійснювати пошук, переглядати результати та переходити до детальної інформації про наукові статті.

Запропонована система може бути успішно використана в науковій та освітній діяльності, а також має потенціал для подальшого розвитку й удосконалення.

Загальні висновки

Результатом виконання кваліфікаційної роботи бакалавра є досягнення поставленої мети, що полягала в покращенні процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання. Покращення процесу адаптивної фільтрації наукових статей полягає у підвищенні релевантності та персоналізації рекомендацій завдяки застосуванню моделі all-MiniLM-L6-v2 для створення семантично точних векторних представлень тексту, що враховують контекст і змістовну схожість статей.

Для досягнення поставленої мети були успішно виконані наступні задачі:

- проведено аналіз існуючих інформаційних моделей адаптивної фільтрації наукових статей, що дозволило визначити основні підходи, труднощі та потреби користувачів;

- розроблено метод адаптивної фільтрації наукових статей, що базується на машинному навчанні та враховує інтереси користувача;

- реалізовано прикладну програмну систему, яка поєднує функціональність збору, обробки, індексації та персоніфікованого пошуку наукових текстів;

- проведено дослідження методу з використанням запропонованого веб-застосунку, зокрема шляхом порівняння з системою Google Scholar, що підтвердило переваги в релевантності та адаптивності рекомендацій.

Також було створено інтелектуальну систему, яка забезпечує персоніфікований пошук наукових публікацій на основі семантичної близькості текстів. Система включає повний цикл роботи, а саме зчитування даних із БД, їхнє векторне подання, індексацію за допомогою FAISS і подальше формування рекомендацій.

Результати дослідження показали ефективність запропонованого методу адаптивної фільтрації, що підтверджує доцільність застосування розробленого методу у науково-освітньому середовищі. За результатами кваліфікаційної роботи було виконано наукову публікацію (тези конференції) [43].

Перелік посилань

1. Adaptive filter. *Sciencedirect*. URL: <https://www.sciencedirect.com/topics/computer-science/adaptive-filter> (date of access: 04.06.2025).
2. Публікація наукової статті. *Наукові публікації*. URL: <https://spubl.com.ua/uk/blog/publishing-a-scientific-article-main-stages-and-tips> (дата звернення: 04.06.2025).
3. Наукові статті. Поняття та види. *Академія статей*. URL: <https://a-articles.com/ua/nauchnye-stati-ponyatie-i-vidy> (дата звернення: 04.06.2025).
4. Структура наукової публікації. *Бібліотека КІІ*. URL: <https://www.library.kpi.ua/struktura-naukovoyi-publikatsiyi> (дата звернення: 04.06.2025).
5. Різні типи наукових статей. *Na5ku*. URL: https://na5ku.com.ua/uk/rizni-tipi-naukovih-statej-oglyadova-eksperimentalna-doslidniczka-ta-in/?srsId=AfmBOoqZTrb2kY4C763_LJUxPrIlgYwau6V_VdPNQ2Uu3Pg60Q6vSDWuT (дата звернення: 04.06.2025).
6. Як влучно підібрати для наукової статті ключові слова?. *Наукові публікації*. URL: <https://spubl.com.ua/uk/blog/kak-metko-podobrat-klyuchevyye-slova-dlya-nauchnoy-stati> (дата звернення: 04.06.2025).
7. Як написати анотацію до наукової статті?. *Наукові публікації*. URL: <https://spubl.com.ua/uk/blog/kak-napisat-annotatsiyu-k-nauchnoy-statye-mysli-uchenikh-so-vsego-mira> (дата звернення: 04.06.2025).
8. Оформлення списку використаних джерел. *Бібліотека рівненського державного гуманітарного університету*. URL: <https://library.rshu.edu.ua/index.php/informatsiini-resursy-posluhu/posluhu/spiski-dzherel> (дата звернення: 04.06.2025).
9. Штучна нейронна мережа. *Wikipedia*. URL: https://uk.wikipedia.org/wiki/Штучна_нейронна_мережа (дата звернення: 04.06.2025).

10. Розкриття потужності великих мовних моделей (LLM). Unite.AI. URL: <https://www.unite.ai/uk/large-language-models> (дата звернення: 04.06.2025).
11. Векторні представлення. *Machine Learning*. URL: <https://developers.google.com/machine-learning/crash-course/embeddings/embedding-space?hl=uk> (дата звернення: 04.06.2025).
12. Tf-idf. *Wikipedia*. URL: <https://uk.wikipedia.org/wiki/TF-IDF> (date of access: 04.06.2025).
13. Огляд чотирьох популярних NLP-моделей. *Senior.ua*. URL: <https://senior.ua/articles/oglyad-chotiroh-populyarnih-nlpmodeley> (дата звернення: 04.06.2025).
14. Семантичний пошук JavaFX, Word2Vec, Doc2Vec. *Прикладне програмування з нуля*. URL: <https://www.ap-impulse.com/semantichnij-poshuk-javafx-word2vec-doc2vec-step-129> (дата звернення: 04.06.2025).
15. BERT (модель мови). *Wikipedia*. URL: [https://uk.wikipedia.org/wiki/BERT_\(модель_мови\)](https://uk.wikipedia.org/wiki/BERT_(модель_мови)) (дата звернення: 04.06.2025).
16. Brief Review – SciBERT. *Medium*. URL: <https://sh-tsang.medium.com/brief-review-scibert-a-pretrained-language-model-for-scientific-text-57c44772f03e> (дата звернення: 04.06.2025).
17. Використання векторних баз даних у генеративному штучному інтелекті. *Ерам*. URL: <https://careers.epam.ua/blog/using-vector-databases-in-generative-ai> (дата звернення: 04.06.2025).
18. Як використовувати векторні бази даних з розширеним пошуком для потужних LLM – додатків. *Skim AI*. URL: <https://skimai.com/uk/як-використовувати-векторні-бази-дан/> (дата звернення: 04.06.2025).
19. Google scholar. *Google scholar*. URL: <https://scholar.google.com> (date of access: 04.06.2025).
20. Mendeley. *Mendeley*. URL: <https://www.mendeley.com> (date of access: 04.06.2025).
21. ResearchGate. *ResearchGate*. URL: <https://www.researchgate.net> (date of access: 04.06.2025).

22. Google scholar як агрегатор наукової літератури. *Наука та метрика*. URL: <https://nim.media/articles/google-scholar-yak-globalny-agregator-naukovoyi-literaturi-okhopennya-pokazniki-ta-printsipi-roboti> (дата звернення: 04.06.2025).
23. Що таке google scholar. *Soer publishing*. URL: <https://soer-publ.com.ua/uk/blog/shho-take-google-scholar-akademiya-google-dlya-naukovih-publikatsij/> (дата звернення: 04.06.2025).
24. Бібліографічний менеджер Mendeley. *Наукові публікації*. URL: <https://spubl.com.ua/uk/blog/bibliografichesky-menedzher-mendeley-osnovnyue-funktsii> (дата звернення: 04.06.2025).
25. Що таке ResearchGate. *Mindthegraph*. URL: <https://mindthegraph.com/blog/uk/what-is-researchgate/> (дата звернення: 04.06.2025).
26. Deep Adaptive Interest Network: Personalized Recommendation with Context-Aware Learning / S. Huang et al. arXiv, 2024. URL: <https://arxiv.org/abs/2409.02425> (date of access: 24.12.2024).
27. Beyond Relevance: An Adaptive Exploration-Based Framework for Personalized Recommendations / E. Bianchi. arXiv, 2025. URL: <https://arxiv.org/abs/2503.19525> (date of access: 25.03.2025).
28. MovieLens. *Wikipedia*. URL: <https://en.wikipedia.org/wiki/MovieLens> (date of access: 04.06.2025).
29. Intelligent personalized content recommendations based on neural networks / HeQiang Zhou. *International Journal of Intelligent Networks*, 2023. Vol. 4. P. 231–239. URL: <https://doi.org/10.1016/j.ijin.2023.09.001> (date of access: 01.09.2023).
30. Adaptive Collaborative Filtering with Personalized Time Decay Functions for Financial Product Recommendation / A. Ghiye, B. Barreau, L. Carlier, M. Vazirgiannis. arXiv, 2023. URL: <https://arxiv.org/abs/2308.01208> (date of access: 01.08.2023).
31. Afoudi Y., Lazaar M., Al Achhab M. Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network / Y. Afoudi, M. Lazaar, M. Al Achhab.

Simulation Modelling Practice and Theory, 2021. Vol. 113, 102375. URL: <https://doi.org/10.1016/j.simpat.2021.102375> (date of access: 23.12.2021).

32. Research papers dataset. *Kaggle*. URL: <https://www.kaggle.com/datasets/nechbamohammed/research-papers-dataset> (date of access: 04.06.2025).

33. All-MiniLM-L6-v2. *Hugging face*. URL: <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2> (date of access: 04.06.2025).

34. Psycopg2. *PyPI*. URL: <https://pypi.org/project/psycopg2> (date of access: 04.06.2025).

35. Sentence-transformers. *PyPI*. URL: <https://pypi.org/project/sentence-transformers> (date of access: 04.06.2025).

36. Facebook AI Similarity Search. *GitHub*. URL: <https://github.com/facebookresearch/faiss> (date of access: 04.06.2025).

37. Postgresql. *Postgresql*. URL: <https://www.postgresql.org/docs> (date of access: 04.06.2025).

38. JSON. *Wikipedia*. URL: <https://uk.wikipedia.org/wiki/JSON> (date of access: 04.06.2025).

39. Python. *Python*. URL: <https://www.python.org/> (date of access: 04.06.2025).

40. Laravel. *Laravel*. URL: <https://laravel.com/docs/12.x> (date of access: 04.06.2025).

41. ORM Eloquent. *Laravel*. URL: <https://laravel.com/docs/12.x/eloquent> (date of access: 04.06.2025).

42. Docker. *Dockerdocs*. URL: <https://docs.docker.com/get-started/docker-overview> (date of access: 04.06.2025).

43. Andrushchenko D., Klimenko V., Mazurets O. Vector Databases Search for Adaptive Filtering of Scientific Articles / D. Andrushchenko, V. Klimenko, O. Mazurets. Scientific trends in the development of modern technologies. Proceedings XXII International Scientific and Practical Conference. 2025. Krakow, Poland. Pp. 189-195.

ДОДАТКИ

Додаток А

Програмні коди

Вихідний код, використаний у дослідженні, доступний у репозиторії GitHub: <https://github.com/ANBISH/diploma/> (дата звернення: 13.05.2025).

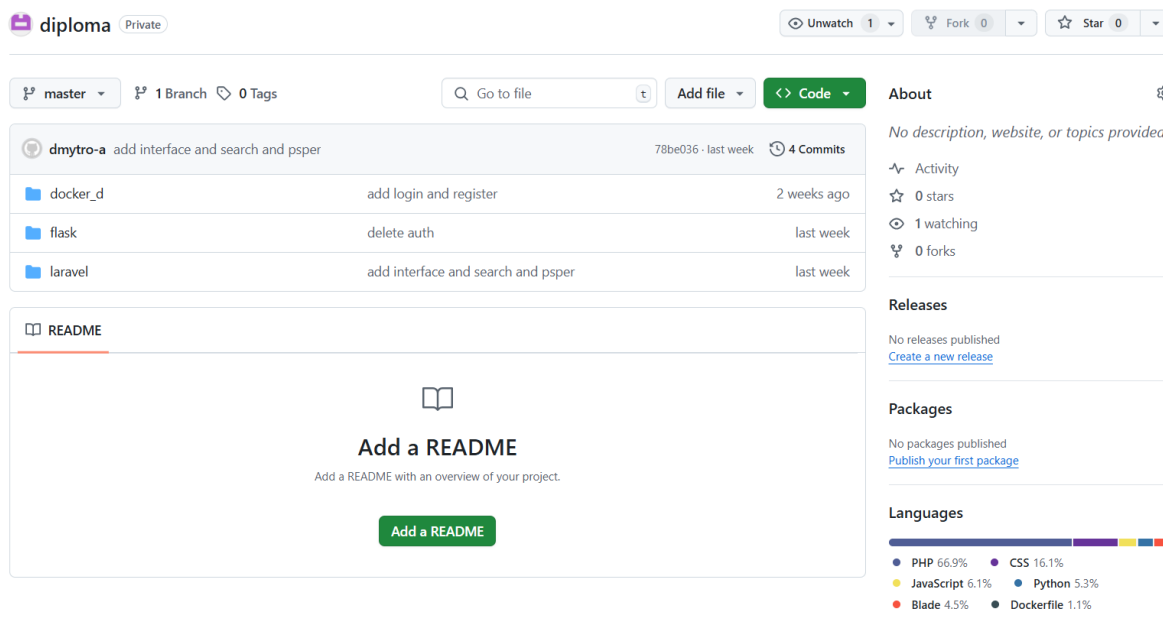


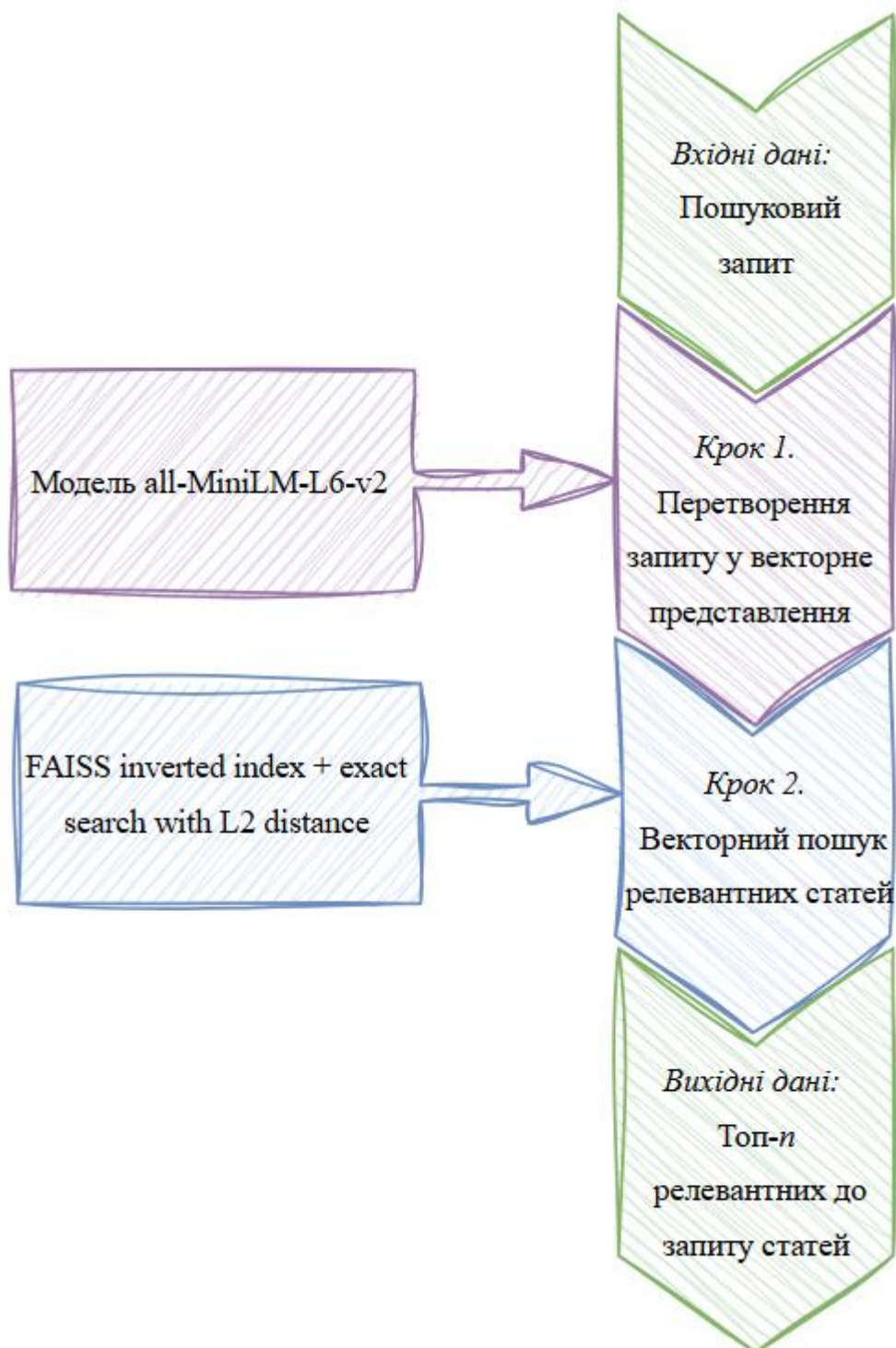
Рисунок А.1 – Головна сторінка репозиторію

Структура репозиторію наступна:

- директорія `docker_d` містить код для збірки всіх сервісів.
- директорія `flask` містить код який відповідає за адаптивну фільтрацію, зокрема це підключення до БД щоб працювати з текстовими даними та опрацьовувати їх ембедінг.
- директорія `laravel` містить код, який взаємодіє з користувачем. Це веб-додаток, який працює за запитом до сервісу `flask` щоб передати ключові слова і на базі цього запиту вивести рекомендації наукових статей.

Додаток Б

Схема методу адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання



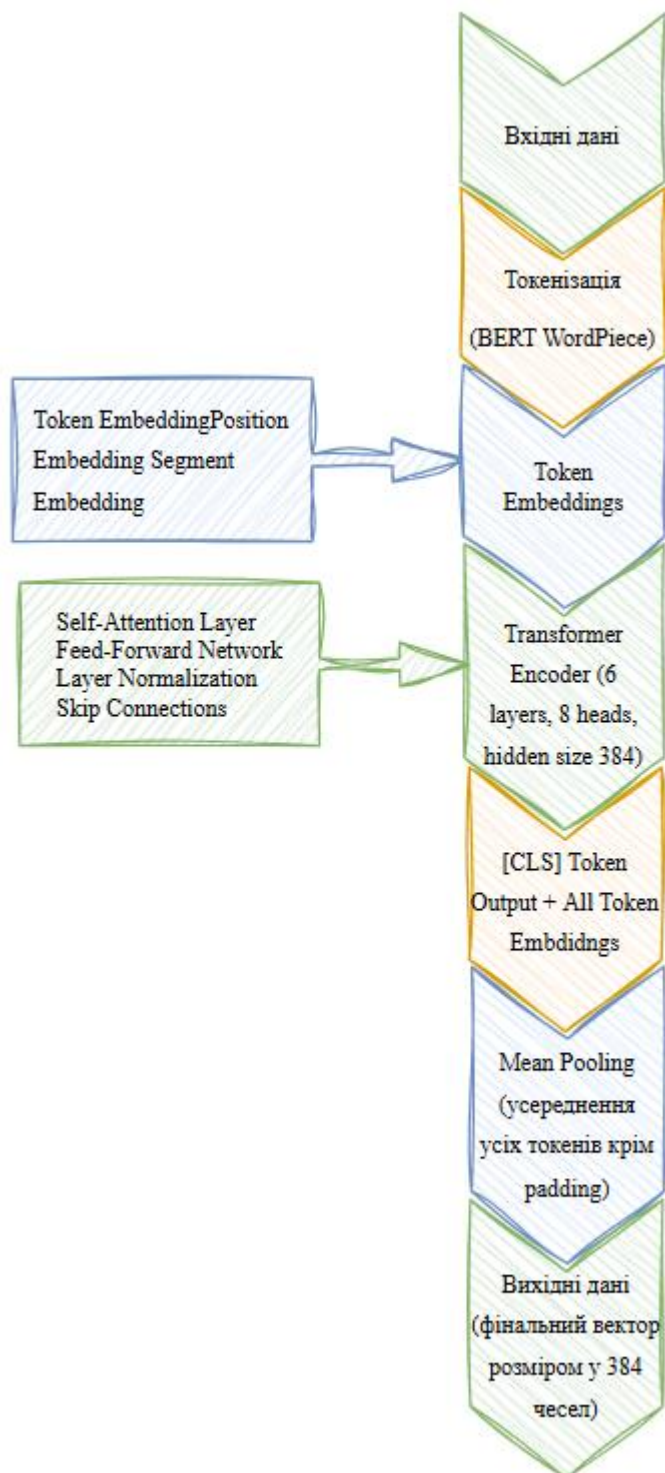
Додаток В

Архітектура інтелектуальної системи адаптивної фільтрації наукових статей





Додаток Г

Схема архітектури моделі all-MiniLM-L6-v2



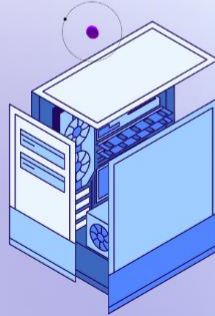
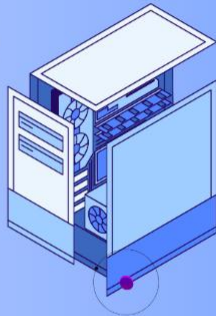
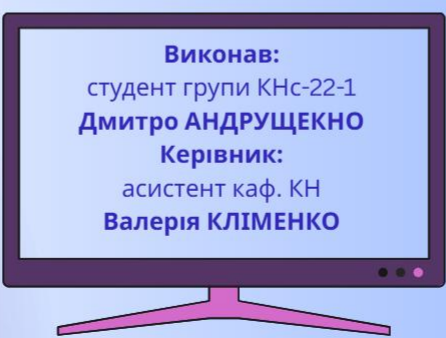
Додаток Д

Презентаційний матеріал





КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

МЕТОД АДАПТИВНОЇ ФІЛЬТРАЦІЇ НАУКОВИХ СТАТЕЙ ДЛЯ ПЕРСОНІФІКОВАНИХ РЕКОМЕНДАЦІЙ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

Виконав:
студент групи КНС-22-1
Дмитро АНДРУЩЕКО
Керівник:
асистент каф. КН
Валерія КЛІМЕНКО



АКТУАЛЬНІСТЬ

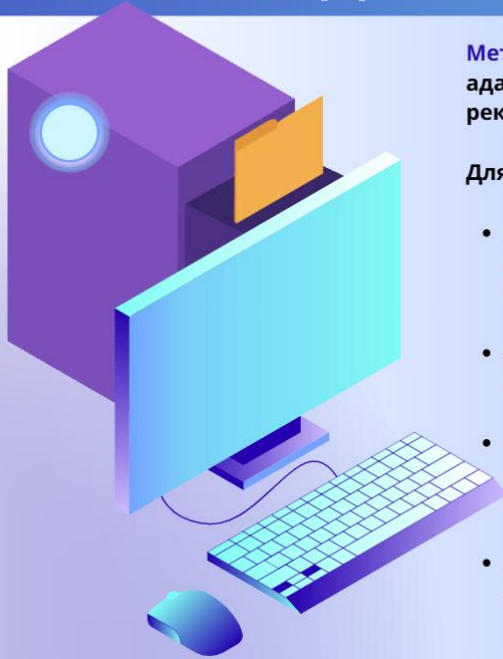
Актуальність дослідження, присвяченого методам адаптивної фільтрації наукових статей для персоналізованих рекомендацій, зумовлена як безпрецедентним зростанням обсягів наукової інформації, так і зміною способів її споживання в умовах цифровізації наукового простору.

У сучасному інформаційному середовищі науковці стикаються з проблемою інформаційного перевантаження, коли кількість доступних публікацій значно перевищує можливості їх осмислення та аналізу. Ця ситуація створює суттєві виклики для формування релевантного знання, ускладнюючи процес виявлення цінних і змістовних джерел, необхідних для проведення досліджень.

Дослідження у сфері адаптивної фільтрації наукових статей є не лише відповіддю на потреби сучасної наукової спільноти, але й важливим внеском у розвиток інтелектуальних інформаційних систем, здатних підтримувати ефективну наукову комунікацію, сприяти формуванню міждисциплінарних зв'язків та пришвидшувати генерацію нових знань.

Таким чином, вдосконалення алгоритмів персоналізованої рекомендації наукових статей має суттєве практичне значення для підвищення ефективності наукової діяльності та доступності знань у сучасному інформаційному середовищі.

МЕТА І ЗАДАЧІ РОБОТИ



Метою кваліфікаційної роботи бакалавра є покращення процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання.

Для досягнення поставленої мети слід вирішити наступні завдання:

- виконати аналіз інформаційних моделей області автоматизованої адаптивної фільтрації наукових статей для персоніфікованих рекомендацій;
- створити метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання;
- виконати прикладну програмну реалізацію інтелектуальної системи адаптивної фільтрації наукових статей для персоніфікованих рекомендацій;
- виконати дослідження ефективності розробленого методу з використанням прикладної програмної реалізації.

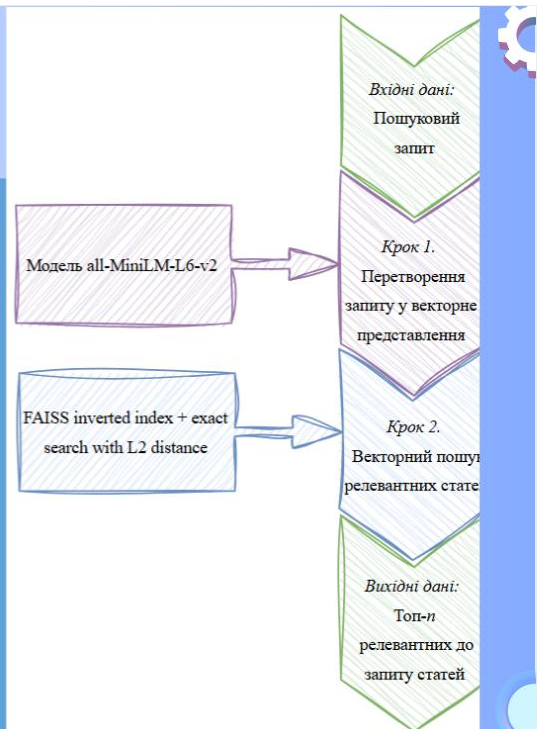
СХЕМА ТА КРОКИ МЕТОДУ АДАПТИВНОЇ ФІЛЬТРАЦІЇ НАУКОВИХ СТАТЕЙ ЗАСОБАМИ МАШИНОГО НАВЧАННЯ

Вхідними даними методу є пошуковий запит користувача, який формується на основі його інформаційних потреб

На першому кроці здійснюється перетворення запиту у векторне представлення за допомогою моделі «all-MiniLM-L6-v2», яка є результатом попереднього тренування з використанням знань більших трансформерних архітектур, у межах підходу «knowledge distillation», що дозволяє зберегти семантичну інформативність при значно меншій кількості параметрів.

На другому кроці здійснюється векторний пошук релевантних наукових статей. Для цього використовується бібліотека FAISS, яка дозволяє ефективно виконувати пошук найближчих сусідів у великих векторних просторах.

Вихідними даними є топ-*n* найбільш релевантних до запиту наукових статей, упорядкованих за ступенем схожості.



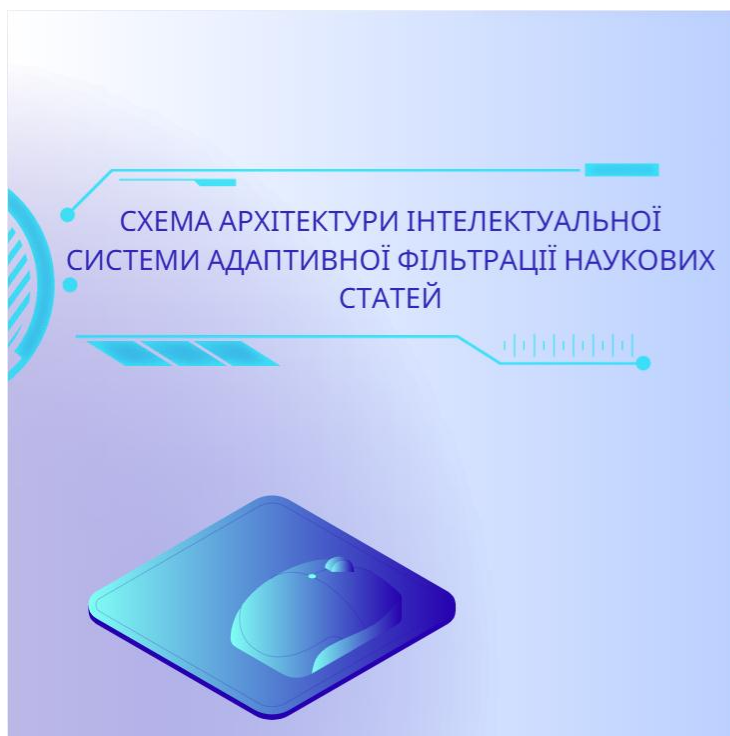


СХЕМА АРХІТЕКТУРИ МОДЕЛІ ALL-MINILM-L6-V2

Початковим етапом обробки є токенизація тексту за допомогою токенизатора WordPiece, який використовується в архітектурі BERT. Він розбиває речення на послідовність токенів, зрозумілих моделі. На початку та в кінці додаються спеціальні токени CLS і SEP.

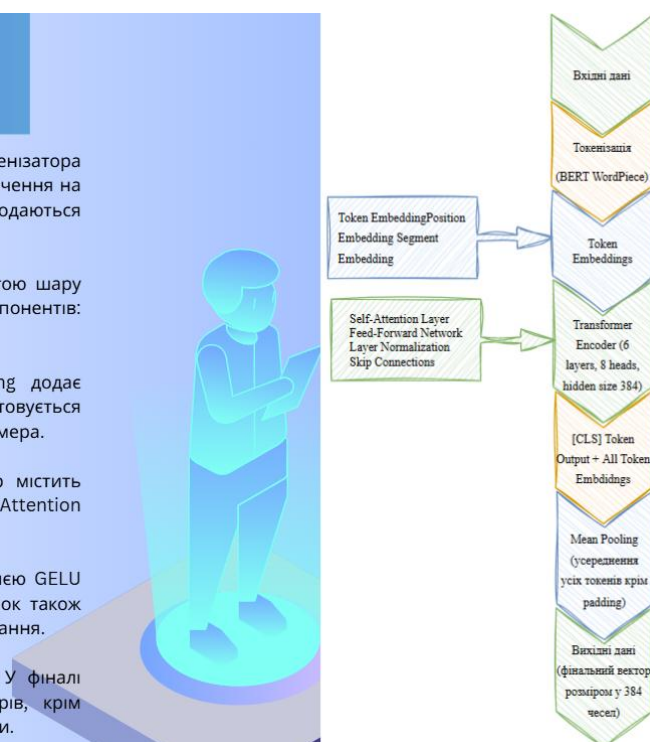
Після токенизації кожен токен перетворюється на вектор за допомогою шару ембедінгів. Векторне представлення формується як сума трьох компонентів: Token Embedding, Position Embedding та Segment Embedding.

Token Embedding відображає значення слова, Position Embedding додає інформацію про порядок слів у реченні, а Segment Embedding використовується для розділення речень у парі. Разом вони формують вхід для трансформера.

У моделі all-MiniLM-L6-v2 є 6 трансформерних шарів. Кожен шар містить механізм Self-Attention (з 8 головами) і Feed-Forward Network. Self-Attention аналізує зв'язки між токенами в контексті речення.

Після уваги токени проходять через повнозв'язну мережу з активацією GELU або ReLU, що допомагає моделі виділяти важливі ознаки. Кожен блок також включає залишкові зв'язки та Layer Normalization для стабільного навчання.

На виході кожного шару модель уточнює представлення токенів. У фіналі обчислюється середнє значення (mean pooling) всіх токен-векторів, крім службових. Це і є sentence embedding — вектор довжиною 384 елементи.



НАБІР ДАНИХ ДОСЛІДЖЕННЯ

Для реалізації методу адаптивної фільтрації було обрано датасет Research Papers Dataset, що налічує близько 500 тис. записів, з яких було обрано 5000, що містять наукові статті.

# abstract	# authors	# citation	# references	# title	# venue	# year	# id
[null] 17%	865804 unique values	0 42.5k	[null] 12%	999064 unique values	[null] 18%	1937 2017	1000000 unique values
Quality and Reliabi... 0%			Other (819188) 82%	International confe... 1%	Other (810413) 87%		
Other (827395) 83%							
We studied an autoassociative neural network with dynamic synapses which include a facilitating mech...	['Joquin J. Torres', 'Jesus M. Cortes', 'Joquin Marro', 'Hilbert J. Kappen']	58	['4817c9d2-9845-4a42-a85b-8a5523727c5', 'b1187381-c024-471e-8288-56a9c3d3ca9', 'ecf8e778-ca06-47ea...']	Attractor neural networks with activity-dependent synapses: The role of synaptic facilitation	Neurocomputing	2007	4ab3772a-e777-46f7-a6d2-16984f81c1db
It is well-known that Sturmian sequences are the non ultimately periodic sequences that are balanced...	['Genevieve Paquin', 'Laurent Vuillon']	58	['1c655ae2-867d-4bc4-b8cc-8c779e9a7f18', '24e457ca-d788-4bc7-57a5-2a33aca2f574', '588ea83a-835b-454d...']	A characterization of balanced episturmian sequences	Electronic Journal of Combinatorics	2007	4ab34c4f-1d96-4ce5-ab6f-b3e19f2d68de
One of the fundamental challenges of recognizing actions is accounting for the variability that arises...	['Yasser Sheikh', 'Munazzar Shah', 'Mubarak Shah']	221	['856116c1-9a7a-4f9b-a918-44a0199a7598', '85a452a1-dfad-4c32-9a54-6c4e58cfca5', '8c2f71cc-6797-445a...']	Exploring the space of a human action	International conference on computer vision	2005	4ab3a58c-3628-47ec-b578-884ecf4a6206
This paper generalizes previous optimal upper bounds on the minimum Euclidean distance for phase shi...	['Efraim Lohman', 'Hakan Lernerstad', 'Magnus Nilsson']	8	['91a79508-8cb3-405c-998f-29c36750b435', '58c0c0b-0814-4c88-8728-e23a5d48d718', '71a4d872-4b37-489e...']	Generalized upper bounds on the minimum distance of PCK block codes	Isis Journal of Mathematical Control and Information	2015	4ab35585-82b4-4207-9188-b80ce7a27c4e
Queueing networks with multiple classes of customers	['Simoneeta Balasoo', 'Gian-Luca Del Rossi', 'Andrea']	6	['1c264228-57d2-4b0c-b4c9-8d5851c17fac']	Applying BCP multi-class queueing networks for the	International Journal of Computer Aided Engineering	2015	4ab3a768-78c9-4407-888e-9e954cb5f2e4

ПРОГРАМНА СТРУКТУРА ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ

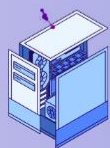


Веб-застосунок для персоналізованого пошуку наукових статей.

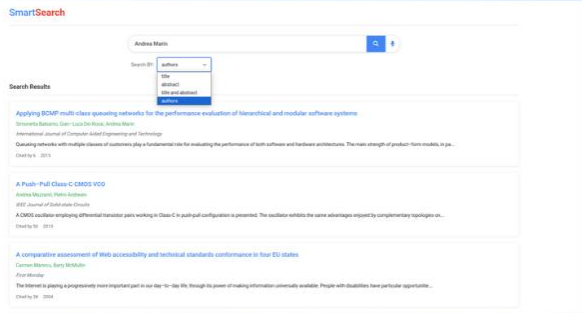
Користувач взаємодіє з клієнтським сервісом, що відображає інтерфейс, обробляє запити та передає їх до сервісу пошуку. Пошук здійснюється за ключовими словами або голосом.

Сервіс пошуку використовує FAISS для перетворення запиту у вектор, порівнює його з векторами статей у бази даних PostgreSQL та повертає релевантні результати.

Уся інформація зберігається в єдиній БД, з якою працюють обидва сервіси. Система забезпечує персоналізований, гнучкий та масштабований пошук наукових статей.



ІНТЕЛЕКТУАЛЬНА СИСТЕМА РОЗУМНОЇ ФІЛЬТРАЦІЇ НАУКОВИХ СТАТЕЙ



Детальна інформація про вибрану наукову статтю

Інтерфейс інтелектуальної системи дозволяє зручно обирати критерії фільтрації для пошуку, а також переглядати топ-5 найбільш релевантних наукових статей.



ПЛАН ПРОВЕДЕННЯ ДОСЛІДЖЕННЯ



Для перевірки ефективності методу адаптивної фільтрації було проведено експертне дослідження.

Два фахівці з комп'ютерних наук оцінили якість рекомендацій за п'ятибальною шкалою.

Результати порівнювались із рекомендаціями Google Scholar.

Оцінювання враховувало релевантність, персоналізацію, швидкість, зручність інтерфейсу та загальне задоволення користувача.

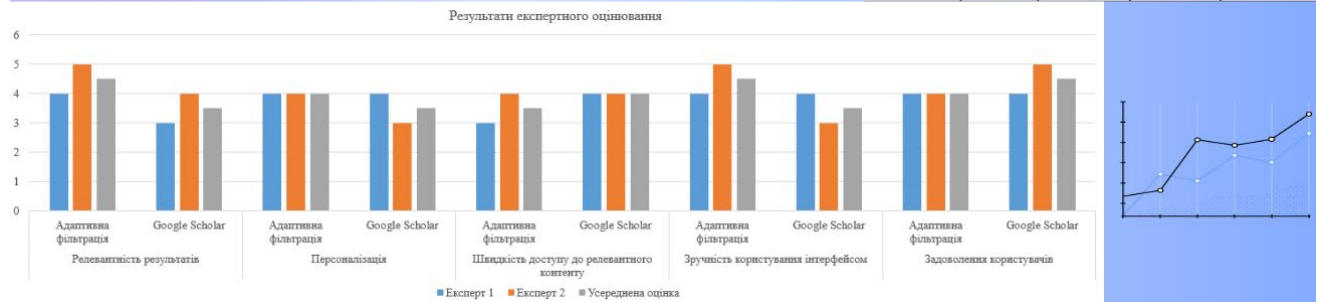
РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Таблиця демонструє порівняння двох підходів: метод адаптивної фільтрації та Google Scholar. За п'ятьма ключовими критеріями, такими як релевантність результатів, персоналізація, швидкість доступу до контенту, зручність інтерфейсу та задоволення користувачів.

Діаграма показує результати експертного оцінювання адаптивної фільтрації у порівнянні з Google Scholar.

Згідно з проведеним експертним оцінюванням, метод адаптивної фільтрації демонструє кращі результати у критеріях релевантності, персоналізації та зручності користування. Водночас Google Scholar показав вищі оцінки за критеріями швидкості доступу до релевантного контенту та рівня задоволення користувачів, що пояснюється його стабільністю, широкою базою даних та відомістю серед науковців.

Критерій	Підхід	Експерт 1	Експерт 2	Усереднена оцінка
Релевантність результатів	Адаптивна фільтрація	4	5	4,5
	Google Scholar	3	4	3,5
Персоналізація	Адаптивна фільтрація	4	4	4
	Google Scholar	4	3	3,5
Швидкість доступу до релевантного контенту	Адаптивна фільтрація	3	4	3,5
	Google Scholar	4	4	4
Зручність користування інтерфейсом	Адаптивна фільтрація	4	5	4,5
	Google Scholar	4	3	3,5
Задоволення користувачів	Адаптивна фільтрація	4	4	4
	Google Scholar	4	5	4,5



ВИСНОВКИ

Результатом виконання кваліфікаційної роботи бакалавра є досягнення поставленої мети, що полягала в покращенні процесу адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання.

Для досягнення поставленої мети були успішно виконані наступні задачі:

- проведено аналіз існуючих інформаційних моделей адаптивної фільтрації наукових статей, що дозволило визначити основні підходи, труднощі та потреби користувачів;
- розроблено метод адаптивної фільтрації наукових статей, що базується на машинному навчанні та враховує інтереси користувача;
- реалізовано прикладну програмну систему, яка поєднує функціональність збору, обробки, індексації та персоналізованого пошуку наукових текстів;
- проведено дослідження ефективності методу з використанням запропонованого веб-застосунку, зокрема шляхом порівняння з системою Google Scholar, що підтвердило переваги в релевантності та адаптивності рекомендацій.

Anti-Plagiarism (UA) v-15.281 Educational

The maximum coincidence with one document 2.0%

Dictionaries check: en_US, ru_RU, ua_UA. **Errors in the documents: 12%**

ID: 246614 Title: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання Added in a DB: 2025-06-17 Authors: Дмитро АНДРУЩЕНКО Heads: Валерія КЛІМЕНКО Consultants: Opponents:	Document		Sum coincidence on the DB	
	Symbols	Lexemes	Symbols	Lexemes
	63547	940	3982 (6%)	61 (6%)

Plagiarism sources

ID	Description	Plagiarism presence in the document	
		Symbols	Lexemes

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Дмитро АНДРУЩЕНКО

Співавтор:

Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод адаптивної фільтрації наукових статей для персоналізованих рекомендацій засобами машинного навчання

Науковий керівник: Валерія КЛІМЕНКО, асистент каф. КН

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1: 5.8%

Коефіцієнт подібності 2: 2.8%

Мікропробіли: 0

Заміна букв: 1

Інтервали: 0

Білі знаки: 190

Дата створення звіту: 2025-06-17 20:38:20.0

Після аналізу Звіту подібності констатую наступне:

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.

Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.

Обґрунтування:

2025-06-17

Дата

експерт

Петровська С.С.

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи Метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання.

Автор студент групи КНс-22-1 Дмитро Андрущенко

Освітня програма Комп'ютерні науки

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 – Комп'ютерні науки

Науковий керівник: асистент каф. комп'ютерних наук Валерія КЛІМЕНКО

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмними засобами комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	<i>відповідає</i>
1.2	Виявлені запозичення не є академічним плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	<i>відсутні</i>

Підтвердження:

Запозичення, виявлені в роботі Дмитра Андрущенко, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, які не мають авторства і містять поширені конструкції та загальновідомі терміни, скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

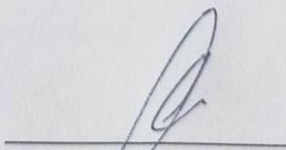
Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості:

- за системою Anti-Plagiarism: 2%;

- за системою StrikePlagiarism КП1: 5,8%, КП2: 2,8%.

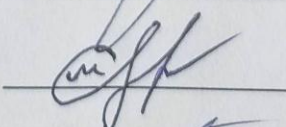
18.06.2025

Завідувач кафедри



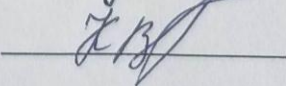
Олександр БАРМАК

Гарант освітньої програми



Олександр МАЗУРЕЦЬ

Керівник кваліфікаційної роботи



Валерія КЛІМЕНКО



**ВІДГУК НАУКОВОГО КЕРІВНИКА
на кваліфікаційну роботу бакалавра**

студента *гр. КНс-22-1 Андрущенко Дмитра Петровича*

за темою Метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання

1. Актуальність теми

У контексті зростаючого обсягу наукової інформації та цифровізації академічного середовища проблема ефективного доступу до релевантних джерел набуває особливої гостроти. Інформаційне перевантаження ускладнює процес формування наукової картини світу, що обумовлює необхідність у використанні адаптивних методів фільтрації. Персоніфіковані рекомендаційні системи, побудовані на основі машинного навчання, здатні динамічно враховувати інтереси користувача, забезпечуючи цільову доставку наукового контенту. Дослідження в цьому напрямі має прикладне значення для оптимізації наукової діяльності та розвитку інтелектуальних інформаційних технологій.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

За стандартом, а саме описом предметної області, об'єктом роботи є процес адаптивної фільтрації наукових статей для персоніфікованих рекомендацій. Метою роботи є покращення процесу адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання. При вирішенні поставленої задачі використано методи та технології машинного навчання адаптивної фільтрації наукових статей для персоніфікованих рекомендацій. Тому результати виконання кваліфікаційної роботи бакалавра відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

3. Професійні та особистісні якості бакалавра

У процесі виконання кваліфікаційної роботи бакалавра студент продемонстрував відповідальний підхід до поставлених завдань, належну дисциплінованість та здатність самостійно організовувати дослідницьку діяльність. На всіх етапах роботи – від аналізу літературних джерел до реалізації прикладного програмного забезпечення – здобувач виявив належний рівень професійної підготовки, відповідний освітній програмі за спеціальністю «Комп'ютерні науки». Також слід

відзначити сформовані комунікативні навички, ефективного управління часом, що свідчить про достатній рівень розвитку як фахових, так і міжособистісних компетентностей.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Результати, викладені у кваліфікаційній роботі, отримані студентом у результаті самостійного виконання усіх етапів дослідження без залучення сторонньої допомоги.

5. Ступінь оволодіння методами дослідження

У процесі виконання роботи здобувач продемонстрував належний рівень оволодіння сучасними методами наукового дослідження, а також практичне вміння застосовувати відповідні технології та інструменти в межах спеціальності.

6. Повнота та якість розкриття теми роботи

Обрана тема розкрита глибоко, послідовно та всебічно. У роботі надано обґрунтування актуальності дослідження, здійснено аналіз сучасних наукових підходів, чітко сформульовано завдання, які були успішно реалізовані, а також розроблено програмний засіб для перевірки ефективності запропонованого рішення щодо адаптивної фільтрації наукових статей.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Робота відзначається логічною структурою, послідовним викладенням матеріалу та належним рівнем мовної грамотності. Усі положення аргументовані й узгоджені з метою дослідження.

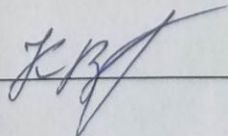
8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Розроблений у роботі метод та інтелектуальна система може бути використана науковцями для покращення якості наукових пошуків, підвищення релевантності рекомендованих матеріалів.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Керівник _____



асистент каф. КН Валерія КЛІМЕНКО



РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КНС-22-1 Андрущенко Дмитра Петровича*

за темою: Метод адаптивної фільтрації наукових статей для персоніфікованих рекомендацій засобами машинного навчання

1. Актуальність обраної теми

Сучасна наукова діяльність характеризується різноманіттям та масовістю доступних публікацій, що ускладнює оперативне виявлення релевантних матеріалів. У цьому контексті адаптивні системи фільтрації, які використовують алгоритми машинного навчання, виступають ефективним інструментом персоніфікованого доступу до знань. Їх застосування дозволяє зменшити вплив інформаційного перевантаження та підвищити якість дослідницької роботи. Актуальність теми зумовлена потребою у вдосконаленні механізмів наукової комунікації в умовах швидкої еволюції цифрового академічного середовища.

2. Повнота розкриття мети та завдань роботи

У межах кваліфікаційної роботи студент продемонстрував розуміння поставленої мети та завдань дослідження, сформулювавши їх чітко й аргументовано. Виконання дослідницької частини у відповідності до визначеної методології підтверджує належний рівень його наукової підготовки.

3. Зміст кожного розділу роботи

Для досягнення поставленої мети кожен розділ пояснювальної записки містить розгорнутий опис відповідних завдань, що забезпечує логічну послідовність дослідження. Перший розділ присвячений характеристиці предметної області, а саме проведенню аналізу моделей, методів та реалізацій. Другий розділ описує метод адаптивної фільтрації наукових статей засобами машинного навчання. У третьому розділі проведено експериментальне дослідження методу.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблена система забезпечує персоналізований підбір наукових матеріалів, що дозволяє користувачам ефективно орієнтуватися в великому інформаційному масиві та оптимізувати власну дослідницьку діяльність. Практична цінність системи полягає у

забезпеченні науковців релевантною інформацією з урахуванням індивідуальних інтересів, що підвищує якість академічного аналізу.

5. Якість оформлення кваліфікаційної роботи бакалавра

Кваліфікаційна робота оформлена відповідно до чинних вимог, що засвідчує належний рівень наукової культури автора. Структурна логіка викладу та стилістична узгодженість змісту свідчать про ґрунтовну підготовку студента до фахової діяльності.

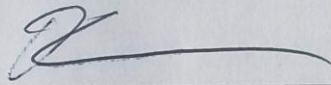
6. Недоліки кваліфікаційної роботи бакалавра

У деяких фрагментах записки застосовано складну термінологію без належного пояснення. У тексті трапляються неуніфіковані позначення одних і тих самих понять. У деяких частинах тексту спостерігається зміщення уваги з основної теми на другорядні аспекти. Зазначені недоліки мають несуттєвий характер і не знижують наукової цінності одержаних результатів.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Рецензент



проф. каф. ЗЛІС Сабенко О.С.