МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА

ГРИНЮК АРТЕМ ІГОРОВИЧ

Допускається до захисту:
в. о. завідувача кафедри
інформаційних технологій,
канд. техн. наук, доцент
_____ О. В. Зелінська
«_____»_____ 2024 р.

РЕКОМЕНДАЦІЙНА СИСТЕМА ДЛЯ АВТОМАТИЗОВАНОЇ ОБРОБКИ ТА АНАЛІЗУ ТЕКСТОВОГО КОНТЕНТУ

Спеціальність 122 «Комп’ютерні науки»

Кваліфікаційна (магістерська) робота

Науковий керівник:
Т. В. Січко, доцент кафедри
інформаційних технологій,
канд. техн. наук, доцент
_____
(підпис)

Оцінка: _____/_____/_____
(бали/за шкалою ЄКТС/за національною шкалою)
Голова ЕК: _____
(підпис)

Вінниця – 2024

АНОТАЦІЯ

Гринюк А. І. Рекомендаційна система для автоматизованої обробки та аналізу текстового контенту. Спеціальність 122 «Комп’ютерні науки», освітня програма «Комп’ютерні технології обробки даних (Data Science)», Донецький національний університет імені Василя Стуса, Вінниця, 2024.

У кваліфікаційній роботі проаналізовано використання технології рекомендаційних систем для аналізу текстового контенту. Висвітлено функціональні можливості системи та її переваги над традиційними методами обробки та аналізу даних. Доведено, що за допомогою рекомендаційної системи можна значно покращити процеси обробки та аналізу текстового контенту.

Ключові слова: рекомендаційна система, текстовий контент, аналіз даних, обробка інформації.

Табл. 22. Рис. 12. Бібліограф.: 50 найм.

ABSTRACT

Hryniuk A. I. Recommender System for Automated Processing and Analysis of Text Content. Speciality 122 «Computer science», educational program «Computer data processing technologies (Data Science)», Vasyl' Stus Donetsk National University, Vinnytsia, 2024.

The qualification work analyses the use of recommender system technology for analysing textual content. The functionality of the system and its advantages over traditional methods of data processing and analysis are highlighted. It is proved that the recommender system can significantly improve the processes of processing and analysing textual content.

Keywords: recommender system, textual content, data analysis, information processing.

Tabl. 22. Fig. 12. Bibliography: 50 items.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ 4

ВСТУП 5

РОЗДІЛ 1 ТЕОРЕТИЧНИЙ ОГЛЯД ПИТАНЬ РОЗРОБКИ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ДЛЯ АВТОМАТИЗОВАНОЇ ОБРОБКИ ТА АНАЛІЗУ ТЕКСТОВОГО КОНТЕНТУ 7

1.1 Автоматизація інформаційних процесів 7

1.2 Огляд процесів обробки та аналізу текстового контенту 11

1.3 Обґрунтування необхідності розробки рекомендаційної системи 14

1.4 Визначення наукової гіпотези дослідження 17

РОЗДІЛ 2 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ 20

2.1 Визначення та концепція рекомендаційної системи 20

2.2 Огляд методів, що використовуються у рекомендаційних системах 26

2.3 Опис механізмів роботи рекомендаційної системи 33

2.4 Опис програмно-апаратного інструментарію 36

2.5 Виклики та проблеми у розробці рекомендаційної системи 41

РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ 46

3.1 Загальна архітектура та принцип роботи системи 46

3.2 Структура бази даних 49

3.3 Розробка основного функціоналу 56

3.4 Розробка конфігурації 65

3.5 Оцінка ефективності розробленої рекомендаційної системи 68

ВИСНОВКИ 73

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ 74

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ЗМІ – Засоби масової інформації

ІАД – Інтелектуальний аналіз даних

ІС – Інформаційні системи

ООП – Об'єктно-орієнтоване програмування

СКРБД – Система керування реляційними базами даних

СУБД – Система управління базами даних

ШІ – Штучний інтелект

API (Application Programming Interface) – Інтерфейс програмування застосунків

GPT (Generative Pre-trained Transformer) – Генеративний попередньо навчений трансформер

GUI (Graphical User Interface) – Графічний інтерфейс користувача

HTTP (HyperText Transfer Protocol) – Протокол передачі гіпертексту

NLP (Natural Language Processing) – Обробка природної мови

ORM (Object-Relational Mapping) – Об'єктно-реляційне відображення

SQL (Structured Query Language) – Мова структурованих запитів

ВСТУП

Актуальність теми дослідження. З розвитком комп'ютерних технологій обробки даних та необхідністю подальшого його аналізу в різних сферах, зокрема в медіаіндустрії, виникла дедалі більша потреба у створенні рекомендаційної системи, яка б автоматизувала процеси обробки й аналізу та надавала рекомендації з покращення на основі отриманих результатів. Традиційна ручна обробка та неточний аналіз великих масивів тексту є занадто трудомісткими і не можуть впоратися з сучасними вимогами до швидкості й точності. На противагу застарілим методам використання рекомендаційної системи дозволить не лише автоматизувати ці процеси, але й забезпечити експертні рекомендації з покращення текстового контенту.

Метою кваліфікаційної роботи є розробка рекомендаційної системи, спрямованої на збір, автоматизовану обробку та аналіз текстового контенту.

Відповідно до поставленої мети сформульовано основні **завдання**:

1. Розглянути теоретичні аспекти процесів обробки та аналізу текстового контенту.
2. Провести огляд та аналіз наявних підходів до обробки та аналізу текстового контенту. Виявити переваги та недоліки різних методів.
3. Узагальнити технологічний процес створення рекомендаційної системи.
4. Обґрунтувати необхідність розробки рекомендаційної системи для медіаіндустрії.
5. Навести переваги рекомендаційної системи у порівнянні з традиційними підходами.
6. Виявити основні виклики та проблеми у розробці рекомендаційної системи.
7. Розробити рекомендаційну систему, яка відповідатиме усім заявленим вимогам.
8. Провести оцінку ефективності розробленої рекомендаційної системи.

Об’єкт дослідження — процеси автоматизованої обробки та аналізу текстового контенту з використанням технології рекомендаційних систем.

Предмет дослідження — методи та інструменти автоматизованої обробки та аналізу текстового контенту з метою отримання експертних рекомендацій.

Отримані результати мають як теоретичне, так і практичне значення. Теоретично, робота сприятиме розширенню розуміння процесів обробки та аналізу даних в медіаіндустрії та розвитку методів використання рекомендаційних систем для розв’язання проблем досліджуваної галузі. Вона зробить внесок до досліджень з автоматизації медіапроцесів та сприятиме подальшій розробці та вдосконаленню інструментів аналізу даних.

Практично, розроблена рекомендаційна система буде корисною для медіакомпаній, редакцій та аналітичних відділів, даючи змогу їм здійснювати автоматизовану обробку великих обсягів даних та швидкий аналіз текстового контенту.

Апробація результатів дослідження. Результати кваліфікаційної роботи було апробовано на Міжнародній науково-практичній конференції «Відновлення України у повоєнні часи: виклики, стратегічні пріоритети, ресурсне забезпечення, потенціал майбутнього розвитку», яка відбулася 10–11 жовтня 2024 року на базі Донецького національного університету імені Василя Стуса. Тези на тему: «Використання технології рекомендаційних систем для аналізу текстового контенту» опубліковано в електронному збірнику наукових праць.

Впровадження результатів кваліфікаційної роботи в практичну діяльність організації. Результати кваліфікаційної роботи були впроваджені в практичну діяльність ТОВ «РЕДАКЦІЯ ВІННИЦЬКОЇ РЕГІОНАЛЬНОЇ ГАЗЕТИ „ПОДІЛЬСЬКА ЗОРЯ“», що дозволило автоматизувати робочі процеси, підвищити ефективність діяльності редакції, а також покращити якість підготовки та випуску текстового контенту. Теоретичні напрацювання та результати впроваджено на практиці.

РОЗДІЛ 1
ТЕОРЕТИЧНИЙ ОГЛЯД ПИТАНЬ РОЗРОБКИ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ДЛЯ АВТОМАТИЗОВАНОЇ ОБРОБКИ ТА АНАЛІЗУ ТЕКСТОВОГО КОНТЕНТУ

1.1 Автоматизація інформаційних процесів

У сучасному світі, охопленому хвилею технологічного розвитку, важко уявити галузь, яка залишається поза впливом великих обсягів даних, або так званих Big Data. З допомогою високотехнологічних інструментів і методів обробки даних Big Data стає ключовим елементом у різних сферах – від бізнесу та медицини до науки і громадського управління [1].

Зростаючі обсяги інформації потребують апаратних і програмних засобів, здатних ефективно і швидко обробляти великі обсяги інформації, зі значним зниженням вартості збору, обробки, зберігання і передачі інформації. Ці процеси стали підґрунтям у новому та перспективному напряму розвитку інформаційних послуг – Big Data [2].

Big Data – це структуровані й неструктуровані дані величезних обсягів і значного розмаїття, що піддаються ефективній обробці програмними інструментами, які горизонтально масштабуються. Основні принципи роботи з великими даними [3]:

- горизонтальна масштабованість;
- відмовостійкість;
- локальність даних.

Сучасними програмними інструментами роботи з великими даними є R, Python, комплексний підхід Business Intelligence, реляційні системи управління базами даних із підтримкою мови SQL, а також NoSQL бази даних.

Big Data працює за таким принципом: чим більшою кількістю інформації володіє людина, тим точніший прогноз вона може зробити. Також можливість порівняння певних даних та взаємозв'язків між ними дозволяє знайти

закономірності, які були приховані до цього. Усе це забезпечує глибинне розуміння проблем та, в кінцевому результаті, дозволяє знайти рішення або можливості керування потрібними процесами [4]. Завдяки Big Data відкриваються нові можливості для обробки та аналізу інформації, яка раніше була недоступна через обмежені обчислювальні ресурси та недостатню кількість даних.

Сучасний світ значною мірою залежить від комп'ютерних технологій, які стали невіддільною частиною майже всіх сфер людської діяльності. Завдяки впровадженню Big Data у різні сфери діяльності, комп'ютерні технології набули нового значення та функціонального навантаження, адже вони дають змогу ефективно збирати, обробляти й аналізувати інформацію у величезних обсягах. Це особливо актуально в епоху цифрової інформатизації, коли дані стають ключовим ресурсом, що підтримує прийняття рішень на всіх рівнях. Інформатизація охопила різні галузі, зокрема й засоби масової інформації (ЗМІ), де комп'ютерні технології відіграють важливу роль у зборі, обробці, аналізі, поширенні та зберіганні інформації.

З кожним днем обсяг інформації, що потребує обробки та аналізу, невпинно зростає, стимулюючи розвиток нових технологій, спрямованих на полегшення цих процесів. З розвитком інформаційних технологій кількість даних зростає в геометричній прогресії. Швидке поширення обчислювальної техніки, поява нових і розвиток уже наявних програмних засобів дають змогу вирішити безліч питань, пов’язаних з автоматизацією обробки та аналізу текстового контенту.

Зростання обсягу інформації й розвиток технологій привели до необхідності автоматизації процесів, які раніше вимагали значних людських ресурсів. Автоматизація стає ключовим елементом у роботі з великими масивами даних, адже завдяки їй можна значно підвищити ефективність і швидкість обробки інформації.

Під поняттям «автоматизація» науковці розуміють використання технічного і програмного забезпечення, яке частково, або повністю звільняє

людину від участі в багатьох робочих процесах. Отримання, зберігання, перетворення, передача та використання інформації здійснюється за допомогою комп'ютеризованих систем, які дають змогу швидко досягнути запланованих результатів, ефективно структурувати та систематизувати великі обсяги даних. До переваг автоматизації можна віднести:

- підвищення точності та якості виконання операцій;
- скорочення витрат на робочу силу;
- зменшення ймовірності людських помилок;
- підвищення якості контенту в медіа;
- збільшення продуктивності;

Автоматизація інформаційних процесів у медіа призводить до значного покращення ефективності редакційних операцій, зокрема до скорочення часу на повторювані завдання та підвищення точності контенту. Автоматизовані системи допомагають усунути людський фактор у рутинних завданнях, мінімізуючи ймовірність помилок та забезпечуючи високу якість публікацій.

До основних етапів, що потребують автоматизації в ЗМІ, належать такі завдання, як збір, обробка, фільтрація, аналіз і публікація. Проте автоматизація цих етапів у ЗМІ супроводжується викликами, зокрема необхідністю структурування тексту, що надходить з різних джерел. Різноманітність джерел, форматів та типів даних ускладнює завдання інтеграції інформації в єдину систему, що потребує використання адаптивних технологій та алгоритмів. Це підкреслює важливість розробки нових підходів до управління даними та їх аналізу.

Для подолання цих викликів важливо розробляти методи обробки даних, які здатні забезпечити ефективну роботу з великими масивами інформації. В умовах постійного зростання обсягів інформації потрібні інструменти, які не лише автоматизували процеси збору та аналізу, а й допомагали структуровано організувати дані для інтелектуальної обробки та аналізу.

У сучасному інформатизованому суспільстві удосконалення технологій збирання та зберігання даних призвело до накопичення величезних обсягів різноманітної інформації. Накопичені дані утворюють масиви неструктурованої інформації, які без спеціального оброблення не мають корисного використання, навпаки, ускладнюють процес пошуку дійсно потрібної інформації. Дані потребують особливого оброблення, інструменти якого мають бути простими у застосуванні, а результати – конкретними та зрозумілими. Одним із методів такого оброблення є інтелектуальний аналіз даних. Серед найбільш популярних алгоритмів ІАД можна виділити кластеризацію, класифікацію, асоціативний аналіз, регресію та методи прогнозування.

Сучасні методи інтелектуального аналізу даних (ІАД, Data Science) охоплюють такі задачі [5]:

- перетворення даних;
- зберігання даних;
- аналіз даних;
- отримання інформації для прийняття рішень.

Завдяки впровадженню інтелектуальної рекомендаційної системи, яка дає змогу об’єднати процес автоматизованої обробки та ІАД, ЗМІ може отримувати найкращі рекомендації з покращення свого текстового контенту. Можливості сучасних рекомендаційних систем для медіа:

- покращення релевантності матеріалів для користувачів;
- збільшення залученості аудиторії;
- підвищення рівня довіри до ЗМІ;
- інтеграція різних каналів отримання інформації.

Як показує аналіз науково-практичної літератури та емпіричних досліджень, більшість завдань автоматизованої обробки та аналізу текстового контенту є завданнями формування та оптимізації процесів створення, обробки та поширення інформаційного контенту, що дає змогу ЗМІ працювати більш ефективно та забезпечувати високу якість поданої інформації.

1.2 Огляд процесів обробки та аналізу текстового контенту

Згідно з численними дослідженнями, у сучасному медіасередовищі обробка та аналіз текстового контенту є ключовими елементами, що впливають на ефективність роботи ЗМІ. Нині більшість ЗМІ поєднують традиційні та автоматизовані методи обробки та аналізу текстових даних. Проте значна частина обробки контенту виконується вручну, що забезпечує високу якість матеріалу, але обмежує масштабованість. Для оптимізації своїх процесів організації активно запроваджують сучасні інструменти автоматизації, а також базові методи ІАД та штучного інтелекту.

За визначенням науковця О. А. Баранова, штучний інтелект – це певна сукупність методів, способів, засобів та технологій насамперед, комп'ютерних, що імітує (моделює) когнітивні функції, які мають критерії, характеристики та показники еквівалентні критеріям, характеристикам та показникам відповідних когнітивних функцій людини [6].

Одним із головних етапів ШІ є машинне навчання. Машинне навчання – це наука про розробку та застосування алгоритмів, які можуть покращувати свою роботу, аналізуючи попередні результати виконаної роботи. Завдяки машинному навчанню можна оптимізувати процеси, пов'язані з аналізом тексту, виявленням закономірностей та структуруванням даних. Зокрема, це стосується автоматизованих методів кластеризації та категоризації текстів.

Автоматизовані методи кластеризації та категоризації дають змогу автоматично розподілити тексти за тематичними групами, що зменшує навантаження на редакційний персонал. Впровадження кластеризації та категоризації особливо корисне для великих медіаорганізацій, де потік інформації є постійним і об'ємним.

Проте ці інструменти мають свої обмеження, оскільки не здатні забезпечити глибоку персоналізацію контенту та врахувати індивідуальні потреби читачів. Алгоритми кластеризації та автоматизації мають обмеження щодо персоналізації, особливо коли йдеться про контент зі специфічними стилістичними особливостями.

До того ж алгоритми кластеризації можуть іноді помилково розподілити матеріали, особливо коли тематика текстів є багатогранною та перетинається між різними категоріями. Методи машинного навчання, хоча й дають змогу краще адаптувати обробку до певних тематик, також постають перед труднощами при аналізі контенту, що має змістовні й стилістичні відмінності, які потребують глибшої персоналізації.

З метою оптимізації та підвищення якості обробки даних ЗМІ дедалі більше впроваджують цифрові інструменти, що полегшують роботу з матеріалом. Для наочнішого розуміння поточних методів обробки даних, що використовуються, а також їхніх переваг та недоліків, розглянемо порівняльну табл. 1.1, яка допоможе визначити, які аспекти потребують вдосконалення та впровадження нових технологій, зокрема щодо автоматизації аналізу, скорочення часу на рутинні завдання та підвищення точності отриманих даних.

Таблиця 1.1 – Методи обробки текстових даних

Метод	**Переваги**	**Недоліки**
Руна обробка	Висока точність і якість матеріалів, можливість глибокої персоналізації та індивідуального підходу.	Трудомісткий процес, залежність від людського фактора, обмежена масштабованість.
Автоматизовані інструменти	Швидкість обробки, можливість обробки великих обсягів даних, зменшення навантаження на персонал.	Відсутність глибокої персоналізації, можливість помилок у розподілі текстів за темами.
Інтелектуальний аналіз даних	Можливість виявлення прихованих закономірностей, полегшення аналізу великих масивів даних.	Потреба налаштування та розуміння специфіки контенту, складність впровадження.
Штучний інтелект	Адаптивність, здатність навчатися та покращувати якість обробки з часом.	Можливі помилки у класифікації складного контенту, відсутність гнучкості у персоналізації.

Таким чином, кожен із розглянутих методів обробки має свої сильні та слабкі сторони, а їх поєднання може забезпечити ефективнішу обробку текстових даних, що дає змогу покращити якість контенту, а також спростити аналіз великих обсягів інформації, забезпечуючи гнучкість і точність в процесі обробки.

У сучасних медіа та інформаційних системах аналіз текстового контенту має критичне значення, адже дозволяє глибше зрозуміти зміст, виявити приховані закономірності та багато іншого. На відміну від обробки, яка переважно стосується механічних і структурних змін тексту, аналіз фокусується на змістовній та семантичній частині. Крім того, аналіз текстового контенту забезпечує можливість автоматичного визначення тем та ключових понять.

Кожен з методів аналізу окремо дозволяє покращити певні аспекти аналізу текстового контенту, але їхнє комбіноване використання дозволяє досягти більш комплексного та глибшого розуміння даних, забезпечуючи вищу точність і релевантність результатів.

Нові методи обробки та аналізу текстових даних, що з'являються завдяки швидкому розвитку технологій ШІ, відіграють важливу роль у поліпшенні точності та швидкості обробки інформації. Завдяки впровадженню технологій ШІ, процес аналізу великих обсягів текстової інформації став менш залежним від людського фактора, зменшуючи ймовірність помилок та підвищуючи продуктивність. Використання сучасних алгоритмів, таких як трансформери та глибинні нейронні мережі, дозволяє системам краще розуміти контекст тексту та аналізувати його з урахуванням семантичних зв'язків.

Автоматизована обробка та аналіз текстового контенту у медіасередовищі стають ключовими інструментами для підвищення ефективності та релевантності інформації. Поєднання різних методів обробки та аналізу тексту дозволяє досягти найкращих результатів у покращенні якості контенту. Швидка динаміка розвитку інформаційних технологій вимагає високої швидкості адаптації та гнучкості алгоритмів для підтримки якості контенту і точності аналітики.

1.3 Обґрунтування необхідності розробки рекомендаційної системи

Зважаючи на обмеження поточних методів обробки та аналізу текстових даних, які не забезпечують достатнього рівня автоматизації в роботі з великими обсягами даних, виникла нагальна потреба у впровадженні більш просунутих технологій, зокрема систем рекомендаційного типу, які здатні персоналізувати обробку даних та адаптувати аналіз під конкретні потреби. Сучасні рішення в галузі ШІ та машинного навчання можуть суттєво полегшити дані процеси. Використання таких технологій дозволить ЗМІ та іншим організаціям залишатися конкурентоспроможними та швидше реагувати на зміни в інформаційному середовищі.

Зростання обсягу інформації та різноманіття тематики ускладнює процеси ручного сортування, обробки та аналізу даних, що вимагає залучення значних людських ресурсів і часу. Це, своєю чергою, створює необхідність у використанні автоматизованих систем, які здатні не лише пришвидшити процеси обробки інформації, але й підвищити точність та релевантність результатів.

Для розв’язання цієї проблеми доцільно розглянути можливість впровадження рекомендаційної системи, яка б поєднувала ШІ, методи машинного навчання, ІАД та обробку природної мови. Така автоматизована система дозволила б автоматично обробляти та аналізувати великі обсяги текстової інформації, надаючи релевантні рекомендації та висновки, що сприятиме підвищенню продуктивності та точності аналізу інформаційних потоків.

Кожен із розглянутих методів обробки має свої переваги та недоліки. Ручна обробка забезпечує високу точність і якість, але має обмежену масштабованість. Автоматизовані інструменти забезпечують швидкість обробки, але відсутність глибокої персоналізації. Машинне навчання та ШІ пропонують ефективність та автоматизацію, проте можуть мати обмежену персоналізацію та вимагати складної реалізації.

Зважаючи на переваги та недоліки різних методів обробки та аналізу текстового контенту, постає потреба в оптимізації процесів за допомогою

комплексних технологій. Важливим інструментом в цьому є інформаційні системи (ІС), які забезпечують автоматизацію зберігання, обробки та аналізу даних. ІС відіграють ключову роль в управлінні інформаційними потоками та сприяють покращенню якості контенту. Функціональні можливості ІС подано в табл. 1.2, які є важливими для автоматизованої обробки та аналізу даних.

Таблиця 1.2 – Функціональні можливості інформаційних систем

Функція	**Опис**
Надійне зберігання інформації	Сучасні інформаційні системи зазвичай використовують надійні бази даних, які дають змогу зберігати великі обсяги інформації та забезпечують захист від втрати даних.
Обробка інформації	Інформаційні системи часто включають інструменти для автоматизації обробки даних, що значно спрощує процес їхнього аналізу та обробки.
Зручний доступ для користувачів	Розробляються з урахуванням зручності для користувачів, забезпечуючи простий та швидкий доступ до інформації, яка їм потрібна.
Забезпечення інформаційної безпеки	Інформаційні системи повинні забезпечувати захист даних від несанкціонованого доступу, кібератак, втрати чи крадіжки.

Інформаційна система – це програмний комплекс, що забезпечує виконання таких функцій [7]:

- підтримка надійного зберігання інформації у пам’яті комп’ютера;
- виконання специфічних для цього додатка перетворень інформації та/або обчислень;
- надання користувачам зручного інтерфейсу, який легко опанувати.

Завдання інформаційних систем – накопичення інформації, якої потребує організація для забезпечення ефективного управління всіма своїми ресурсами, створення інформаційного і технічного середовища для здійснення управління організацією [8]. Основна роль інформаційних систем у діяльності організацій:

- підтримка збору, зберігання та обробки інформації;
- забезпечення управління ресурсами, процесами та прийняття рішень;
- полегшення керування діяльністю організації.

Для ефективної обробки та аналізу великих обсягів текстової інформації традиційних інформаційних систем може бути недостатньо. Саме тому сучасні інформаційні системи дедалі частіше включають використання рекомендаційних систем. Вони допомагають автоматизувати процеси обробки та аналізу текстового контенту, що значно підвищує ефективність роботи з великими обсягами даних. Такі системи не лише спрощують аналіз, але й роблять його точнішим. Завдяки впровадженню таких систем можливо забезпечити більш персоналізований підхід до роботи з інформацією.

Основні причини необхідності розробки рекомендаційної системи подано в табл. 1.3, де подано детальний опис основних причин, поділених на категорії.

Таблиця 1.3 – Основні причини розробки рекомендаційної системи

Причина	**Опис**
Автоматизована обробка	Використання системи дозволяє значно зменшити час і людські ресурси, необхідні для обробки великих обсягів текстових даних.
Підвищення якості контенту	Система рекомендує релевантний та персоналізований контент, що допомагає підвищити якість і релевантність матеріалів..
Зменшення інформаційного навантаження	Завдяки рекомендаційній системі користувачі отримують лише релевантний контент, що знижує надмірне інформаційне навантаження та полегшує доступ до необхідної інформації.
Оптимізація інформаційних процесів	Рекомендаційна система допомагає організувати та структурувати потоки даних, що робить інформаційні процеси більш прозорими й керованими, мінімізуючи можливість втрати важливої інформації.

Всі ці причини є ключовими для розуміння необхідності розробки рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту в медіаіндустрії, оскільки це забезпечить оптимізацію інформаційних процесів, підвищить якість контенту та зменшить інформаційне навантаження. Такі рішення дають змогу зосередити ресурси на творчих і стратегічних аспектах медіабізнесу, що є критично важливим у сучасних умовах динамічного розвитку галузі.

1.4 Визначення наукової гіпотези дослідження

Гіпотеза є важливим елементом наукового дослідження, оскільки вона формулює припущення про можливі зв'язки між різними змінними або явищами, що підлягають дослідженню. Гіпотеза – це наукове припущення, висунуте для пояснення будь-яких явищ, процесів або причин, які зумовлюють певний наслідок. Гіпотеза, як структурований елемент науки, є спробою на основі узагальнення вже наявних знань вийти за його межі, тобто сформулювати нові наукові положення, достовірність яких потрібно довести [9]. Таким чином, вона задає новий вектор для науки, заохочуючи розширення розуміння відомих процесів.

Гіпотеза потребує перевірки через збір емпіричних даних, що є необхідною умовою для доведення її правдивості. Це положення підкреслює важливість експериментальної або аналітичної фази дослідження. Без належної перевірки гіпотеза залишатиметься лише теоретичним припущенням, яке не матиме наукової ваги. Вона має дві основні функції:

1. Організаційна: допомагає структурувати дослідження, визначаючи напрямок і об'єкти дослідження.
2. Прогнозуюча: забезпечує передбачення можливих результатів на основі теоретичної бази.

Формулювання гіпотези допомагає визначити мету та сфокусуватися на основних питаннях дослідження, виявляючи потенційні зв’язки між змінними. Формулювання гіпотези є критичним для вибору методології дослідження, оскільки визначає тип аналізу даних, методи збору інформації та способи перевірки результатів. Таким чином, гіпотеза не лише спрямовує дослідження, але й впливає на всі його етапи – від розробки до інтерпретації результатів. Гіпотеза визначає структуру дослідження, допомагаючи обрати оптимальні методи для досягнення цілей дослідження.

На основі цього аналізу сформульовано гіпотезу, яка підлягає подальшій емпіричній перевірці в межах цього дослідження. Основна гіпотеза цього дослідження полягає в тому, що впровадження комп’ютерних технологій

обробки даних, зокрема рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту, дасть змогу швидше обробляти великі обсяги тексту з високою точністю та надавати експертні рекомендації, які відповідатимуть найкращим стандартам медіаіндустрії.

У цьому дослідженні увага зосереджена на специфічних особливостях обробки та аналізу текстової інформації, де складність контенту та його різноманітність вимагають особливих підходів. Ця гіпотеза відображає прагнення знайти більш ефективний шлях для розв'язання наявних проблем у галузі обробки інформації. Її формування ґрунтується на проведеному аналізі поточних методів обробки та аналізу текстових даних який виявив низку обмежень, зокрема низьку точність класифікації, недостатню релевантність виділеної інформації та тривалість часу, необхідного для обробки великих обсягів тексту. Впровадження рекомендаційної системи передбачається як рішення, яке здатне підвищити ефективність основних процесів, а саме обробки та аналізу.

Сучасні технології аналізу тексту стали важливим інструментом для обробки великих обсягів інформації, особливо в умовах чимраз більших обсягів даних. Проте, традиційні методи обробки та аналізу тексту стикаються з низкою проблем, які знижують їх ефективність та продуктивність. Це стає помітним при спробах застосування цих методів у сферах, де важлива швидкість, точність та персоналізація обробки інформації, як-от у медіаіндустрії, маркетингу, освіті та інших галузях. Одним з ключових викликів є здатність автоматично розуміти контекст та приховані змісти, що часто впливають на повноту і точність аналізу тексту. Зокрема, до основних недоліків традиційних методів обробки та аналізу тексту належать такі аспекти:

- трудомісткість процесу;
- затратність обробки;
- низька ефективність через брак автоматизації.

Водночас для побудови сучасної рекомендаційної системи необхідно враховувати низку факторів, що впливають на її ефективність. Це особливо важливо для досягнення високого рівня релевантності рекомендацій, що потребують точного аналізу текстових даних і налаштування системи на зміни у вподобаннях користувачів. Основні фактори, які слід враховувати при розробці ефективної рекомендаційної системи, включають:

- обраний алгоритм аналізу тексту;
- якість вхідних даних, що використовуються для навчання моделі;
- кількість вхідних даних;
- репрезентативність вибірки (урахування стилістичних особливостей різних джерел тексту).

Впровадження рекомендаційної системи пропонує широкі можливості для оптимізації роботи з великими обсягами текстової інформації. Це стає особливо цінним для медіаіндустрії, де швидкість і точність обробки контенту мають велике значення, а також для інших галузей, які працюють з великими масивами текстових даних.

Підсумовуючи визначення поставленої наукової гіпотези, можна стверджувати, що впровадження рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту потенційно сприятиме оптимізації даних процесів, підвищенню точності аналізу та релевантності результатів. Дане дослідження відкриває нові можливості в розробці інтелектуальних систем, здатних адаптуватися до різноманітних текстових даних. З урахуванням визначених факторів, таких як якість вхідних даних і вибір методів аналізу, подальше емпіричне дослідження дозволить перевірити обґрунтованість гіпотези та оцінити реальний вплив автоматизації на ефективність аналізу великих обсягів тексту.

РОЗДІЛ 2
АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ

2.1 Визначення та концепція рекомендаційної системи

У сучасному світі обсяги інформації зростають експоненційно, і користувачі стикаються з величезною кількістю даних, з яких слід вибрати найкориснішу та релевантну інформацію. Це призводить до потреби у вдосконаленні методів обробки та аналізу даних, щоб уникнути інформаційного перевантаження, що негативно впливає на процес прийняття рішень. Дані – це складова частина інформації, що являє собою зареєстровані сигнали. Під час інформаційного процесу дані перетворюються з одного виду в інший за допомогою методів. Обробка даних включає багато різних операцій [10]. Для якісного використання таких даних необхідно мати ефективні механізми обробки, щоб унеможливити хаотичне накопичення інформації. Однією з таких операцій в обробці даних є фільтрація. До основних функцій фільтрації відносять:

- видалення зайвих даних, які не потрібні для ухвалення рішень;
- фокусування уваги користувача на значущій інформації;
- відсівання зайвої або малозначущої інформації.

Інформаційна система, виступаючи ефективним засобом керування даними, забезпечує автоматизовану обробку, зберігання та доступ до інформації, що дозволяє оптимізувати інформаційні потоки та сприяє точному прийняттю рішень. Але в умовах стрімкого зростання обсягів даних, що надходять, виникає потреба у додаткових методах обробки, таких як фільтрація. Фільтрація даних допомагає відокремити релевантну інформацію від непотрібної, що зменшує навантаження на аналітичні системи та підвищує їхню продуктивність. Це особливо важливо для мінімізації інформаційного шуму, що сприяє оптимізації подальшого аналізу та підвищенню ефективності прийняття рішень.

Фільтрація даних є лише одним із кроків на шляху до оптимізації процесу взаємодії з інформацією, але для максимально ефективної роботи з великими обсягами даних необхідне використання складнішої системи, а саме інтелектуальної. Інтелектуальна система – це один з видів автоматизованих інформаційних систем, які ґрунтуються на знаннях. Інтелектуальна система є комплексом програмних, лінгвістичних і логіко-математичних засобів для реалізації основного завдання: здійснення підтримки діяльності людини і пошуку інформації в режимі розширеного діалогу природною мовою [11]. Інтелектуальна система має наступні особливості:

- фільтрація інформації;
- аналіз даних та розпізнавання закономірностей;
- надання контекстуально значущих рекомендацій.

Інтелектуальна система включає самонавчання на основі накопичених даних і автоматичне вдосконалення алгоритмів для підвищення точності результатів. Така система забезпечує інтеграцію з іншими інформаційними системами для отримання актуальної інформації, а також адаптацію до змін у запитах і потребах користувача.

Однією з найбільш поширених видів інтелектуальної системи є рекомендаційна. Рекомендаційна система – це інтелектуальна система, яка сприяє формуванню пропозицій, на основі аналізу релевантної інформації, отриманої з віртуального сховища даних, що консолідує велику кількість джерел [12]. Рекомендаційна система – це інтелектуальна система, здатна на основі інформації про користувачів і предметів рекомендувати предмети, що підходять користувачам. Основна мета рекомендаційної системи – збільшення доходів продавця коштом збільшення обсягу продажів рекомендованих предметів [13]. Важливою проблемою, яку вирішує рекомендаційна система, є показ контенту, який із максимальною ймовірністю зацікавить користувача.

Сучасні рекомендаційні системи широко використовують технології глибинного навчання, що є одним із ключових підходів машинного навчання.

Глибинне навчання базується на багатошарових нейронних мережах, які здатні автоматично витягувати значущі ознаки з великих обсягів даних. Глибинні нейронні мережі, що складаються з десятків або навіть сотень шарів, здатні обробляти великі обсяги даних і знаходити складні взаємозв'язки між елементами, що покращує якість прогнозів і релевантність рекомендацій.

Рекомендаційна система здебільшого використовує методи глибинного навчання для підвищення якості рекомендацій. Глибинне навчання – це розділ машинного навчання, спрямований на побудову ієрархічних моделей шляхом використання високорівневих масових абстракцій даних на основі глибинного графа з багатьма обробними шарами, які здійснюють лінійні або нелінійні перетворення, тобто це – прогресуючий високорівневий витяг ознак з сирих (необроблених) первісних вхідних даних [14].

Глибинне навчання є напрямком ШІ, який зосереджений на розробці та використанні нейронних мереж, особливо багатошарових (глибоких) нейромереж. Глибинне навчання відрізняється від традиційних методів програмування тим, що не використовує жорстко прописані правила, а, навпаки, ґрунтується на здатності моделей самостійно вивчати закономірності та адаптуватися до вхідних даних. Це дозволяє нейронним мережам розв'язувати складні завдання, такі як обробка тексту, аналіз, прогнозування тощо, без необхідності ручного програмування кожного конкретного правила.

Для навчання глибоких нейронних мереж зазвичай використовують великі обсяги даних і високі обчислювальні потужності, що робить цей підхід затратним, але ефективним для багатьох завдань. Нейронні мережі зазвичай складаються з одного або кількох шарів нейронів, але не мають великої кількості прихованих шарів. Вони добре працюють для простих завдань, але обмежені в обробці складних даних. Однією з переваг нейронних мереж є їхня гнучкість. Змінюючи кількість шарів і нейронів, можна адаптувати мережу до різних задач і обсягів даних. Це дозволяє налаштувати модель під конкретні потреби. Нейронні мережі можуть бути недетермінованими, тобто при повторному

навчанні можуть давати дещо інші результати залежно від випадкової ініціалізації ваг та інших випадкових факторів у процесі навчання.

Технології глибинного навчання, зокрема глибинні нейронні мережі, значно підвищили ефективність рекомендаційних систем. Глибинні нейронні мережі – це мережі з багатьма шарами, призначені для виявлення складних взаємозв'язків у даних. Даний тип нейронних мереж, який має десятки шарів і може виконувати велику кількість операцій над даними, використовується для класифікації, апроксимації функцій та широкого спектра завдань аналізу даних [15]. Основні функції глибинних нейронних мереж подано в табл. 2.1.

Таблиця 2.1 – Основні функції глибинних нейронних мереж

Функція	**Опис**
Автоматичний витяг ознак з великих обсягів даних	Глибинні нейронні мережі здатні самостійно знаходити релевантні ознаки, що зменшує потребу у використанні ручного інженерінгу ознак.
Виявлення прихованих закономірностей	Глибинні нейронні мережі мають велику кількість шарів, які дають змогу їм аналізувати дані на багатьох рівнях абстракції, що робить їх ефективними для виявлення прихованих, складних закономірностей.
Зниження залежності від ручного інженерінгу ознак	Використання глибинного навчання дозволяє зменшити обсяг ручної роботи, оскільки моделі самостійно вивчають корисні ознаки, що прискорює процес створення нових моделей.

Завдяки наведеним функціям глибинні нейронні мережі значно розширили можливості аналізу даних та автоматизації процесу обробки інформації. Глибинні нейронні мережі здатні знаходити та вивчати складні, приховані закономірності та взаємозв'язки у великих обсягах даних завдяки використанню багатьох рівнів абстракції, що досягається за допомогою ієрархічної структури мережі, яка дозволяє обробляти дані на різних рівнях – від найпростіших ознак до високорівневих концепцій. Крім того, їхнє застосування сприяє підвищенню ефективності рішень, оскільки глибинні мережі можуть самостійно знаходити найкращі ознаки для аналізу, мінімізуючи потребу в ручному втручанні в процес

вибору ознак, що особливо важливо у випадках, коли дані є неоднорідними або мають складну структуру. Це робить їх надзвичайно корисними для завдань, де простіші алгоритми аналізу не працюють належним чином. Глибинні нейронні мережі добре працюють на великих наборах даних, оскільки можуть виявляти дрібні деталі та зв'язки, які могли б залишитися непоміченими при використанні менш складних моделей. В інформаційному світі рекомендаційні системи стрімко еволюціонують, і в їх розвитку можна виділити кілька ключових тенденцій, що представлені в табл. 2.2.

Таблиця 2.2 – Тенденції розвитку рекомендаційних систем

Тенденція	**Опис**
Збільшення використання глибинного навчання	Використання глибинних нейронних мереж для поліпшення якості рекомендацій завдяки виявленню прихованих закономірностей і автоматичному витягу ознак із великих обсягів даних.
Зменшення залежності від ручного інженерінгу ознак	Використання глибинного навчання дозволяє знизити потребу в ручному створенні ознак, що спрощує процес розробки рекомендаційних систем.
Оптимізація обробки великих обсягів даних	Використання алгоритмів для фільтрації та обробки даних з метою зменшення інформаційного перевантаження та забезпечення користувача лише значущою інформацією для прийняття рішень.
Підвищення ефективності рекомендацій	Завдяки застосуванню новітніх методів машинного навчання, рекомендаційні системи забезпечують більш точні рекомендації.

Всі ці ключові тенденції у розвитку рекомендаційних систем сприяють підвищенню якості взаємодії з великими обсягами інформації, забезпечуючи більш точне та персоналізоване підбирання контенту.

Завдяки стрімкому розвитку ШІ та машинного навчання, дослідники та розробники приділяють особливу увагу кільком ключовим напрямам, даючи змогу вдосконалювати алгоритми, роблячи їх більш гнучкими, точними та

придатними для широкого спектра завдань. Основні напрями досліджень у галузі ШІ та машинного навчання подано в табл. 2.3.

Таблиця 2.3 – Основні напрями досліджень

Напрямок досліджень	**Опис**
Створення складніших і точніших моделей	Постійні дослідження дають змогу розробляти моделі, здатні вирішувати більш комплексні завдання, що сприяє розвитку ШІ та машинного навчання.
Урахування найдрібніших деталей у моделях	Сучасні моделі обробляють навіть найменші деталі, що значно покращує точність прогнозів та рекомендацій.
Підвищення універсальності рішень	Технології ШІ адаптуються для різних галузей, що розширює можливості практичного застосування даних у різноманітних сферах діяльності.

З огляду на швидкий темп розвитку ШІ, подальші дослідження зосереджуються не лише на покращенні точності та швидкості обробки даних, але й на забезпеченні етичних стандартів, конфіденційності даних і безпеки використання технологій. Все більше уваги приділяється розробці прозорих алгоритмів, які дають змогу користувачам зрозуміти логіку ухвалення рішень у рекомендаційних системах.

Рекомендаційна система є важливим інструментом для обробки та фільтрації великих обсягів даних, що дозволяє швидко генерувати найкращі рекомендації. Розглянуті концепції підтверджують здатність системи адаптуватися до різних потреб, знижуючи інформаційне перевантаження та сприяючи прийняттю обґрунтованих рішень.

Отже, завдяки технологіям машинного навчання, зокрема глибинному навчанню, рекомендаційна система не лише покращує якість рекомендацій, а й підвищує загальну ефективність інформаційних процесів, що є критично важливим у сучасному цифровому середовищі. Розвиток таких технологій сприяє також підвищенню взаємодії користувача з системою, що в результаті покращує користувацький досвід.

2.2 Огляд методів, що використовуються у рекомендаційних системах

Рекомендаційні системи охоплюють широкий спектр методів, кожен з яких спрямований на підвищення релевантності та персоналізації наданих рекомендацій. З розвитком технологій та збільшенням обсягу доступних даних, ці системи стають дедалі складнішими та більш ефективними. З розширенням сфер їх застосування виникла необхідність у чіткій класифікації наявних методів, що дає змогу краще розуміти та оптимально застосовувати ці системи відповідно до конкретних завдань.

Завдяки різноманіттю підходів, рекомендаційні системи можуть адаптуватися до специфіки предметної області, забезпечуючи водночас високу точність і зручність для користувачів. Залежно від цілей та характеристик контенту, кожен метод має свої унікальні переваги, які сприяють створенню персоналізованих пропозицій. Проте, важливо обирати відповідний метод, враховуючи ресурси та можливості системи, щоб забезпечити оптимальну продуктивність. Отже, класифікація рекомендаційних систем є важливим інструментом для досягнення балансу між ефективністю обробки даних і релевантністю наданих рекомендацій.

Науковці класифікують рекомендаційні системи за трьома типами (рис. 2.1).

Рис. 2.1. Класифікація рекомендаційних систем

Як видно з рис. 2.1, рекомендаційні системи поділяться на три основних типи: колаборативну фільтрацію, рекомендації на основі контенту та гібридні

моделі. Кожен із цих типів використовує різні підходи до формування рекомендацій для користувачів, залежно від аналізованих даних і цілей системи.

Колаборативна фільтрація є одним із найефективніших підходів у рекомендаційних системах, оскільки дає змогу формувати прогнози для користувача на основі аналізу схожих взаємодій, здійснених іншими користувачами. (рис. 2.2).

Рис. 2.2. Колаборативна фільтрація

Як видно з рис. 2.2, такий тип рекомендаційної системи базується на використанні інформації про переваги групи користувачів для формування рекомендацій іншому, активному користувачеві. Такий підхід дозволяє підбирати контент, продукти чи послуги, які з великою ймовірністю будуть цікаві користувачу, засновуючись на схожості його інтересів з інтересами інших учасників групи. Такий спосіб аналізу дозволяє створювати ефективні алгоритми, які враховують унікальність кожного користувача.

Ці методи відповідно, створюють рекомендації лише на основі виявлених попередніх взаємодій між користувачами та елементами. Кожна нова рекомендація залежить від знань про поведінку користувача. Вважається, що даних про минулі взаємодії достатньо для виявлення подібних користувачів та/або подібних елементів і відповідно здійснення прогнозів на основі оцінок встановленої подібності. Водночас виявлені схожості можуть бути використані для надання релевантних рекомендацій навіть у випадках, коли про конкретного користувача мало інформації Сильна сторона колаборативної фільтрації полягає в можливості застосування цього методу до об’єктів, які мають складну або

неоднозначну структуру. Це забезпечує ефективність системи навіть у випадках, коли відсутня точна інформація про контент, що рекомендується.

Ключова перевага підходів колаборативної фільтрації полягає в тому, що вони зовсім не покладаються на аналіз вмісту елементів чи характеристик користувачів. Для пошуку подібності, вироблення рекомендацій не потрібно опрацьовувати сам елемент, щоб зрозуміти його. Це особливо корисно у випадках, коли об'єкти мають складну або неоднозначну структуру, яку важко формально описати. Колаборативна фільтрація може бути ефективною навіть у випадках, коли мало або взагалі немає інформації про самі об'єкти, які рекомендуються. Чим активніше користувачі взаємодіють з елементами, тим точнішими стають нові рекомендації.

Інший тип, а саме контентна фільтрація використовує інформацію про властивості предметів. Ідея контентної фільтрації полягає в тому, що предметам із подібним контентом користувачі надають подібні переваги. Це дозволяє створювати рекомендації на основі аналізу подібностей між об'єктами, а не на основі поведінки інших користувачів. Фільтрацію вмісту можна застосувати у таких системах, де наперед передбачається наявність описових даних.

Фільтрація на основі знань – пропонує продукти на основі висновків про потреби та вподобання користувачів, вибір предметів та основу для рекомендацій. Цей підхід часто використовується у випадках, коли необхідно надати рекомендації на основі конкретних правил або знань про предметну область.

Найбільш ефективні рекомендаційні системи використовують комбінації підходів фільтрації даних і колаборативної фільтрації, так звані гібридними моделями. Такі системи можуть поєднувати переваги обох підходів, зменшуючи їхні недоліки. Гібридні моделі часто використовують глибинні нейронні мережі або інші методи машинного навчання, що дають змогу підвищити точність рекомендацій шляхом обробки великих обсягів даних. Поєднання різних алгоритмів та підходів у гібридних моделях, таких як колаборативна фільтрація, контентна фільтрація, допомагає створити більш персоналізовані та релевантні

рекомендації для користувачів. У деяких випадках гібридні системи також можуть автоматично підлаштовуватись до потреб різних груп користувачів, використовуючи методи кластеризації або багатошарового аналізу. З розвитком ШІ та глибинного навчання гібридні моделі продовжують вдосконалюватись, що дозволяє створювати ще точніші та адаптованіші рекомендаційні системи. Зокрема, сучасні гібридні підходи активно впроваджують трансформери та інші архітектури, які демонструють високу ефективність у задачах прогнозування та персоналізації.

Експерти виділяють кілька основних варіантів комбінування різних методів в рекомендаційних системах гібридного типу (рис. 2.3).

Рис. 2.3. Комбінування методів в рекомендаційних системах гібридного типу

Як видно з рис. 2.3, комбінування методів у рекомендаційних системах гібридного типу дозволяє підвищити точність і релевантність рекомендацій. Цей підхід поєднує переваги різних алгоритмів та компенсує їх недоліки, що сприяє покращенню загальної ефективності системи. Основні частини комбінування методів в рекомендаційних системах гібридного типу подано в табл. 2.4.

Таблиця 2.4 – Основні частини комбінування методів в рекомендаційних системах гібридного типу

Частина	Опис
Побудова єдиної моделі, що використовує характеристики обох методів.	Інтегруються характеристики і user-based, та item-based методів для створення єдиної моделі.
Впровадження item-based характеристик в user-based системи.	Використовуються характеристики item-based методу у системі, яка зазвичай базується на user-based методах.
Впровадження user-based і item-based підходів окремо і використання комбінації отриманих рейтингів.	Обидва підходи застосовуються окремо, а результати їх використання комбінуються для підвищення точності рекомендацій.
Впровадження user-based характеристик в item-based системи.	Використовуються user-based характеристики в системах, які традиційно зосереджені на item-based підходах.

Таке комбінування забезпечує точнішу персоналізацію та гнучкість рекомендаційної системи, даючи змогу враховувати як історію взаємодій користувачів, так і характеристики об'єктів, що робить її універсальною у різних застосуваннях. Однак, для досягнення високої точності рекомендацій необхідно враховувати й новітні підходи, що дозволяють глибше зрозуміти контекст даних та забезпечити адаптивність системи до різних типів інформації.

Сучасні підходи до створення рекомендаційних систем все частіше використовують мовні моделі на основі трансформерів, що забезпечують новий рівень точності та персоналізації рекомендацій. Одним із перспективних напрямів є використання мовних моделей на основі популярних трансформерів, таких як BERT або GPT, які здатні значно покращити якість аналізу та обробки тексту. У контексті глибинного навчання та обробки природної мови трансформери означають спеціальну архітектуру нейронних мереж, розроблену для обробки послідовностей даних, таких як текст. Особливістю трансформерів є їхня здатність паралельно обробляти великі обсяги інформації, що забезпечує високу продуктивність та точність при роботі з довгими послідовностями тексту.

Вони стали проривом завдяки своїй здатності ефективно працювати з великими текстовими даними і враховувати контекст кожного слова, незалежно від його розташування в тексті, що робить їх важливим інструментом у галузі ШІ. Ключові переваги трансформерів включають:

1. Паралельну обробку даних, що значно пришвидшує тренування та застосування моделей.
2. Глобальний контекст: трансформери можуть розглядати всю послідовність даних одразу, що допомагає краще враховувати значення слів у тексті.
3. Універсальність: трансформери застосовуються не лише в обробці тексту, а й у таких галузях, як обробка зображень, аналіз послідовностей тощо.
4. Адаптивність до різних задач: трансформери можуть бути легко налаштовані для різних завдань, таких як класифікація тексту, генерація, машинний переклад та питання-відповідь.

Зокрема, серед таких мовних моделей можна виділити BERT та GPT, які базуються на архітектурі трансформерів і спеціалізуються на обробці природної мови. Вони використовуються для вирішення різних завдань, таких як:

- обробка тексту;
- генерація нового тексту;
- автоматичне завершення тексту;
- машинний переклад;
- відповіді на запитання.

Моделі часто інтегрують у рекомендаційні системи для покращення релевантності та персоналізації. Вони здатні не лише розуміти контекст і значення слів у конкретних завданнях, але й адаптуватися до нових типів даних і задач, що забезпечує їх широка застосовність. Такі моделі здатні:

- обробляти як структури, так і змісту текстів;
- підвищувати якість контентно-орієнтованих рекомендаційних систем.

До основних переваг інтеграції трансформерів у рекомендаційні системи відносять:

1. Покращення прозорості: трансформери дають змогу пояснювати результати, що допомагає розуміти, чому пропонуються певні рекомендації, та підвищує довіру до системи.
2. Збільшення ефективності: машинне навчання сприяє автоматизованій обробці та глибокому аналізу текстового контенту.
3. Підвищення якості рекомендацій: завдяки глибокій обробці та адаптації до нових задач і типів даних.

Таким чином, використання трансформерів у рекомендаційних системах надає суттєві переваги, які сприяють більшій ефективності та кращому користувацькому досвіду, забезпечуючи персоналізовані, зрозумілі та якісні рекомендації.

Огляд методів, що використовуються у рекомендаційних системах, показує, що ефективність рекомендаційних алгоритмів базується на їхній здатності адаптуватися до різних сценаріїв застосування. Колаборативна фільтрація, контентна фільтрація та гібридні моделі пропонують різні підходи для створення персоналізованих рекомендацій. Кожен із цих методів має свої сильні та слабкі сторони: колаборативна фільтрація є ефективною для виявлення прихованих уподобань користувачів, контентна підходить для аналізу характеристик об’єктів, а гібридні системи поєднувати всі переваги, зменшуючи обмеження.

Сучасні алгоритми, зокрема на основі трансформерів, надають можливості для глибшого розуміння контенту та покращення персоналізації, що робить рекомендаційні системи незамінними в умовах чимраз більших обсягів даних. Крім того, трансформери здатні адаптуватися до складних структур і зменшити залежність від явних ознак, виявляючи складніші зв'язки між даними. Однак, впровадження хоча б одного трансформера вимагає значних обчислювальних ресурсів, що ставить питання оптимізації та ефективності їх використання в рекомендаційних системах.

2.3 Опис механізмів роботи рекомендаційної системи

Рекомендаційна система є складним комплексним інструментом, що використовує передові алгоритми й методи для обробки та аналізу великих обсягів даних, з метою надання найкращих рекомендацій з покращення тексту. Завдяки цьому, вона забезпечує систематизований підхід до аналізу контенту, який послідовно вдосконалюється на кожному етапі.

Щоб зрозуміти принципи її роботи, важливо розглянути основні механізми та етапи функціонування такої системи. Механізми роботи рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту базуються на використанні алгоритмів, які виконують серію послідовних кроків, забезпечуючи високу точність і релевантність отриманих результатів. Основні етапи функціонування такої системи включають: збирання даних, фільтрація даних, навчання моделі та формулювання рекомендацій [16]. Кожен етап тісно інтегрований з наступним для підвищення якості та максимальної точності рекомендацій, що сприяє покращенню ефективності та високої продуктивності рекомендаційної системи.

Рекомендаційна система для автоматизованої обробки та аналізу текстового контенту працюватиме за принципом, як це показано на рис. 2.4.

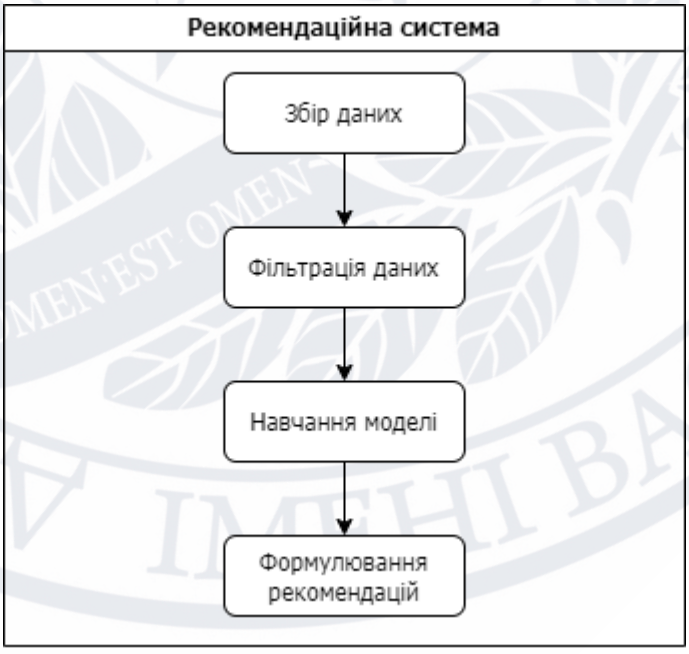

Рис. 2.4. Основні етапи роботи рекомендаційної системи

Як видно з рис. 2.4, на першому етапі рекомендаційна система виконує важливу роль автоматизації процесу збору та первинної підготовки даних. Спершу система отримує текстовий контент з різноманітних джерел, які можуть бути як структурованими, так і неструктурованими. До структурованих джерел належить лише база даних, оскільки вона організована за певними правилами, з чітко визначеною структурою та форматом даних. Ця гнучкість є ключовою в умовах чимраз більшого обсягу текстової інформації, яка потребує автоматизованої обробки. Кожен вхідний текст проходить первинну обробку, де виділяються ключові сегменти, що мають інформативну цінність.

На наступному етапі дані проходять фільтрацію для видалення непотрібної або надлишкової інформації. Фільтрація допомагає зменшити обсяг даних, підвищуючи точність аналізу. Цей етап значно знижує навантаження на систему, оскільки відокремлює релевантну інформацію від вторинної, що мінімізує інформаційне перевантаження й підвищує точність аналізу в подальших етапах роботи системи. Під час процесу фільтрації система використовує інтелектуальні методи відбору, які можуть враховувати як контекстні ознаки, так і значущі ключові слова, щоб забезпечити релевантність інформації для подальшого аналізу.

Таким чином, завдяки якісно виконаним етапам збору та попередньої обробки даних, рекомендаційна система отримує структурований та готовий до аналізу контент, що значно підвищує ефективність подальших етапів роботи системи.

Паралельно з аналізом контенту система виконує семантичний аналіз, що дозволяє враховувати не лише ключові слова, але й значення та контексти, які є критично важливими для покращення точності та релевантності рекомендацій.

На наступному етапі мовна модель GPT, яка є одним із передових досягнень у сфері ШІ, тренується на основі зібраних даних для того, щоб розпізнавати стиль тексту, граматичні конструкції та інші патерни, що впливають на якість контенту. Модель аналізує текстові особливості, щоб визначити, як його можна покращити в аспектах ясності, логіки та стилю.

Інтеграція нових підходів дозволяє рекомендаційній системі не лише автоматизувати процеси, але й підвищити їхню гнучкість та точність, що критично важливо в умовах роботи з великими обсягами інформації.

Завершальний етап – це генерація експертних рекомендацій. На основі результатів аналізу модель GPT пропонує конкретні рекомендації, як покращити текст: спростити складні речення, посилити логічні зв’язки, підібрати точніші вирази або синтаксичні конструкції, щоб зробити текст більш зрозумілим. Кожна рекомендація формується з урахуванням стилю та призначення тексту, що дозволяє зберегти авторське бачення. Рекомендації є персоналізованими й враховують загальний стиль та мету тексту. Рекомендації проходять перевірку на узгодженість і точність, після чого надаються користувачу у вигляді конкретних згенерованих пропозицій з покращення текстового контенту. На цьому етапі, після отримання рекомендацій, система продовжує вдосконалюватися, використовуючи зворотний зв'язок.

Важливо зазначити, що мовна модель GPT була додатково налаштована для конкретних завдань або сфер застосування. У нашому випадку це означає, що її можна адаптувати для допомоги в аналізі текстового матеріалу. Проте слід пам’ятати, що ця модель лише інструмент, а не заміна людської експертизи. Хоча вона може допомагати у різноманітних завданнях, важливо завжди перевіряти надані нею рекомендації.

Таким чином, через зазначені ключові етапи – збір даних, фільтрацію, навчання та формулювання рекомендацій – рекомендаційна система здійснює комплексну автоматизовану обробку та глибокий аналіз тексту і надає корисні рекомендації для підвищення його загальної якості. Завдяки можливостям сучасної моделі, такої як GPT, яка базується на трансформерній архітектурі рекомендаційна система швидко адаптується до тексту, пропонуючи оптимальні рекомендації. До того ж, завдяки автоматичній адаптації, система не лише підвищує якість тексту, але й покращує його сприйняття читачем, що робить її незамінною у сфері медіаіндустрії та інших напрямках, де висока якість контенту має ключове значення.

2.4 Опис програмно-апаратного інструментарію

Ефективність роботи сучасної рекомендаційної системи значною мірою залежить від оптимального використання програмно-апаратного інструментарію. Щоб забезпечити ефективну роботу рекомендаційної системи, важливо досягти оптимального балансу між:

- високою продуктивністю;
- економічністю використання ресурсів.

Для надійної роботи рекомендаційної системи для автоматизованої обробки та аналізу різноманітного текстового контенту не потрібні надмірні апаратні ресурси. Система спроєктована таким чином, щоб оптимально використовувати обчислювальні можливості, забезпечуючи високу швидкість обробки даних і надійну продуктивність навіть на відносно стандартному обладнанні. Це дає змогу запускати систему на локальних серверах або комп’ютерах, не вимагаючи значних витрат на апаратне забезпечення. Особлива увага приділяється:

- оптимізації алгоритмів обробки текстового контенту для забезпечення високої ефективності при мінімальних витратах на ресурси;
- використанню адаптивних алгоритмів оптимізації ресурсів для ефективного розподілу навантаження, що дозволяє підтримувати стабільну продуктивність навіть за пікових навантажень;
- зменшенню експлуатаційних витрат, що знижує залежність від високопотужного обладнання та спрощує масштабування системи при збільшенні обсягів оброблюваних даних.

Для впровадження рекомендаційної системи, яка швидко обробляє та якісно аналізує великі обсяги текстового контенту, використовується:

- оптимально підібрана апаратна архітектура, яка забезпечує високу продуктивність без надмірного використання ресурсів;
- особлива увага до оптимізації процесів зберігання та обробки даних для зниження часу відгуку системи та підвищення її ефективності.

Рекомендаційна система побудована з урахуванням жорстких вимог до обчислювальної ефективності, що дозволяє досягати необхідної швидкості обробки та аналізу даних на стандартному обладнанні. Таким чином, навіть за значного навантаження система здатна зберігати стабільну продуктивність, що є ключовим фактором для підтримки її надійності та доступності. Основні компоненти апаратної частини подано в табл. 2.5, яка містить детальний опис кожного використаного елемента апаратного забезпечення, залученого для реалізації рекомендаційної системи.

Таблиця 2.5 – Апаратна частина

Компонент	**Опис**
Центральний процесор	Процесор 11th Gen Intel Core i7-1165G7. Забезпечує високу продуктивність завдяки багатоядерній архітектурі та підтримці технологій Intel, даючи змогу виконувати паралельні обчислення під час аналізу текстових даних.
Оперативна пам'ять	Оперативна пам'ять 16 ГБ, яка дозволяє обробляти великі набори даних, необхідні для навчання та роботи алгоритмів машинного навчання.
Накопичувачі	Твердотілий накопичувач (SSD) для зберігання та швидкого доступу до даних. SSD забезпечує високу швидкість зчитування та запису, що значно скорочує час завантаження та обробки текстових даних.
Графічний процесор	Вбудована графіка Intel Iris Xe дозволяє оптимізувати певні завдання, такі як візуалізація даних. У деяких випадках може використовуватися для паралельної обробки даних, що підвищує ефективність системи.
Мережева інфраструктура	Підключення до високошвидкісної мережі забезпечує швидкий обмін даними, особливо при обробці та аналізі великого обсягу текстового контенту.

Зазначені компоненти апаратної частини забезпечують стабільну та ефективну роботу системи, даючи змогу обробляти великі обсяги текстових даних у режимі реального часу. Однак, для повноцінного функціонування рекомендаційної системи не менш важливим є програмне забезпечення, яке керує цими апаратними ресурсами, реалізує складні алгоритми машинного

навчання та забезпечує зручну взаємодію користувачів із системою. Програмне забезпечення містить кілька ключових елементів. Кожен елемент розроблений з урахуванням високих стандартів безпеки та ефективності, що забезпечує надійну роботу системи в різних умовах.

Основні компоненти програмної частини подано в табл. 2.6, яка містить детальний опис кожного елемента, використаного для реалізації рекомендаційної системи.

Таблиця 2.6 – Програмна частина

Компонент	**Опис**
База даних	Сукупність структурованих даних, які організовані відповідно до заздалегідь визначеної концепції, яка описує ключові характеристику цих даних і взаємозв'язки між їх елементами [17].
Система керування реляційними базами даних	MySQL – потужна реляційна система керування базами даних, яка дозволяє ефективно зберігати, обробляти та маніпулювати структурованими даними за допомогою стандартизованих SQL-запитів. Підтримує високу продуктивність, надійність та масштабованість.
Адміністрування бази даних	Для управління базою даних використовується MySQL Workbench – зручний десктопний додаток розроблений на мові C++, який надає інтуїтивно зрозумілий графічний інтерфейс, що дозволяє здійснювати розробку, моделювання та адміністрування бази даних. Не потребує вебсервера.
Мова програмування	Мова програмування Python, обрана за її простоту, гнучкість та багатий набір бібліотек для обробки даних і машинного навчання, що дає змогу ефективно реалізувати необхідні алгоритми та функціональність системи.
Графічний інтерфейс користувача	Для побудови графічного інтерфейсу використовується фреймворк Qt у поєднанні з модулем PyQt, який забезпечує зв'язок з мовою програмування Python.
OpenAI API	Доступ до будь-якої моделі GPT, розробленої дослідницькою лабораторією OpenAI. API пропонує універсальний інтерфейс «text in, text out», що дає змогу використовувати його для будь-якого завдання.

Перевагами SQL підходу є незалежність від конкретної системи управління базами даних, наявність стандартів, декларативність, швидка обробка запитів та портативність [18].

Використання MySQL як основної СКРБД забезпечує надійне зберігання та ефективну обробку великих обсягів структурованих даних, дозволяючи зберігати дані у вигляді таблиць та забезпечуючи швидкий доступ до них за допомогою SQL-запитів. MySQL підтримує високий рівень продуктивності, надійності та масштабованості, що є важливим для роботи з великими обсягами інформації в режимі реального часу. Завдяки своїй відкритій архітектурі, MySQL також дозволяє легко інтегруватися з іншими компонентами системи, забезпечуючи гнучкість у роботі з даними.

MySQL Workbench використовується для полегшення адміністрування бази даних, надаючи зручний графічний інтерфейс для розробки та моделювання структури даних, управління SQL-запитами, а також моніторингу продуктивності бази даних. Це значно спрощує роботу з базами даних, роблячи процеси налаштування та адміністрування більш доступними та зрозумілими навіть для користувачів без глибоких знань SQL.

Мова програмування Python, своєю чергою, надає потужний набір інструментів для обробки даних і розробки алгоритмів машинного навчання, що робить її ідеальним вибором для створення рекомендаційної системи.

Використання фреймворку Qt з модулем PyQt для побудови графічного інтерфейсу користувача дозволяє зручно взаємодіяти з системою через інтуїтивно зрозумілий інтерфейс, що значно спрощує роботу для кінцевого користувача, особливо в умовах обробки великих обсягів текстових даних. Завдяки Qt забезпечується підтримка різних платформ, що дозволяє запускати програму на різних операційних системах, забезпечуючи кросплатформенність і зручність доступу до системи.

PyQt надає широкі можливості для створення графічних елементів, таких як кнопки, меню, панелі інструментів, а також для налаштування поведінки елементів залежно від контексту роботи системи. Це забезпечує адаптивність

інтерфейсу та його інтерактивність, дозволяючи системі швидко реагувати на дії користувача та знижуючи час виконання повторюваних операцій.

Крім того, PyQt підтримує налаштування стилів та тем, що дозволяє зробити інтерфейс не лише функціональним, але й естетично привабливим, з можливістю персоналізації під потреби користувача. Цей підхід підвищує зручність і комфорт роботи з системою, забезпечуючи її інтерактивність та зручність використання навіть для користувачів без спеціальних технічних навичок.

Додатково, Python підтримує інтеграцію з сучасними інструментами для обробки природної мови, такими як OpenAI API, що забезпечує доступ до всіх доступних мовних моделей GPT. Це дозволяє автоматично генерувати рекомендації, покращуючи персоналізацію та релевантність контенту. Python також має вбудовані засоби для роботи з API та інтеграції з іншими сервісами, що полегшує підключення зовнішніх даних та інструментів для збагачення рекомендацій.

Для обробки складної природної мови та глибокого аналізу текстового контенту інтегровано OpenAI API, що забезпечує доступ до потужних мовних моделей GPT, що дозволяє генерувати релевантні рекомендації, оптимізуючи процес обробки та підвищуючи точність рекомендацій шляхом використання сучасних методів машинного навчання.

Завдяки своїй універсальності та зручності у використанні, Python прискорює розробку прототипів і дозволяє швидко тестувати різні мовні моделі, що сприяє постійному вдосконаленню інтелектуальної рекомендаційної системи.

Завдяки використанню зазначеного програмно-апаратного інструментарію рекомендаційна система має можливість ефективно обробляти та аналізувати текстовий контент, забезпечуючи високу швидкість, гнучкість та надійність роботи. Таким чином, завдяки продуманій архітектурі програмно-апаратного інструментарію система здатна забезпечити стабільну продуктивність і високу ефективність навіть в умовах інтенсивних навантажень.

2.5 Виклики та проблеми у розробці рекомендаційної системи

Будь-яка рекомендаційна система має низку стандартних проблем та викликів, дослідження яких та розробка методів подолання яких є актуальною науково-практичною задачею. До найпоширеніших проблем рекомендаційної системи можна віднести проблему холодного старту та бульбашки фільтрів [19]. Важливо розуміти, що ці проблеми не лише знижують якість рекомендацій, але й можуть призвести до втрати користувачів, які не отримали релевантних пропозицій від системи. Вони можуть суттєво вплинути на ефективність та точність рекомендацій, що робить їх вирішення пріоритетним завданням для дослідників та розробників.

Одна з головних проблем рекомендаційної системи – це проблема холодного старту. Вона виникає тоді, коли в системі з'являються нові елементи – або нові користувачі, історія вподобань яких порожня, або нові об'єкти, у яких ще немає оцінок та/або набору ознак. Це призводить до ситуацій, коли система не може надійно передбачити, які об'єкти можуть бути цікавими для користувача, що, своєю чергою, може вплинути на якість його досвіду з використанням системи. Крім того, проблема холодного старту ускладнюється в системах, які працюють у реальному часі, де швидкість адаптації до нових користувачів або елементів є критично важливою для підтримки актуальності та своєчасності рекомендацій.

Ця проблема ускладнює процес генерації точних рекомендацій на початковому етапі, що може знижувати задоволення користувачів від взаємодії з системою. Одним із можливих підходів до подолання цієї проблеми є використання контекстуальних та демографічних даних нових користувачів для побудови початкових рекомендацій. Також, для нових об'єктів можна застосовувати методи контентного аналізу, які дають змогу визначити їхні характеристики та зіставити з подібними об'єктами, що вже мають оцінки. Це може включати аналіз текстового опису, метаданих та інших атрибутів об'єктів, що дозволяє створити більш структуровану модель їх взаємозв'язків.

Для розв’язання проблеми Cold-start Problem зазвичай застосовують наступні підходи [20]:

1. Гібридизація рекомендаційної системи з поєднанням контентної та колаборативної фільтрації.
2. Використання контексту, в якому створюються та надаються рекомендації (демографічні дані, час та дата тощо).

Не менш важливою проблемою рекомендаційної системи є проблема бульбашки фільтрів. Класична рекомендаційна система пропонує користувачам об’єкти, виходячи лише з їхніх попередніх уподобань. Отже, користувач потрапляє в інформаційне середовище, в якому спостерігає лише обмежену кількість однотипних об’єктів. Це обмежує не тільки інтереси користувача, а й можливість системи розвиватися та пропонувати більш різноманітний контент. Наслідки, викликані бульбашкою фільтрів [20].

1. Користувач не одержує альтернативної інформації, яка може бути йому корисною (наприклад, види об’єктів, про які він зовсім не знає, але які ефективніше розв’яжуть задачі його пошуку).
2. У користувача формується викривлений погляд на інформаційне середовище через те, що він не бачить цілісної картини (наприклад, при рекомендації новин).
3. Користувач може втратити інтерес до списку рекомендацій через те, що йому весь час пропонують однотипні об’єкти.

Для розв’язання проблеми бульбашки фільтрів, як правило, застосовують виконання додаткових вимог до формування списку рекомендацій. Одним із підходів – впровадження механізмів різноманітності та новизни у процес генерації рекомендацій, щоб користувачі отримували ширший спектр об’єктів, включаючи ті, які вони зазвичай не обирають. Додатково, системи можуть впроваджувати елементи випадковості або експерименти з рекомендаціями, щоб періодично пропонувати користувачам об'єкти, які не пов'язані з їхнім попереднім досвідом, тим самим стимулюючи інтерес до нового контенту.

Основні властивості, які допомагають уникнути проблеми бульбашки фільтрів, наведені в табл. 2.7

Таблиця 2.7 – Властивості для розв’язання проблеми бульбашки фільтрів у рекомендаційній системі

Властивість	**Опис**
Різноманітність	Міра схожості елементів списку. Елементи у списку рекомендацій не повинні бути майже однаковими, вони повинні містити різнотипні об’єкти (наприклад, фільми різних жанрів, а не лише одного жанру, чи однієї трилогії).
Неочікуваність	Неочікуваність, несхожість на історію користувача. Елемент, який має властивість неочікуваності, повинен бути важливим, новим та непрогнозованим для користувача.
Новизна	Новизна для користувача, поширеність продукту, частка його рейтингів. Якщо користувачу рекомендувати лише популярні об’єкти, скоріше за все він їх і так знає та обере без рекомендаційної системи, такі рекомендації не будуть містити для нього нової інформації.

Виконання наведених вище вимог допомагає забезпечити ефективнішу роботу рекомендаційної системи, зменшуючи вплив проблеми бульбашки фільтрів. Урахування різноманітності, неочікуваності та новизни дозволяє системі надавати користувачам більш збалансовані та корисні рекомендації, що покращує загальне враження від користування системою.

Розробка рекомендаційної системи є складним і багатоаспектним завданням, яке виходить далеко за межі простого розв’язання конкретних проблем. Цей процес передбачає глибоке розуміння користувацьких потреб та постійного вдосконалення алгоритмів. Алгоритми повинні адаптуватися до змін у поведінці користувачів та враховувати їхні індивідуальні вподобання та контекст. Окрім розв’язання конкретних проблем, розробка рекомендаційної системи стикається з низкою різноманітних викликів. Серед них – необхідність забезпечення високої точності та релевантності рекомендацій у динамічних умовах, коли потреби та інтереси користувачів можуть змінюватися дуже швидко.

Для того, щоб зробити перелік викликів у розробці рекомендаційних систем більш наочним, нижче подано табл. 2.8, яка відображає основні аспекти та їх опис.

Таблиця 2.8 – Основні виклики в розробці рекомендаційних систем

Виклик	**Опис**
Масштабованість	Система повинна ефективно обробляти та аналізувати великі обсяги даних у реальному часі.
Персоналізація	Забезпечення точного урахування індивідуальних вподобань кожного користувача без втрати якості рекомендацій.
Проблема неоднозначних та специфічних запитів	Система повинна ефективно обробляти неоднозначні запити або контент із вузькою спеціалізацією, що вимагає складних алгоритмів обробки природної мови.
Приватність та безпека даних	Захист персональних даних користувачів та забезпечення конфіденційності інформації, що використовується для рекомендацій.
Адаптивність	Система повинна бути здатна адаптуватися до змін вподобань користувачів та нових тенденцій у даних.
Етичні аспекти	Уникнення упереджень та забезпечення справедливості у рекомендаціях, щоб уникнути дискримінації чи несправедливого трактування даних користувачів.
Забезпечення прозорості та довіри	Система повинна забезпечувати зрозумілу логіку рекомендацій для користувачів, щоб підвищити їх довіру. Це стає особливо важливим для систем, що використовують глибинні нейронні мережі або трансформери, де важко пояснити рішення алгоритму.

Всі перелічені виклики вимагають постійного вдосконалення алгоритмів, тестування, а також дотримання етичних стандартів для створення надійної та ефективної рекомендаційної системи.

Проте, окрім загальних викликів, розробники стикаються з певними проблемами при використанні конкретних моделей, таких як GPT. Мовна модель GPT має свої переваги, надаючи рекомендаційній системі можливість розуміти та інтерпретувати природну мову на глибшому рівні, що є значним досягненням при обробці великих масивів тексту. Однак вона створює певні виклики, які

необхідно долати. Основні проблеми, які пов’язані з використанням мовної моделі GPT у рекомендаційних системах подано в табл. 2.9.

Таблиця 2.9 – Основні проблеми використання мовної моделі GPT

Проблема	**Опис**
Обчислювальна складність	Мовна модель GPT вимагає значних обчислювальних ресурсів, оскільки робота з великими масивами текстових даних потребує високої продуктивності.
Контекстуальна релевантність	GPT може генерувати узагальнені або надмірно детальні відповіді, які не завжди відповідають індивідуальним потребам користувача. Важливою проблемою є налаштування системи таким чином, щоб рекомендації були адаптовані до конкретних цілей і вимог.
Обробка неоднозначних та специфічних запитів	GPT, хоча і є потужною мовною моделлю, іноді має труднощі з обробкою вузькоспеціалізованих або контекстуально неоднозначних запитів. Це може вимагати додаткового навчання моделі на специфічних наборах даних для підвищення точності рекомендацій.
Етичні та приватні аспекти	Використання GPT вимагає дотримання принципів конфіденційності, особливо при роботі з персональними даними.

Отже, для подолання проблем у розробці рекомендаційної системи необхідно використовувати гібридизацію різних методів її побудови, враховувати контекстну інформацію, забезпечувати виконання додаткових вимог до формування списку рекомендацій. Це, як правило, зменшує точність прогнозування, проте підвищує якість рекомендацій. Водночас розробники повинні вирішувати виклики масштабованості, персоналізації, приватності, адаптивності та етики для створення ефективних і надійних систем.

Важливою частиною розробки рекомендаційної системи є впровадження мовної моделі GPT, яка забезпечує глибше розуміння тексту та контексту. Проте її використання також пов'язане з певними труднощами, зокрема з високими вимогами до обчислювальних ресурсів і необхідністю налаштування моделі для специфічних запитів.

РОЗДІЛ 3
ПРОГРАМНА РЕАЛІЗАЦІЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ

3.1 Загальна архітектура та принцип роботи системи

Основний принцип роботи рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту полягає в тому, що перед введенням будь-якого тексту відбувається процес автентифікації користувача для забезпечення безпеки та цілісності даних. Після цього текст автоматизовано обробляється за допомогою інтелектуальних алгоритмів та MySQL, яка зберігає дані в реляційній базі даних. Потім за допомогою найостаннішої версії мовної моделі GPT-4o аналізується текст, визначаються ключові аспекти для покращення, та за допомогою заздалегідь прописаного в ній промпту (інструкції) генеруються рекомендації з його покращення та новий оброблений текст (рис. 3.1).

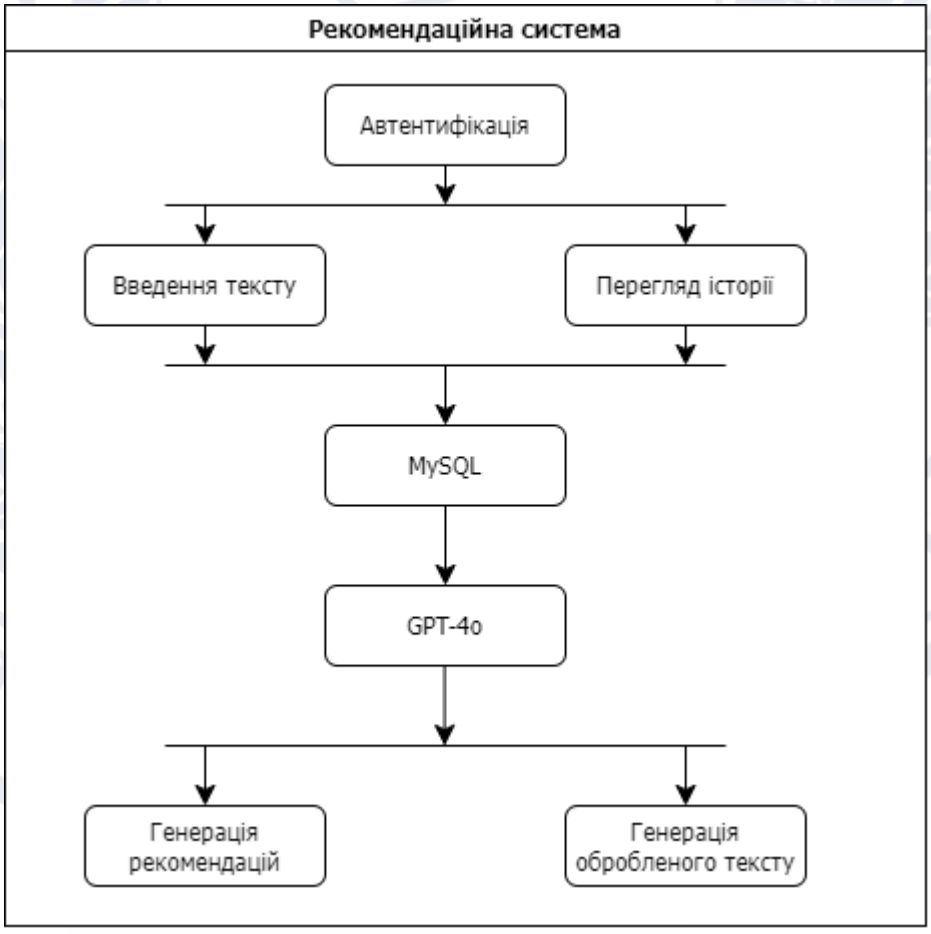

Рис. 3.1. Блок-схема роботи рекомендаційної системи

Як видно з рис. 3.1, архітектура рекомендаційної системи побудована за принципом модульності, що дозволяє розбити її на незалежні компоненти, кожен з яких виконує окрему частину функціоналу. Така комплексна взаємодія всіх частин дозволяє врешті-решт отримати найкращу версію текстового контенту. Принцип модульності є основою сучасних архітектур, зокрема рекомендаційних систем, дозволяючи розділити систему на менші компоненти. Кожен модуль виконує специфічний функціонал і може бути розроблений, протестований та оновлений незалежно.

Кожен варіант використання показує окремий функціональний процес, такий як завантаження текстового контенту, отримання рекомендацій та перегляд результатів аналізу. Такий підхід відображає типову структуру роботи систем з рекомендаційною функцією. Кожен із варіантів використання також включає взаємодію з різними модулями системи, що дозволяє відстежувати етапи обробки тексту та забезпечує логічний зв'язок між ними. Модулі, що відповідають за різні функції (завантаження, аналіз, рекомендації) інтегруються через інтерфейс для забезпечення зв'язку між процесами.

Рекомендаційна система для автоматизованої обробки та аналізу текстового контенту побудована на основі ключових характеристик, як це показано на рис. 3.2.

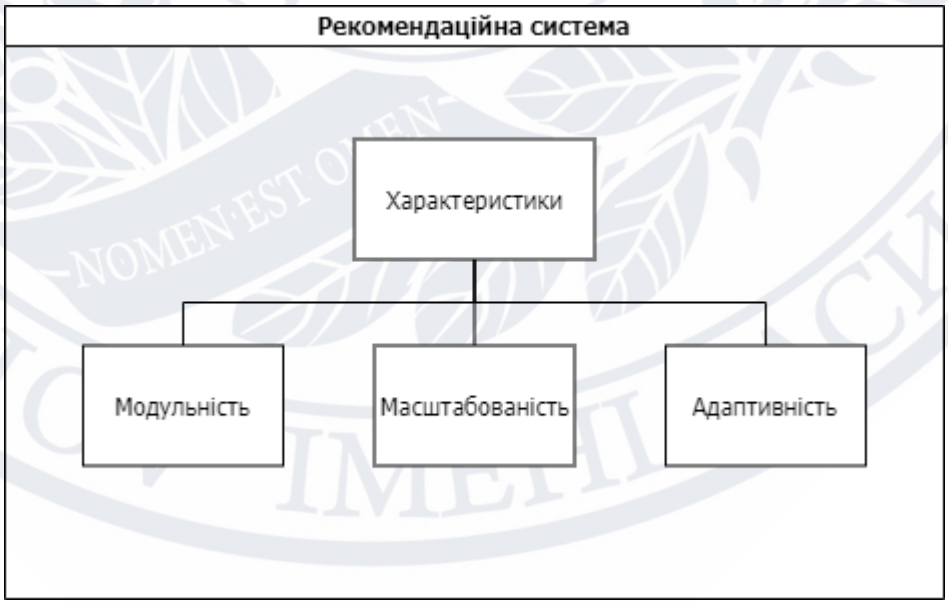

Рис. 3.2. Основні характеристики побудови рекомендаційної системи

Як видно з рис. 3.2, рекомендаційна система базується на трьох характеристиках:

1. Модульність: рекомендаційна система розбита на незалежні компоненти, кожен з яких відповідає за конкретну частину функціоналу.
2. Масштабованість: архітектура рекомендаційної системи спроєктована таким чином, щоб вона могла обробляти збільшення обсягу даних і кількості користувачів без значного зниження продуктивності.
3. Адаптивність: рекомендаційна система здатна динамічно змінювати свої алгоритми та рекомендації на основі аналізу нових даних, що дозволяє підтримувати актуальність та релевантність результатів.

Потік даних у рекомендаційній системі забезпечує оперативний збір та аналіз інформації, яка швидко надходить до неї, даючи змогу постійно оновлювати моделі рекомендацій. Потік даних у рекомендаційній системі відбуватиметься, як це показано на рис. 3.3.

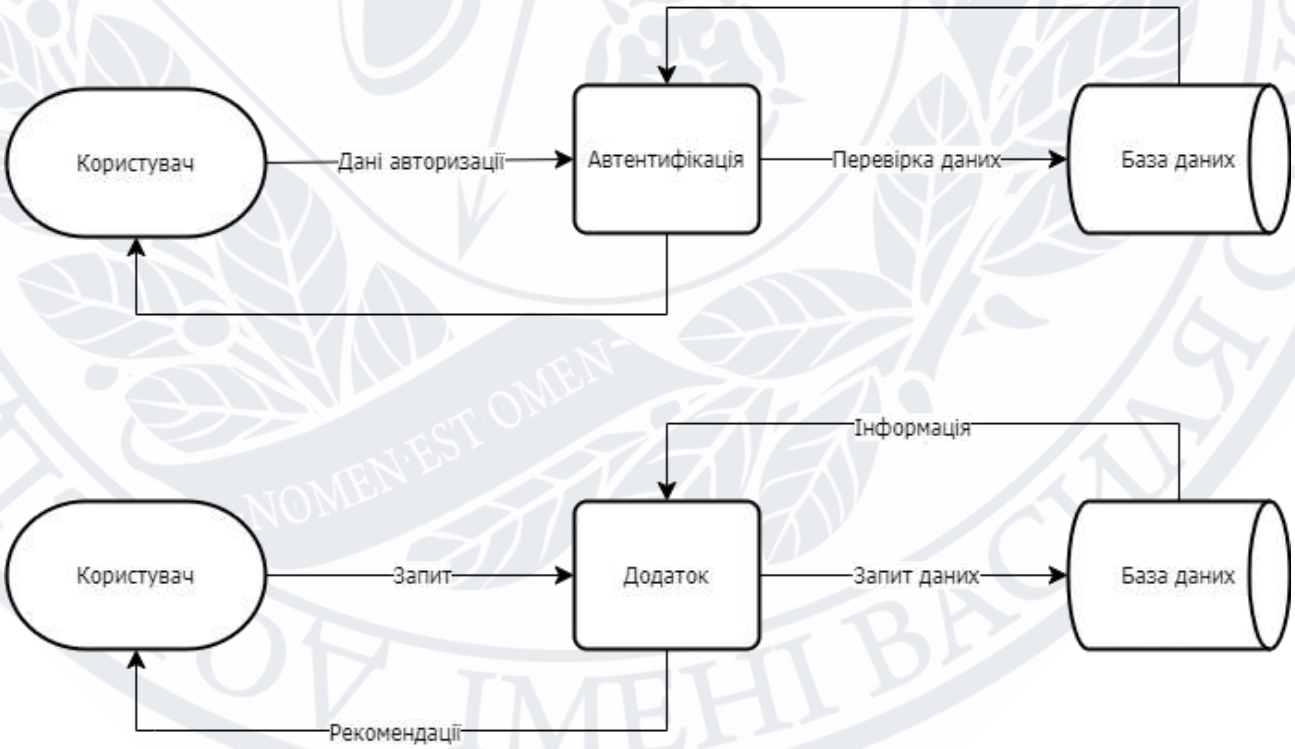

Рис. 3.3. Потік даних у рекомендаційній системі

Завдяки такому структурованому потоку даних рекомендаційна система постійно оновлює моделі рекомендацій, адаптуючись до нових даних та підтримуючи актуальність результатів. Система здатна ефективно обробляти дані завдяки використанню розподілених обчислень та алгоритмів, оптимізованих для паралельної обробки інформації.

Розглянута архітектура рекомендаційної системи відображає важливість інтеграції гнучких і модульних рішень у сучасних інформаційних системах. Поєднання ключових компонентів для збору, обробки та аналізу текстових даних забезпечує не лише автоматизацію рутинних завдань, а й отримання найточніших результатів. Рекомендаційна система використовує потокову обробку даних для постійного оновлення моделей рекомендацій, що дозволяє їй адаптуватися до нових вхідних даних і підтримувати актуальність рекомендацій.

3.2 Структура бази даних

У сучасному світі, де кількість інформації стрімко зростає, постає актуальне питання організації знань в одному універсальному просторі з можливістю динамічно додавати, оновлювати, редагувати та багаторазово використовувати дані. Сьогодні такий функціонал повністю надають бази даних.

База даних – це єдине велике сховище даних, яке визначаться один раз, а потім використовується одночасно багатьма користувачами. Замість розрізнених файлів із надмірними даними (що застосовується у файлових системах) тут усі дані зібрані разом із мінімальною часткою надмірності. Базу даних також називають набором інтегрованих записів з самоописом. У сукупності опис даних називається системним каталогом або словником даних, а самі елементи опису заведено називати метаданими. Саме наявність самоопису даних у базі даних забезпечує у ній незалежність програм від даних [21].

У загальному випадку база даних містить різноманітні схеми, таблиці, подання, збережені процедури та інші об'єкти. Дані у базі організовують відповідно до конкретної моделі організації даних. Таким чином, сучасна база

даних, крім саме даних, містить їх детальний опис і може містити засоби для їх обробки.

На сьогодні в інформаційному світі існують два основних напрями в базах даних: реляційні та нереляційні бази даних, тобто SQL і NoSQL, які відрізняються за своєю структурою, способами зберігання та обробки даних, а також підходами до забезпечення цілісності та доступності інформації.

Реляційні бази даних є одним з найпоширеніших способів зберігання і обробки даних в інформаційних системах. Вони дають змогу структурувати дані у вигляді таблиць та зв’язків між ними. Однією з ключових особливостей реляційних баз даних є підтримка транзакцій – групи операцій, які виконуються як єдине ціле. Транзакції дають змогу забезпечити цілісність та узгодженість даних у разі одночасної роботи багатьох користувачів [22].

Переваги реляційної моделі даних полягають в простоті, надійності, гнучкості, зрозумілості та зручності фізичної реалізації на комп'ютерах. Реляційний підхід ґрунтується на невеликій кількості інтуїтивно зрозумілих абстракцій, на основі яких можливо просте моделювання найбільш поширених наочних областей; реляційний підхід забезпечує можливість ненавігаційного маніпулювання даними без необхідності знання конкретної фізичної організації баз даних у зовнішній пам’яті. Основними недоліками реляційної моделі є відсутність стандартних засобів ідентифікації окремих записів, складність опису ієрархічних і мережевих зв'язків і відповідно повільна розробка.

Для рекомендаційної системи з автоматизованої обробки та аналізу текстового контенту було обрано реляційну базу даних, що забезпечує ефективне управління даними в структурованому вигляді. Реляційна база дозволяє зберігати дані у формі таблиць, де інформація чітко організована за стовпцями та рядками. Завдяки цьому стає можливим зберігання, обробка та аналіз великих обсягів текстових даних з високим рівнем точності та швидкості доступу до них.

Проте сама по собі реляційна база даних не зможе забезпечити повноцінну багатокористувацьку роботу та контроль над одночасним доступом кількох користувачів. Для цього потрібна спеціальна СКРБД, яка не лише зберігатиме

дані, а й забезпечуватиме надійне управління доступом, транзакційність і захист даних від пошкоджень під час паралельних запитів. СКРБД дозволяє керувати структурою бази даних, контролювати права доступу користувачів, а також гарантувати цілісність даних та коректність обробки запитів.

Зважаючи на ці потреби, для реалізації реляційної бази даних обрано MySQL з відкритим вихідним кодом. Вона дозволяє зберігати та організовувати дані у формі таблиць, що полегшує взаємозв'язок і доступ до них, а також підтримує транзакційність, яка є критично важливою для забезпечення надійної роботи рекомендаційної системи в умовах багатокористувацького середовища.

MySQL є системою керування реляційними базами даних, яка працює на основі серверного принципу зберігання та обробки даних. Це означає, що MySQL дозволяє зберігати, організовувати та управляти великими обсягами даних за допомогою сервера, до якого можуть підключатися різні клієнтські додатки.

Однією з переваг MySQL є можливість розміщення кількох баз даних на одному сервері. Це забезпечує централізоване управління даними, полегшує адміністрування та оптимізує використання апаратних ресурсів. Кожна база даних на сервері може бути ізольованою від інших, що дозволяє зберігати дані різних проєктів, додатків або клієнтів на одній фізичній машині без змішування або перетину даних.

Для підключення до сервера MySQL використовуються різні клієнтські додатки та інтерфейси, такі як командний рядок, графічні інтерфейси або програмні інтерфейси для мов програмування, включаючи PHP, Python, Java та інші. Це дозволяє гнучко керувати базами даних, виконувати SQL-запити, управляти таблицями та користувачами.

Іншою перевагою використання MySQL як СКРБД є її висока продуктивність і підтримка транзакцій, що дозволяє забезпечити ефективний доступ і управління великими обсягами даних. MySQL надає потужний інструментарій для оптимізації роботи з даними, завдяки чому система здатна

швидко обробляти запити навіть при великих навантаженнях. Важливими особливостями MySQL є:

1. Підтримка транзакцій: MySQL дозволяє забезпечити цілісність даних завдяки транзакціям, які гарантують, що всі операції виконуються повністю або не виконуються взагалі.
2. Швидкий доступ до даних: Використання SQL-запитів дозволяє швидко отримувати доступ до необхідної інформації, а також виконувати складні операції над даними, такі як агрегування, сортування та фільтрація, що є необхідним для роботи рекомендаційних систем.
3. Оптимізація запитів: MySQL забезпечує можливості для оптимізації запитів, наприклад, за допомогою індексації таблиць, що дозволяє значно скоротити час виконання запитів.

MySQL забезпечує надійне управління даними завдяки розширеним можливостям роботи із запитами. Бази даних MySQL легко обробляють великі обсяги даних, що важливо для сучасних систем. Висока продуктивність і гнучкість налаштувань роблять MySQL популярним вибором серед розробників.

Для спрощення моделювання, налаштування структури та адміністрування бази даних використовується MySQL Workbench – інтуїтивно зрозумілий інструмент для роботи з MySQL. MySQL Workbench дозволяє:

1. Проєктувати та візуалізувати структуру бази даних: Завдяки вбудованому редактору можна швидко створювати та змінювати схему бази даних, додавати нові таблиці, налаштовувати зв'язки між ними, що полегшує проєктування оптимальної структури бази даних.
2. Адмініструвати базу даних: MySQL Workbench дозволяє керувати даними, виконувати SQL-запити, а також моніторити продуктивність бази даних. Це забезпечує швидкий доступ до інструментів управління та спрощує процес адміністрування.
3. Аналізувати продуктивність: Засоби для моніторингу продуктивності дозволяють відстежувати та оптимізувати роботу запитів, що сприяє підвищенню загальної ефективності бази даних.

Крім того, MySQL Workbench підтримує функціонал для автоматизації певних завдань адміністрування бази даних, таких як регулярне оновлення даних або оптимізація запитів. Це зменшує кількість рутинних дій і зосередитися на складніших аналітичних завданнях, що дозволяє ефективніше використовувати ресурси бази даних і підвищити її загальну продуктивність.

Ефективна структура бази даних сприяє швидкому доступу до інформації, що особливо важливо при роботі з великими обсягами даних. Це дозволяє уникати перевантаження системи й підтримувати стабільність її роботи. Для забезпечення високої ефективності обробки великих обсягів інформації база даних організована у вигляді кількох взаємопов'язаних таблиць, кожна з яких виконує конкретну функцію.

Мета правильного проєктування таблиць – видалення надлишкових (повторюваних даних). Правильне проєктування бази даних допомагає мінімізувати дублювання інформації та знижує ризик несумісності даних. Коли дані не дублюються і впорядковані за тематичними таблицями, це значно полегшує процес їх обробки, а також знижує потребу в коригуванні записів у різних місцях одночасно. Відсутність дублювання сприяє підвищенню швидкості обробки даних, оскільки система працює лише з необхідними елементами інформації, а також зменшує розмір бази даних, що оптимізує використання дискового простору.

Для досягнення цієї мети таблицю було розділено на багато тематичних таблиць, щоб кожний факт було представлено лише один раз. Кожна таблиця тепер відповідає за конкретний набір даних, що спрощує управління інформацією. Таке розділення таблиць забезпечує цілісність даних і уникнути помилок. Потім в пов’язані між собою таблиці встановлені спільні поля, які є ключами, що пов’язують записи між таблицями. Структура бази даних рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту показана на рисунку 3.4.

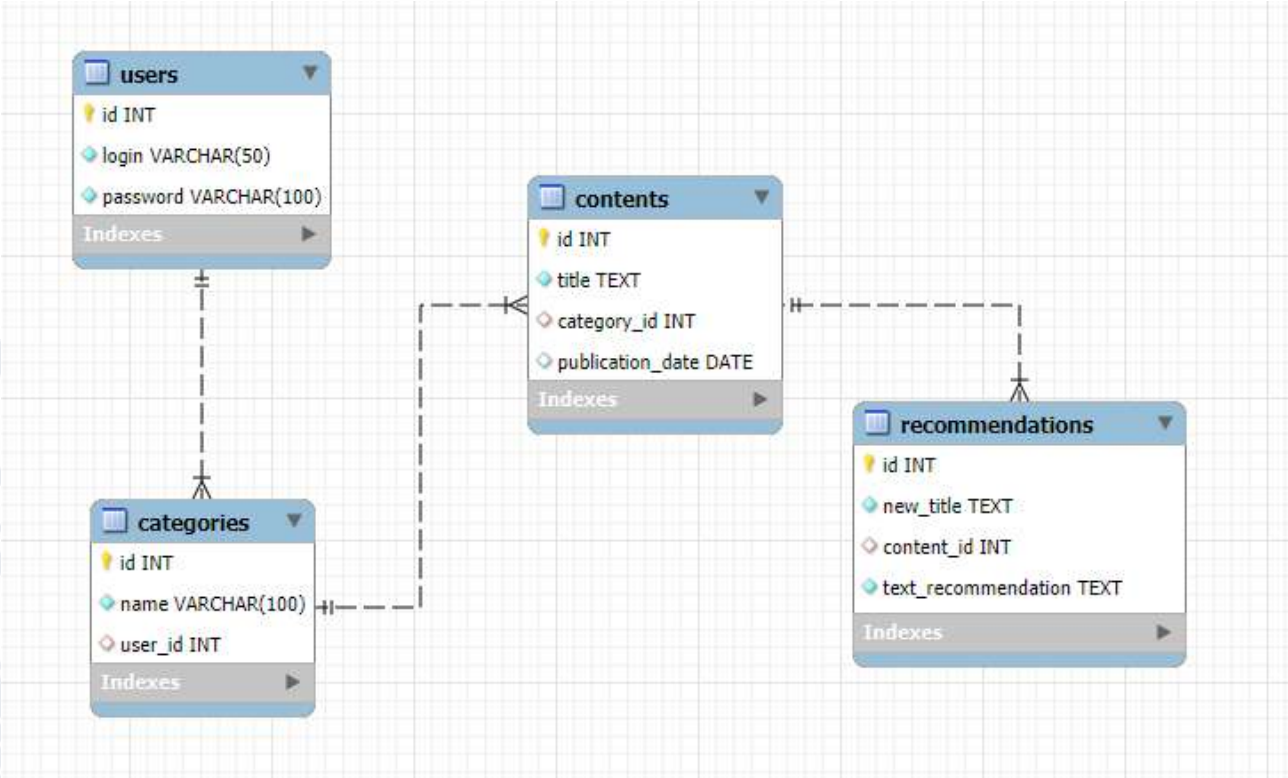

Рис. 3.4. Структура бази даних рекомендаційної системи

Як видно з рис. 3.4, в розробленій базі даних використовується найчастіше застосовуваний тип зв’язку «Один-до-багатьох», який до того ж, є основним для реалізації цілісності структури даних. Такий тип зв’язку означає, що одному запису в таблиці може відповідати від нуля до багатьох записів іншої таблиці. Водночас кожен запис у залежній таблиці пов’язаний лише з одним записом у головній таблиці. Тобто, може бути як два, так і один або навіть нуль записів, які відповідають одному запису з іншої таблиці.

Зв'язок «Один-до-багатьох» є типовим для ситуацій, коли один запис в головній таблиці має логічний зв’язок із кількома записами в іншій таблиці. Детальний опис усіх таблиць, що забезпечують функціонування бази даних показано в табл. 3.1–3.4.

Таблиця 3.1 – Таблиця «Users»

Field	Type	Null	Key	Default	Extra
id	int	NO	PRI	NULL	auto_increment
login	varchar(50)	NO	–	NULL	–
password	varchar(100)	NO	–	NULL	–

Таблиця 3.2 – Таблиця «Categories»

Field	Type	Null	Key	Default	Extra
id	int	NO	PRI	NULL	auto_increment
name	varchar(100)	NO	–	NULL	–
user_id	int	YES	MUL	NULL	–

Таблиця 3.3 – Таблиця «Contents»

Field	Type	Null	Key	Default	Extra
id	int	NO	PRI	NULL	auto_increment
title	text	NO	–	NULL	–
category_id	int	YES	MUL	NULL	–
publication_date	date	YES	–	NULL	–

Таблиця 3.4 – Таблиця «Recommendations»

Field	Type	Null	Key	Default	Extra
id	int	NO	PRI	NULL	auto_increment
new_title	text	NO	–	NULL	–
content_id	int	YES	MUL	NULL	–
text_recommendation	text	NO	–	NULL	–

Важливим аспектом архітектури є використання первинних та зовнішніх ключів, що дають змогу ефективно пов'язувати таблиці між собою та оптимізувати процеси швидкої вибірки даних. Первинний ключ відповідає за абсолютну унікальність та точну ідентифікацію записів у таблиці й має бути не NULL, тоді як зовнішній допомагає швидко знаходити дані, допускаючи можливість повторів у значеннях цих колонок. Первинний ключ автоматично індексується, що значно прискорює пошук за ним, а зовнішній ключ може потребувати додаткового індексування для прискорення з'єднання таблиць у запитах.

Правильне використання чітких первинних та зовнішніх ключів є критично важливим для якісного проєктування бази даних. Таким чином, база даних рекомендаційної системи виступає не лише як сховище інформації, але і як ключовий елемент, що забезпечує основу для точного функціонування алгоритмів автоматизованої обробки та глибокого аналізу текстового контенту.

3.3 Розробка основного функціоналу

Рекомендаційна система для автоматизованої обробки та аналізу текстового контенту є комплексним рішенням, яке не обмежується лише роботою з даними в MySQL. Система також складається зі зручного додатка з інтуїтивно зрозумілим графічним інтерфейсом користувача, який безпосередньо працює з даними. Програмна реалізація була виконана потужною мовою програмування Python. Додаток виконує такі завдання, як швидка обробка запитів, глибокий аналіз даних і генерування рекомендацій. Python – це інтерпретована високорівнева мова програмування загального призначення з відкритим вихідним кодом.

В основі Python лежить декілька парадигм програмування: об’єктноорієнтовна, імперативна, функціональна (функційна), структурна. Проте досить часто про Python говорять зокрема як про об’єктно-орієнтовну мову програмування. В об'єктно-орієнтованих мовах програмування програми будуються на основі об'єктів, які об'єднують в собі дані й функціонал. Об’єктів в мові Python багато, можна навіть сказати, що все є об’єкт (це не дуже коректно, але допустимо) [23].

Основними архітектурними особливостями є такі функції, як автоматичне управління пам’яттю, динамічне введення тексту, механізм обробки винятків, повний самоаналіз, гнучкі високорівневі структури даних та підтримка багатопотокових обчислень. Код на Python організований у класи та функції, які можна об'єднати в модулі [24]. Для значних обчислень використовуються багатоядерні можливості Python для паралельної обробки великих обсягів даних.

Для розробки додатка було обрано об'єктно-орієнтоване програмування (ООП) як основну парадигму. ООП є одним із найпоширеніших підходів у сучасному програмуванні, що забезпечує структурованість і гнучкість розробки, завдяки чому полегшується підтримка й подальше розширення масштабування функціональності системи. За допомогою ООП додаток зручно розбитий на модулі, що дозволяє легко підтримувати та розширювати систему. Ключові концепції ООП подано в табл. 3.5.

Таблиця 3.5 – Ключові концепції ООП

Концепція	Опис
Інкапсуляція	Процес приховування внутрішніх деталей об'єкта, що забезпечує захист даних від зовнішніх впливів та дозволяє взаємодіяти з об'єктом лише через визначений інтерфейс.
Успадкування	Механізм, який дозволяє створювати нові класи на основі наявних, успадковуючи їхні властивості та методи, що сприяє повторному використанню коду.
Поліморфізм	Здатність об'єктів різних класів реагувати на однакові виклики методів по-різному, даючи змогу реалізувати єдиний інтерфейс для різних типів об'єктів.
Абстракція	Процес виділення значущих характеристик об'єкта та приховування несуттєвих, що спрощує складність програмного забезпечення та фокусує увагу на суттєвих аспектах.

Ці чотири концепції разом формують основу об'єктно-орієнтованого програмування, забезпечуючи ефективну організацію, підтримку та масштабованість програмних систем. Завдяки ООП рекомендаційна система стає більш гнучкою, оскільки можна легко додавати нові модулі, розширювати функціональність та підтримувати стабільність коду.

Вибір мови програмування Python для реалізації основного функціоналу системи обумовлений її численними перевагами, особливо в контексті обробки даних, ООП, швидкого прототипування та тестування мовних моделей.

Python підтримує об’єктно-орієнтоване програмування, що дозволяє структуровано організовувати код і забезпечує легкість підтримки й масштабування. Підхід ООП в Python спрощує створення модульних компонентів, які можна використовувати повторно, а також дає змогу легко розробляти та тестувати окремі частини системи. Це особливо важливо для складних проєктів, таких як рекомендаційна система, де кожен модуль створений у вигляді окремого класу з чітко визначеним функціоналом.

Python є високорівневою мовою з простою і зрозумілою синтаксичною структурою, що значно пришвидшує розробку та прототипування. Мова дозволяє створювати робочі прототипи функціоналу та проводити їх тестування

без необхідності витрачати багато часу на низькорівневі деталі. Це важлива перевага при розробці мовних моделей, адже Python має сумісність з бібліотеками для машинного навчання, що забезпечує зручне тестування та оптимізацію алгоритмів. Модулі та бібліотеки, які підключені до додатка, показано на рис. 3.5.

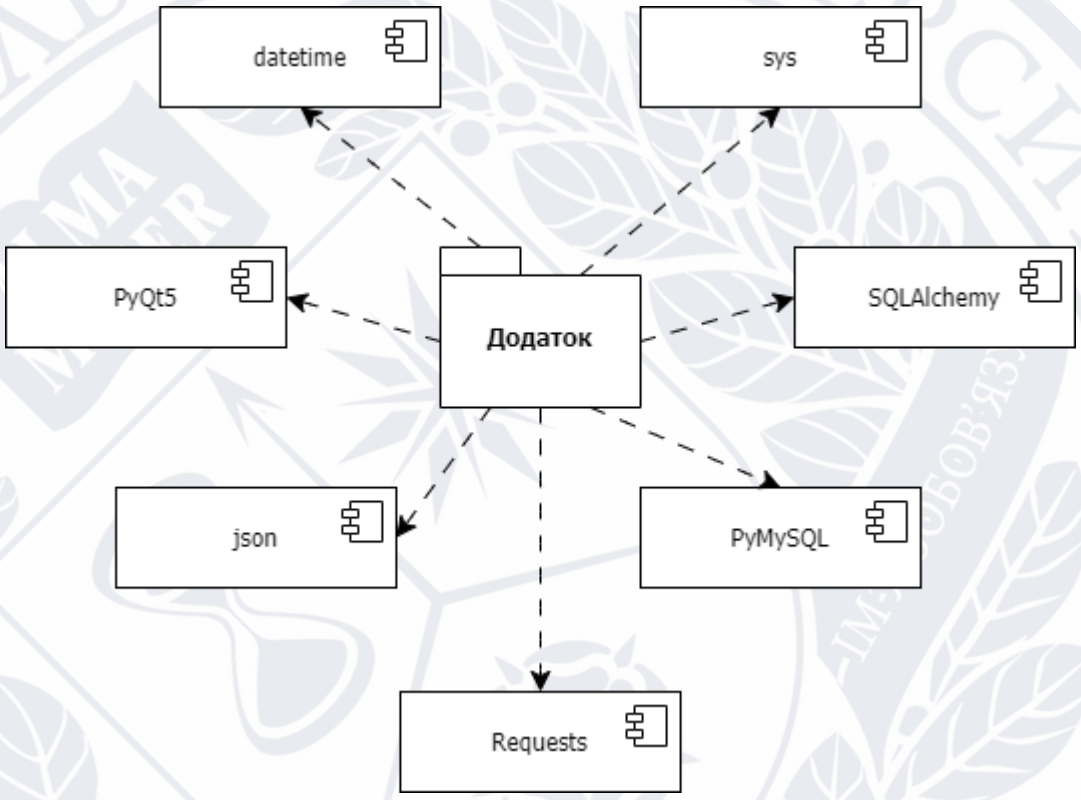

Рис. 3.5. Підключені модулі та бібліотеки Python

Опис використаних у додатку стандартних модулів подано в табл. 3.6.

Таблиця 3.6 – Стандартні модулі

Модуль	Опис
sys	Забезпечує доступ до деяких змінних і функцій, що взаємодіють з інтерпретатором Python.
json	Інструмент для роботи з JSON-даними у Python, використовується для серіалізації та десеріалізації даних.
datetime	Надає класи для опрацювання часу і дати різними способами.

Крім стандартних модулів, які встановлені за замовчуванням і забезпечують базову функціональність, додаток використовує додаткові

спеціалізовані бібліотеки, які розширюють функціональність. Перелік та опис додаткових бібліотек подано в табл. 3.7.

Таблиця 3.7 – Додаткові бібліотеки

Бібліотека	**Опис**
SQLAlchemy	ORM-бібліотека, що забезпечує роботу з базами даних на основі реляційних моделей.
PyMySQL	Бібліотека для підключення та виконання запитів до бази даних MySQL з Python.
Requests	Бібліотека для здійснення HTTP-запитів, що спрощує взаємодію з API.
PyQt5	Бібліотека для створення графічного інтерфейсу користувача, що інтегрується з фреймворком Qt й забезпечує зв’язок з Python.

Використані модулі та спеціалізовані бібліотеки взаємодоповнюють одна одну, забезпечуючи стабільну та ефективну роботу додатка.

Бібліотека SQLAlchemy є інструментом для роботи з базою даних у додатку, що забезпечує об'єктно-реляційне відображення. Вона дозволяє взаємодіяти з базою даних за допомогою Python-класів та об'єктів, перетворюючи їх у таблиці, рядки, і поля в реляційній базі даних. Це значно спрощує роботу з SQL-запитами, оскільки замість написання SQL-коду для виконання операцій, таких як створення, читання, оновлення та видалення даних, можна використовувати Python-код, який буде автоматично конвертований у SQL.

Бібліотека SQLAlchemy також включає систему керування сесіями для управління транзакціями та підтримує гнучкі способи створення зв’язків між таблицями. Це робить її потужним інструментом для роботи з реляційними базами даних в даній системі.

Бібліотека PyMySQL виконує функцію драйвера для підключення до бази даних MySQL через SQLAlchemy. Вона є чисто Python-реалізацією клієнта MySQL, що не вимагає встановлення додаткових бінарних модулів. Оскільки SQLAlchemy є ORM і може працювати з різними базами даних, йому потрібен конкретний драйвер для зв’язку з MySQL, і саме цю функцію виконує PyMySQL.

Бібліотека Requests використовується для посилання HTTP-запитів до OpenAI API. Вона виконує функцію зв'язку між локальним Python-скриптом і сервером OpenAI API, що дозволяє відправляти дані (в нашому випадку текст і параметри для генерації експертних рекомендацій) та отримувати відповіді (текст рекомендацій або виправлений текст) від сервера. Бібліотека підтримує різні методи HTTP, такі як GET, POST, PUT, DELETE, що надає можливість гнучкого налаштування запитів. Таким чином, бібліотека requests дозволяє автоматизувати процес надсилання HTTP-запитів до API та отримувати відповіді, які далі можна обробити в програмі. Вона також забезпечує підтримку налаштування заголовків, параметрів запиту та автентифікації, що є важливим для роботи з API, які вимагають авторизації. Таким чином, бібліотека Requests полегшує взаємодію з API, надаючи інструменти для відправлення запитів та отримання відповідей.

Додаток розроблений з використанням графічного фреймворку Qt та бібліотекою PyQt5, яка забезпечує можливість створення графічного інтерфейсу користувача мовою Python. Фреймворк Qt надає широкий набір інструментів для створення адаптивних та інтуїтивно зрозумілих інтерфейсів, завдяки чому додаток легко налаштовується під вимоги користувача. PyQt5 забезпечує доступ до функціоналу Qt, даючи змогу використовувати різні модулі.

Для забезпечення необхідної функціональності додатка використано основні модулі фреймворку Qt, що містять класи та функції для створення елементів графічного інтерфейсу і роботи з даними, які наведені в табл. 3.8.

Таблиця 3.8 – Основні модулі фреймворку Qt

Модуль	**Опис**
QtCore	Ядро фреймворку.
QtGUI	Компоненти для створення інтерфейсів.
QtWidgets	Модуль для роботи з віджетами.

Ознайомившись з основними модулями фреймворку Qt, які забезпечують широкий функціонал для розробки сучасних програмних рішень, доцільно перейти до аналізу графічного інтерфейсу користувача, реалізованого в цьому

додатку. Одним із ключових елементів графічного інтерфейсу додатка є механізм автентифікації для забезпечення безпеки даних. Оскільки захист даних є однією з функціональних можливостей інформаційних систем, рекомендаційна система вимагає обов'язкової автентифікації користувача під час входу, що забезпечує доступ лише авторизованим користувачам (рис. 3.6).

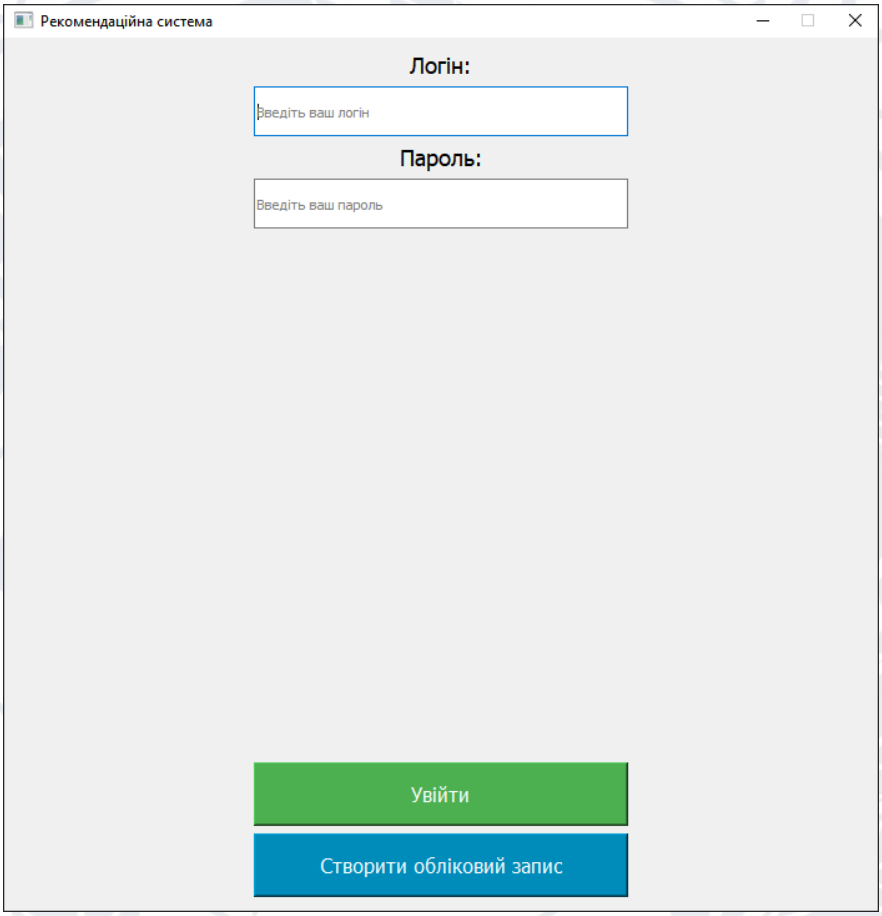

Рис. 3.6. Вікно автентифікації

Автентифікація у галузі інформаційних технологій – це перевірка достовірності користувача інформаційної системи через порівняння введеного ним пароля з паролем, збереженим у базі даних користувачів. Позитивним результатом автентифікації є авторизація користувача інформаційної системи, тобто надання йому прав доступу до ресурсів, необхідних для виконання його завдань. У такий спосіб процес автентифікації створює додатковий рівень безпеки, запобігаючи несанкціонованому доступу та захищаючи конфіденційну

інформацію від можливих загроз. Це дозволяє ідентифікувати користувачів та контролювати доступ до ресурсів, забезпечуючи цілісність даних.

Після успішного проходження автентифікації користувач отримує доступ до функціоналу рекомендаційної системи, що дозволяє виконувати подальші дії, пов'язані з введенням, обробкою та аналізом даних. Потім користувач переходить до вікна створення категорій для тексту, де може налаштувати їх відповідно до тематики або цілей аналізу, забезпечуючи структурування контенту для подальшого ефективного опрацювання системою (рис. 3.7).

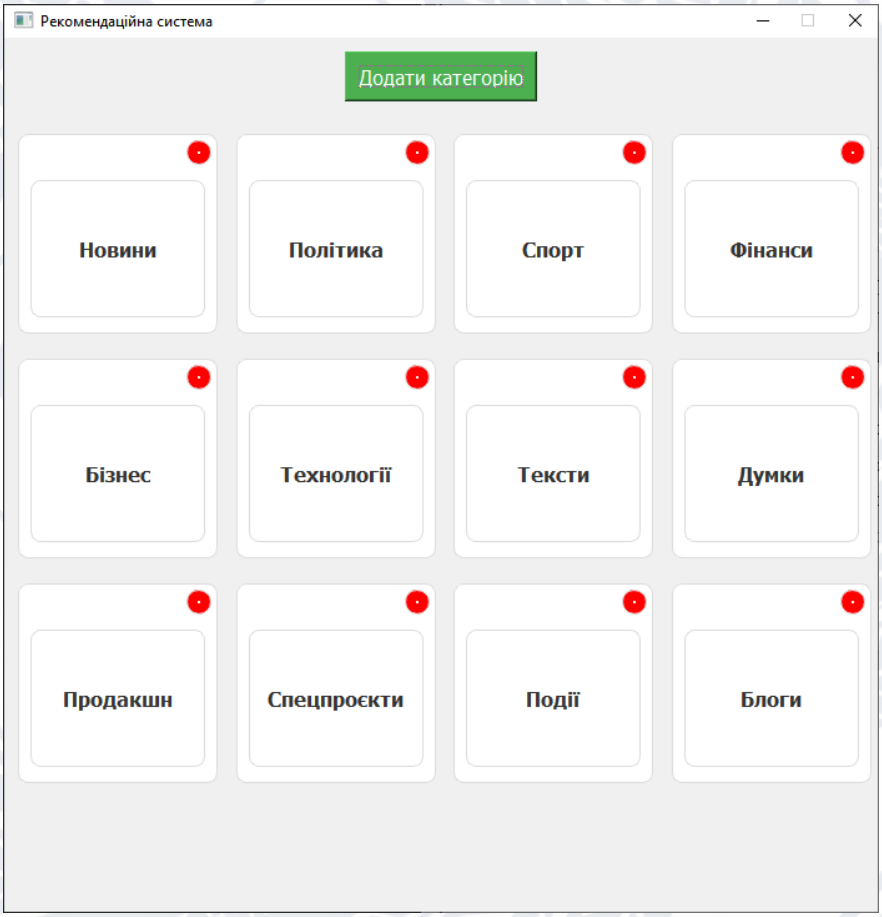

Рис. 3.7. Налаштування категорії текстового контенту

Як видно з рисунка 3.7, користувач може додати нову категорію або видалити вже наявну, що забезпечує гнучкість і зручність у процесі управління категоріями текстового контенту. Така функціональність дозволяє налаштовувати категорії відповідно до індивідуальних потреб користувача, що є важливим аспектом персоналізації та спрощує доступ до потрібних даних у

системі. Таке структурування тексту дозволяє отримати точніші результати аналізу та спростити пошук інформації за ключовими темами, підвищуючи якість рекомендацій. Розподіл тексту на категорії також сприяє адаптації системи під специфічні потреби користувача, що робить процес обробки персоналізованим та результативним. Алгоритм класифікації працює на основі попередньо визначених правил, дозволяючи адаптувати систему під різні типи тексту.

Після створення та обрання найбільш відповідної категорії користувач переходить до головного вікна, де він може ввести або вставити текст, з яким будуть виконуватися наступні автоматизовані дії, пов’язані з обробкою та аналізом (рис. 3.8).

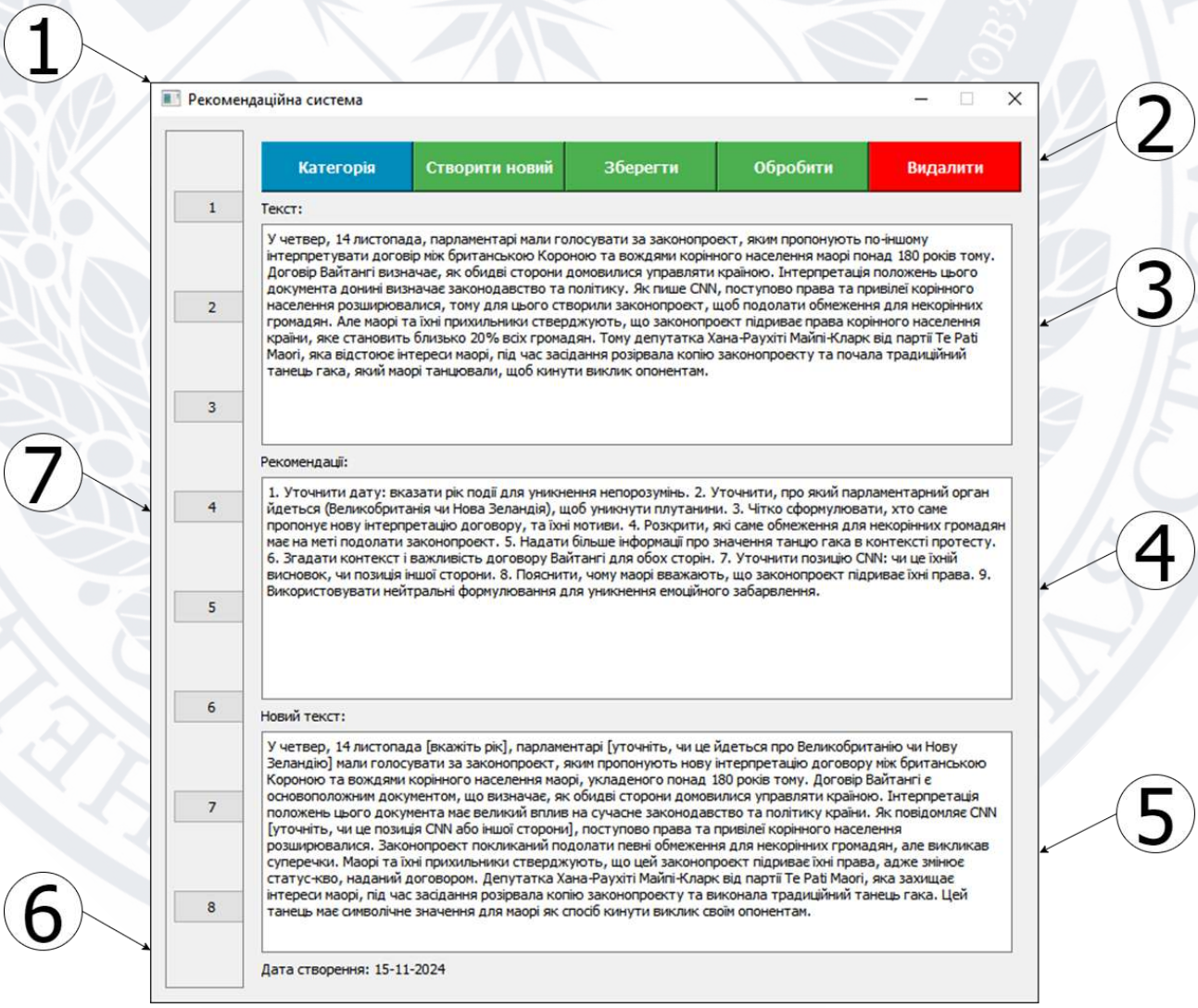

Рис. 3.8. Головне вікно рекомендаційної системи

Як видно з рис. 3.8, головне вікно рекомендаційної системи є ключовим елементом користувацького інтерфейсу, яке забезпечує взаємодію користувача із системою, надаючи доступ до її функціональних можливостей. У табл. 3.9 наведено перелік елементів інтерфейсу, що формують структуру та функціонал цього вікна, а також опис їх призначення.

Таблиця 3.9 – Елементи інтерфейсу головного вікна

№	Елемент інтерфейсу	Опис
1	Заголовок вікна	Рядок безпосередньо під верхньою межею вікна, що містить назву вікна.
2	Панель інструментів	Це ряд кнопок у верхній частині екрана, які мають свою функцію.
3	Необроблений текст	Поле для введення чи вставки тексту, який буде оброблятися.
4	Згенеровані рекомендації	Секція, де система надає свої експертні пропозиції з покращення тексту.
5	Згенерований текст	Поле для відображення варіанту нового автоматизовано обробленого тексту.
6	Дата створення	Інформація про дату виконання операції.
7	Історія	Збереження всіх проміжних та фінальних версій тексту після кожного етапу обробки, дозволяючи відстежувати зміни, порівнювати варіанти та повертатися до попередніх версій за потреби.

Історія в рекомендаційній системі виконує функцію збереження всіх проміжних та фінальних версій тексту після кожного етапу обробки. Це дає користувачеві можливість відстежувати всі зміни, порівнювати різні варіанти тексту, які були згенеровані під час процесу обробки, і, за потреби, повернутися до будь-якої з попередніх версій. Історія також зберігає всі рекомендації, які система пропонувала на кожному етапі, що допомагає краще розуміти, які саме покращення були запропоновані та як вони вплинули на кінцевий результат. Завдяки цьому користувач отримує повний контроль над редагуванням тексту та може зберігати найвдаліші версії для подальшого використання або аналізу.

Історія відіграє роль не лише архіву, але і дозволяє користувачам аналізувати ефективність внесених змін. Це забезпечує зворотний зв’язок про вплив кожної рекомендації та дозволяє оцінювати якість системи.

Таким чином, основний функціонал розробленої рекомендаційної системи забезпечує високий рівень автоматизації та зручності, поєднуючи обробку тексту з аналізом. Завдяки гнучкій модульній структурі система дозволяє легко додавати нові функціональні можливості або вдосконалювати наявні модулі, зберігаючи стабільність та ефективність роботи.

Вибір потужної мови програмування Python для реалізації функціоналу дозволив ефективно оптимізувати обробку тексту за допомогою об’єктно-орієнтованого підходу, що сприяє високій структурованості та гнучкій можливості подальшого розширення функцій. Використання об’єктно-орієнтованого підходу сприяло раціональній оптимізації алгоритмів обробки даних, що, своєю чергою, створює перспективні передумови для подальшого розвитку і застосування розробки в різних галузях, включаючи динамічну медіаіндустрію, аналіз великих даних та інші сфери автоматизації інформаційних процесів.

3.4 Розробка конфігурації

Конфігурація рекомендаційної системи є критично важливим етапом розробки, адже саме на цьому етапі встановлюються основні параметри та налаштування, які значно впливають на продуктивність системи, її здатність адаптуватися до різних обсягів і типів даних, а також на точність рекомендацій. Особливо це стосується систем, що обробляють текстовий контент, де вибір та налаштування мовної моделі є вирішальними.

Рекомендаційна система для автоматизованої обробки та аналізу текстового контенту використовує конфігурацію мовної моделі GPT, окрім MySQL та додатка з графічним інтерфейсом користувача, для забезпечення автоматизованої обробки, високої точності аналізу та генерації рекомендацій. GPT (Generative Pre-trained Transformer) – це нейронна мережа, спеціально

розроблена як мовна модель. Тобто, це нейронна мережа, навчена для роботи з мовними даними. Конфігурація мовної моделі GPT є ключовим елементом системи для забезпечення автоматизованої обробки тексту, а також генерації експертних рекомендацій на основі контексту.

У своїй основі GPT є глибокою нейронною мережею, яка складається з багатьох шарів і мільярдів параметрів. Це специфічний тип нейронної мережі, званий трансформером, який був розроблений спеціально для обробки послідовних даних, таких як текст. Даний тип використовує архітектуру трансформер для обробки та генерації тексту. Така архітектура базується на ідеї самоуважності, що дозволяє моделі обробляти вхідні послідовності паралельно, а не послідовно. Це дає змогу моделі фіксувати довгострокові залежності та зв’язки між словами в речення, що є важливим для створення зв’язного тексту. Архітектура трансформера складається з шарів нейронів, що виконують обробку даних через зважені з'єднання, що відповідає базовій концепції нейромереж.

GPT є саме мовною моделлю, оскільки вона натренована на величезних обсягах текстових даних і здатна передбачати наступне слово в реченні, враховуючи попередні слова. Завдяки цьому GPT здатна генерувати зрозумілий текст, відповідати на питання, вести діалог і виконувати багато інших завдань, пов'язаних з розумінням мови.

Для інтеграції мовної моделі у рекомендаційну систему використовується прикладний програмний інтерфейс, який надає доступ до технологій ШІ, розроблений і підтримуваний дослідницькою лабораторією OpenAI.

Саме OpenAI API дозволяє інтегрувати передову мовну модель, таку як GPT-4o, безпосередньо в рекомендаційну систему. GPT-4o – це найдосконаліша та найшвидша мовна модель сімейства GPT-4 з найбільшою кількістю токенів, яку можна використати для прямих запитів до OpenAI API. Однією з ключових особливостей GTP-4o для рекомендаційної системи є її здатність генерувати зв’язний і схожий на людину текст. Це досягається шляхом використання механізмів уваги, які дають змогу моделі враховувати контекст і зв’язки між словами в реченні.

На додаток до самоуважності, мовні моделі на основі трансформаторів також використовують інші методи, такі як моделювання замаскованої мови та динамічне керування. Моделювання замаскованої мови передбачає маскування частини вхідного тексту та передбачення відсутніх слів на основі контексту, наданого незамаскованими словами. Це допомагає моделі навчитися розуміти зв'язки між словами та створювати текст, який є зв'язним і має сенс у контексті. Динамічний контроль, з іншого боку, дозволяє моделі регулювати рівень деталізації та конкретності вихідних даних на основі вхідних даних. Це дозволяє моделі генерувати текст, який підходить для даного контексту та завдання [25].

У порівнянні з попередніми моделями та іншими аналогічними рішеннями, GPT-4o демонструє вищу продуктивність завдяки розширеному контекстному вікну та можливості обробляти більші обсяги даних у реальному часі. Вона також відзначається здатністю працювати з багатомовними текстами й надавати високу якість відповіді навіть у складних мовних завданнях, що робить її ідеальним вибором для використання в рекомендаційній системі.

Мова програмування Python є найкращим вибором для розробки й тестування обраної мовної моделі завдяки сумісності з бібліотеками глибинного навчання та обробки природної мови. Ці інструменти полегшують процес інтеграції моделей, таких як GPT, та дають можливість гнучко налаштовувати та оптимізувати їх відповідно до потреб конкретної системи. Python також підтримує багатопоточність та роботу з великими масивами даних, що підвищує ефективність роботи з мовною моделлю.

Для досягнення високоякісного результату роботи мовної моделі важливо не лише налаштувати параметри її конфігурації у системі, але й забезпечити чіткі та детально продумані інструкції для виконання завдань. Правильне налаштування промптів та адаптація контексту допомагають моделі краще орієнтуватися у структурі завдання, а також розуміти конкретні вимоги до формату відповіді. Формулювання промптів відіграє критичну роль у процесі, адже саме точність і відповідність інструкцій безпосередньо впливають на якість та релевантність результатів.

Даючи моделі точні інструкції, приклади та необхідну контекстну інформацію, можна підвищити якість і точність результатів роботи моделі. Завдяки таким підходам, модель отримує більше контексту для точного виконання завдання. Процес створення промпту для отримання правильного результату від мовної моделі полягав у формулюванні чітких інструкцій, які допоможуть їй зрозуміти контекст та бажаний формат відповіді. Цей підхід дозволяє зменшити можливість неоднозначних чи неправильних інтерпретацій запиту. Формулювання чіткого і структурованого промпту включало кілька ключових етапів: визначення мети запиту, надання відповідного контексту, вказівка бажаного формату відповіді, а також, прикладів правильно сформульованих відповідей. Також важливим аспектом було налаштування тональності відповіді. У системі тональність взаємодії є нейтрально-дружньою, з акцентом на професійність та універсальність.

Загалом, правильна конфігурація мовної моделі GPT-4o та її інтеграція у рекомендаційну систему дозволила забезпечити високий рівень автоматизованої обробки, точності аналізу та генерації експертних рекомендацій, що є ключовим для функціонування системи.

3.5 Оцінка ефективності розробленої рекомендаційної системи

Оцінка ефективності розробленої рекомендаційної системи є важливим етапом дослідження, що дозволяє визначити її здатність розв’язувати поставлені задачі з автоматизованої обробки та аналізу текстового контенту. Без ґрунтовної оцінки неможливо підтвердити переваги системи над наявними рішеннями та забезпечити її надійність у практичному застосуванні. Така оцінка не лише підкреслює значущість розробки, але й сприяє її подальшому вдосконаленню.

Окрім кількісних показників, важливо враховувати якісні характеристики, такі як зручність використання системи для кінцевих користувачів та інтуїтивність її інтерфейсу. Основною метою цього підпункту є порівняння результатів роботи рекомендаційної системи з наявними методами обробки та аналізу тексту, зокрема оцінка точності та релевантності наданих рекомендацій.

Порівняння роботи рекомендаційної системи з іншими методами аналізу текстового контенту дозволяє оцінити її ефективність у контексті конкурентних рішень.

Ключові метрики, які були обрані для проведення оцінки ефективності розробленої рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту наведені в табл. 3.10.

Таблиця 3.10 – Ключові метрики

Метрика	**Опис**
Швидкість обробки даних	Показує час, необхідний для аналізу та надання рекомендацій системою у порівнянні з традиційними методами.
Якість аналізу даних	Оцінює здатність системи виявляти основні аспекти текстового контенту для глибшого розуміння тексту.
Кількість релевантних рекомендацій	Відображає кількість корисних рекомендацій серед загального набору даних, що були оброблені системою.
Релевантність рекомендацій	Визначає відповідність рекомендованих матеріалів інтересам та потребам користувача.
Точність рекомендацій	Відображає частку правильних рекомендацій серед усіх запропонованих.

Розрахунок ефективності розробленої рекомендаційної системи здійснювався на основі даних, отриманих від ТОВ «РЕДАКЦІЯ ВІННИЦЬКОЇ РЕГІОНАЛЬНОЇ ГАЗЕТИ „ПОДІЛЬСЬКА ЗОРЯ“», що уможливило дослідити реальні умови функціонування системи в межах редакційної діяльності. Це дозволило забезпечити практичну значущість отриманих результатів.

Для оцінки ефективності обробки розроблена рекомендаційна система була протестована з іншими наявними методами обробки текстового контенту, зокрема ручною, напівавтоматизованою та автоматизованою на великих обсягах даних, аби досягати найбільш реалістичного сценарію. Порівняння включало швидкість обробки великих обсягів тексту, що дозволило оцінити переваги та недоліки кожного підходу в різних сценаріях застосування.

Середній час обробки одного тексту для кожного методу визначали за формулою:

$$T_{avg} = \frac{T_{total}}{N}, \tag{3.1}$$

де T_{avg} – середній час, необхідний для обробки одного тексту при використанні певного методу;

T_{total} – загальний час обробки всіх текстів для певного методу;

N – загальна кількість текстів, які були оброблені.

$$T_{avg,\text{ручний}} = \frac{167\ \text{год}}{1000} = \frac{0.167\ \text{год}}{\text{текст}} \approx \frac{10\ \text{хв}}{\text{текст}};$$

$$T_{avg,\text{напівавтоматизований}} = \frac{117\ \text{год}}{1000} = \frac{0.117\ \text{год}}{\text{текст}} \approx \frac{7\ \text{хв}}{\text{текст}};$$

$$T_{avg,\text{автоматизований}} = \frac{67\ \text{год}}{1000} = \frac{0.067\ \text{год}}{\text{текст}} \approx \frac{4\ \text{хв}}{\text{текст}};$$

$$T_{avg,\text{рекомендаційна система}} = \frac{2.8\ \text{год}}{1000} = \frac{0.0028\ \text{год}}{\text{текст}} \approx \frac{10\ \text{с}}{\text{текст}}.$$

Прискорення відносно ручного методу обробки визначали за формулою:

$$a = \frac{T_{\text{ручний}}}{T_{\text{новий}}}, \tag{3.2}$$

де a – коефіцієнт прискорення, що показує, у скільки разів новий метод швидший;

$T_{\text{ручний}}$ – загальний час обробки ручним методом;

$T_{\text{новий}}$ – загальний час обробки новим методом.

Цей підхід ілюструє ефективність кожного нового методу обробки порівняно з ручним методом. Чим більший коефіцієнт a, тим суттєвіше новий метод зменшує час обробки текстів.

$$a_{\text{напівавтоматичний}} = \frac{10 \text{ год}}{5 \text{ год}} = 2;$$

$$a_{\text{автоматичний}} = \frac{10 \text{ год}}{1 \text{ год}} = 10;$$

$$a_{\text{рекомендаційна система}} = \frac{10 \text{ год}}{0.5 \text{ год}} = 20.$$

Аби оцінити точність та експертність наданих рекомендацій з покращення тексту розроблена рекомендаційна система була протестована в порівнянні з рекомендаціями, наданими групою експертів, аби визначити, наскільки рекомендації системи відповідають експертним оцінкам та чи може вона слугувати ефективним інструментом для покращення текстів у практичному застосуванні. Для цього було обрано кілька текстів різної складності, які потребували вдосконалення, та проаналізовано якість змін, запропонованих як системою, так і експертами. Порівняння проводилося за такими критеріями, як відповідність стилю, логіка та структурованість тексту, зменшення граматичних, лексичних і стилістичних помилок, а також загальна читабельність.

У процесі тестувань було встановлено:

1. Середній час обробки одного тексту рекомендаційною системою суттєво зменшився порівняно з іншими наявними методами обробки, зокрема ручною, напівавтоматизованою та автоматизованою.
2. Прискорення обробки рекомендаційною системою у 20 разів порівняно з ручним методом свідчить про значну ефективність.
3. Рекомендації з покращення тексту, надані системою, збігалися з експертними висновками у 89% випадків, підтверджуючи експертність та надійність системи.

Для оцінки масштабованості проводилися ретельні аналізи продуктивності, включаючи метрики часу відповіді, пропускної здатності, а також оцінку точності й повноти рекомендацій. Завдяки тестам навантаження та стресу, система показала свою здатність адаптуватися до дедалі більшого навантаження, зберігаючи при цьому високу якість результатів.

Таким чином, рекомендаційна система виявилася потужним інструментом для автоматизації обробки та аналізу тексту, здатною підвищити якість контенту завдяки точним і експертним рекомендаціям. Це підкреслює її значущість для впровадження у практичних умовах, особливо в медіаіндустрії та інших галузях, де обробка великих обсягів тексту є критично важливою.

ВИСНОВКИ

У процесі виконання кваліфікаційної роботи створено рекомендаційну систему для автоматизованої обробки та аналізу текстового контенту, спрямовану на підвищення його якості.

Перед початком безпосереднього процесу розробки було розглянуто теоретичні аспекти процесів обробки та аналізу текстового контенту. Проведено огляд та аналіз наявних підходів до обробки та аналізу текстового контенту, виявлені переваги та недоліки різних методів. Узагальнено технологічний процес створення рекомендаційної системи: поняття, головна мета, напрямки відповідальності, можливості застосування та переваги використання.

Обґрунтовано необхідність розробки рекомендаційної системи для медіаіндустрії. Наведено переваги рекомендаційної системи у порівнянні з традиційними підходами. Виявлено основні виклики та проблеми у розробці рекомендаційної системи.

Після огляду всієї необхідної інформації, розроблено рекомендаційну систему з використанням мовної моделі GPT-4o на основі штучного інтелекту, яка відповідає усім заявленим вимогам. Теоретичні напрацювання та результати впроваджено на практиці.

Проведена оцінка ефективності розробленої рекомендаційної системи шляхом порівняння з іншими наявними методами обробки й аналізу текстового контенту, а також у порівнянні з рекомендаціями, наданими групою експертів, продемонструвала, що система показує себе вкрай результативно.

Отже, доведено основну гіпотезу дослідження, яка полягає в тому, що впровадження комп’ютерних технологій обробки даних, зокрема рекомендаційної системи для автоматизованої обробки та аналізу текстового контенту, дає змогу швидше обробляти великі обсяги тексту з високою точністю та надавати експертні рекомендації, які відповідають найкращим стандартам досліджуваної галузі, що загалом сприяє підвищенню якості контенту в медіаіндустрії.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ

1. Підруцький Д. А., Січко Т. В. Big Data в обробці і аналізі даних з різних сфер. Комп'ютерні технології обробки даних : матеріали IV Всеукр. наук.-практ. конф., Вінниця, 2023. С. 183–187.

2. Ланде Д. В., Субач І. Ю., Гладун А. Я. Оброблення надвеликих масивів даних (Big Data) : навч. посіб. Київ, 2021. 168 с.

3. Смерічевська С. В., Савченко Л. В., Автомонов О. А. Бізнес-аналіз і обробка даних : методичні рекомендації до самостійної роботи. К. : НАУ, 2024. 34 с.

4. Кобзев І. В., Кобзева Н. М., Тесленко О. В. Технології роботи з Big Data: методичні рекомендації. Х. : ХНЕУ, 2023. 83 с.

5. Гороховатський В. О., Творошенко І. С. Методи інтелектуального аналізу та оброблення даних : навч. посіб. Х. : ХНУРЕ, 2021. 92 с.

6. Баранов О. А. Визначення терміну «штучний інтелект». Інформація і право. 2023. С. 32–49.

7. Кучеров Д. П. Методи аналізу великих даних: навч. посіб. Київ, 2020. 237 с.

8. Зелінська О. В., Потапова Н. А., Волонтир Л. О. Інформаційні системи та технології в галузі : навч. посіб. Вінниця : ВНАУ, 2020. 253 с.

9. Рашкевич Н. В., Отрош Ю. А. Методологія та організація наукових досліджень: навч. посіб. Харків, 2022. 291 с.

10. Проценко Н. М. Економічна інформатика : навч. посіб. Харків, 2020. 212 с.

11. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань : навч. посіб. Львів, 2021. 276 с.

12. Іванчук Я. В., Месюра В. І., Яровий А. А., Манжілевський О. Д. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних. Вінниця : ВНТУ, 2021. 69 с.

13. Нескородєва Т. В., Федоров Є. Є., Січко Т. В., Нескородєва А. Р. Експертні та рекомендаційні системи : навч. посіб. Вінниця : ДонНУ імені Василя Стуса, 2023. 224 с.

14. Субботін С. О. Нейронні мережі: теорія та практика : навч. посіб. Житомир : Вид. О. О. Євенок, 2020. 184 с.

15. Берко А. Ю., Товма Н. В., Пушкаренко С. А. Основи великих даних : методичні аспекти навчання. Вінниця : ВНТУ, 2021. 120 с.

16. Ляшенко В. І., Шевчук М. О. Технології великих даних : концепції, інструменти, застосування. Харків : ХНУРЕ, 2022. 142 с.

17. Іванов О. А., Васильєв А. В. Аналіз даних великих обсягів: практичний підхід. Київ : КПІ ім. Ігоря Сікорського, 2020. 98 с.

18. Степанюк О. С., Січко Т. В. Особливості використання реляційних та нереляційних баз даних в Big Data. Комп'ютерні технології обробки даних : матеріали Всеукр. наук.-практ. конф., Вінниця, 2020. С. 103–106.

19. Якимчук В. С., Демиденко О. В. Візуалізація великих даних : теорія та практика. Вінниця : ВНАУ, 2023. 88 с.

20. Сухоруков А. І., Дяченко О. В. Big Data: основи, інструменти та застосування: навч. посіб. Київ : НАУ, 2022. 180 с.

21. Ковтун Т. В., Матвієнко І. С. Обробка великих даних у бізнес-аналітиці : навч. посіб. Львів : ЛНУ ім. І. Франка, 2021. 144 с.

22. Шевченко В. П., Головко П. О. Методи аналізу великих даних : навч. посіб. Харків : ХНУ, 2023. 128 с.

23. Пилипенко О. М., Кузьменко Т. В., Пашенко Ю. В. Застосування Big Data у сучасних інформаційних системах. Дніпро : ДНУ ім. Олеся Гончара, 2022. 156 с.

24. Щербань В. Ю., Краснитський С. М., Астістова Т. І., Яхно В. М. Методи представлення, збереження та аналізу даних інформаційних систем. К. : ТОВ «Фастбінд Україна», 2023. 472 с.

25. Голубенко О. І., Підмогильний О. О. Generative pre-trained transformer. Науковий журнал «IT Synergy» Київ, 2022. С. 19-27.

26. Ткаченко О. І., Золотарьова Л. В. Вступ до технологій Big Data: навч. посіб. Дніпро : ДНУ, 2020. 96 с.

27. Морозова Н. А., Литвиненко Ю. О. Хмарні технології у Big Data : навч. посіб. Одеса : ОНПУ, 2021. 150 с.

28. Алексюк В. В., Січко Т. В. Дослідження особливостей програмного забезпечення для аналізу даних. Комп'ютерні технології обробки даних : матеріали Всеукр. наук.-практ. конф., Вінниця, 2022. С. 249–251.

29. Олещенко Л. М. Технології оброблення великих даних. Комп'ютерний практикум : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2021. 85 с.

30. Гуменюк К. В., Січко Т. В. Основи машинного навчання. Комп'ютерні технології обробки даних : матеріали Всеукр. наук.-практ. конф., Вінниця, 2022. С. 142–144.

31. Андрійченко К. А., Січко Т. В. Аналіз даних як складова інформаційних технологій. Комп'ютерні технології обробки даних : матеріали Всеукр. наук.-практ. конф., Вінниця, 2020. С. 8–10.

32. Зелінський О. О., Січко Т. В. Огляд та особливості інтелектуальних систем. Комп'ютерні технології обробки даних : матеріали III Всеукр. наук.-практ. конф., Вінниця, 2022 С. 149-152.

33. Мельник В. Р., Січко Т. В. Використання методів оптимізації для обробки та аналізу великих даних. Комп'ютерні технології обробки даних : матеріали IV Всеукр. наук.-практ. конф., Вінниця, 2023 С. 88–90.

34. Гриценко Г. Г., Бондаренко Б. Б. Рекомендаційні системи : методи та алгоритми : навч. посіб. Дніпро : ДНУ, 2023. 200 с.

35. Ткаченко Т. Т. Практикум з розробки рекомендаційних систем. Київ : КПІ ім. Ігоря Сікорського, 2024. 150 с.

36. What Is Artificial Intelligence (AI)? URL: https://www.ibm.com/topics/artificial-intelligence (дата звернення 11.11.2024).

37. What is artificial intelligence and how is it used? URL: https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used (дата звернення 11.11.2024).

38. OpenAI API: Overview. URL: https://platform.openai.com/docs/overview (дата звернення: 11.11.2024).

39. GPT-4o від OpenAI: мультимодальна модель ШІ, що трансформує взаємодію людини та машини. URL: https://www.unite.ai/uk/openais-gpt-4o-the-multimodal-ai-model-transforming-human-machine-interaction/ (дата звернення: 11.11.2024).

40. OpenAI's GPT-4o is the biggest GPT-4 update yet – and it's free URL: https://www.xda-developers.com/openai-gpt-4o-spring-update-announcement/ (дата звернення: 11.11.2024).

41. Троцько В. В. Методи штучного інтелекту: навчально-методичний і практичний посібник. Київ, 2020. 86 с.

42. Когут Ю. І. Штучний інтелект і безпека : практ. посіб. за ред. д.т.н., проф. Довгополого А. С. Київ, 2024. 294 с.

43. Кублій Л. І. Алгоритми та структури даних. Основи алгоритмізації. Київ : КПІ ім. Ігоря Сікорського, 2022. 528 с.

44. В. Ю. Щербань, С. М. Краснитський, Т. І. Астістова, В. М. Яхно. Методи представлення, збереження та аналізу даних інформаційних систем : колективна монографія. Київ, 2023. 473 с.

45. Шаров С.В. Сучасний стан розвитку штучного інтелекту та напрямки його використання : зб. наук. пр. Інноваційні обрії України. 2023. С. 136-144.

46. Величко О. М., Гордієнко Т. Б. Інтелектуальні інформаційні системи: структура і застосування : навч. посіб. Київ, 2022. 728 с.

47. Литвин В. В., Пасічник В. В., Яцишин Ю. В. Інтелектуальні системи : підручник. Львів, 2024. 364 с.

48. Олещенко Л. М. Машинне навчання : навч. посіб. Київ: КПІ ім. Ігоря Сікорського, 2019. 200 с.

49. Гладун А. Я., Рогушина Ю. В. Data Mining: пошук знань в даних : навч. посіб. Київ, 2023. 452 с.

50. Кононова К. Ю. Машинне навчання: методи та моделі : навч. посіб. Харків : ХНУ імені В. Н. Каразіна, 2020. 301 с.

ДЕКЛАРАЦІЯ

про дотримання академічної доброчесності

Я, _____

_____

_____

_____

_____

Повністю вказується ПІБ та статус (освітня (освітньо-наукова) програма – для здобувачів вищої освіти, назва кваліфікаційної роботи)

що нижче підписалась/підписався, розуміючи та підтримуючи загальновизнані засади справедливості, доброчесності та законності,

ЗОБОВ'ЯЗУЮСЬ:

дотримуватися принципів та правил академічної доброчесності, що визначені законодавством України, локальними нормативними актами Донецького національного університету імені Василя Стуса, положеннями, правилами, умовами, визначеними іншими суб'єктами, та не допускати їх порушення.

ПІДТВЕРДЖУЮ:

що мені відомі положення статті 42 Закону України «Про освіту»;

що у даній роботі не представляла/представляв чиїсь роботи повністю або частково як свої власні. Там, де я скористалася/скористався працею інших, я зробила/зробив відповідні посилання на джерела інформації;

що дана робота не передавалась іншим особам і подається вперше, не порушує авторських та суміжних прав закріплених статтями 21-25 Закону України «Про авторське право та суміжні права», а дані та інформація не отримувались в недозволений спосіб.

УСВІДОМЛЮЮ:

що ця робота може бути перевірена університетом на плагіат або інші порушення академічної доброчесності, в тому числі з використанням спеціалізованих сервісів;

що у разі порушення академічної доброчесності, до мене можуть бути застосовані процедури, передбачені законодавством України та Кодексом академічної доброчесності та корпоративної етики Донецького національного університету імені Василя Стуса, іншими локальними нормативними актами університету, та я можу бути притягнута/притягнутий до академічної відповідальності.

_____ _____

(дата) (підпис)