

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ПОЛІСЬКИЙ НАЦІОНАЛЬНИ УНІВЕРСИТЕТ

Факультет інформаційних технологій, обліку та фінансів
Кафедра комп'ютерних технологій
і моделювання систем

Кваліфікаційна робота
на правах рукопису

Данильчук Назарій Вадимович
(прізвище, ім'я, по батькові здобувача освіти)

УДК 004.4:004.8

КВАЛІФІКАЦІЙНА РОБОТА

Система крос-перевірки фактів
(тема роботи)

122 «Комп'ютерні науки»
(шифр і назва спеціальності)

Подається на здобуття освітнього ступеня бакалавр

кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

(підпис, ініціали та прізвище здобувача вищої освіти)

Керівник роботи
Николук Ольга Миколаївна,
д.е.н., проф., керівник ННЦ
цифрових технологій та моделювання систем
Поліського національного університету

Житомир – 2025

Висновок кафедри _____

за результатами попереднього захисту: _____

Протокол засідання кафедри _____

№ _____ від «_____» _____ 20____ р.

Завідувач кафедри _____

(науковий ступінь, вчене звання)

(підпис)

(прізвище, ім'я, по батькові)

«_____» _____ 20____ р.

Результати захисту кваліфікаційної роботи

Здобувач вищої освіти _____ захистив (ла)

(прізвище, ім'я, по батькові)

кваліфікаційну роботу з оцінкою:

сума балів за 100-бальною шкалою _____

за шкалою ECTS _____

за національною шкалою _____

Секретар ЕК

(науковий ступінь, вчене звання)

(підпис)

(прізвище, ім'я, по батькові)

АНОТАЦІЯ

Данильчук Н. В. Інформаційна система крос-перевірки фактів – Кваліфікаційна робота на правах рукопису.

Кваліфікаційна робота присвячена розробці інформаційної системи для крос-перевірки фактів на основі даних з відкритих джерел. Метою системи є надання інформації, необхідної для аналізу достовірності текстової інформації через аналіз, пошук і аналіз тверджень за допомогою SERP API та мовних моделей.

Розроблена інформаційна система передбачає два основних сценарії її використання. У першому сценарії користувач здійснює пошук твердження одночасно в п'яти пошукових системах (Google, Bing, Yahoo, DuckDuckGo, Yandex). Це дозволяє отримати широкий спектр результатів з різноманітних джерел та порівняти знайдену інформацію для попередньої перевірки. У другому сценарії система аналізує наданий користувачем текст за допомогою мовної моделі OpenAI GPT-4, яка виконує семантичний аналіз, виокремлює потенційно сумнівні твердження.

Наукова новизна полягає в поєднанні механізмів веб-скрапінгу, машинного навчання та мовних моделей для формалізованої перевірки текстових тверджень. Така система може бути ефективно застосована в медіааналітиці, освітньому процесі, наукових дослідженнях та журналістських розслідуваннях.

Ключові слова: фактчекінг, крос-перевірка, GPT-4, інформаційна система, пошукова система, штучний інтелект, веб-скрапінг, посилання.

SUMMARY

Danilchuk N. V. Information System for Cross-Fact-Checking – Qualification Manuscript.

The qualification work is dedicated to the development of an information system for cross-fact-checking based on data from open sources. The goal of the system is to provide the necessary information for analyzing the reliability of textual data through the examination, search, and analysis of claims using SERP API and language models.

The developed information system involves two main usage scenarios. In the first scenario, the user searches for a claim simultaneously across five search engines (Google, Bing, Yahoo, DuckDuckGo, Yandex). This allows obtaining a wide range of results from diverse sources and comparing the retrieved information for preliminary verification. In the second scenario, the system analyzes the user-provided text using the OpenAI GPT-4 language model, which performs semantic analysis and extracts potentially dubious claims.

The scientific novelty lies in the combination of web scraping mechanisms, machine learning, and language models for the formalized verification of textual claims. Such a system can be effectively applied in media analytics, educational processes, scientific research, and journalistic investigations.

Keywords: fact-checking, cross-verification, GPT-4, information system, search engine, artificial intelligence, web scraping, references.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	7
ВСТУП.....	8
Розділ 1. АНАЛІЗ ОСОБЛИВОСТЕЙ ПРЕДМЕТНОЇ ОБЛАСТІ.....	11
1.1 Роль, інформаційні потреби та визначення інформаційної системи крос-перевірки фактів	11
1.2 Моделювання інформаційної системи крос-перевірки фактів	12
Висновки до розділу 1.....	15
Розділ 2. ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ КРОС-ПЕРЕВІРКИ ФАКТІВ	16
2.1 Структура інформаційної системи крос-перевірки фактів	16
2.2 Моделювання інформаційної систем крос-перевірки фактів	18
Висновки до розділу 2.....	22
Розділ 3. РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ КРОС-ПЕРЕВІРКИ ФАКТІВ	23
3.1 Програмна реалізація інформаційної системи крос-перевірки фактів	23
3.2 Інтерфейс інформаційної системи крос-перевірки фактів.....	24
3.3 Інструкція користувачу інформаційної системи крос-перевірки фактів	26
Висновки до розділу 3.....	27
ВИСНОВКИ.....	29
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	30
ДОДАТКИ	36

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

API – Application Programming Interface

SERP – Search Engine Results Page

GPT – Generative Pre-trained Transformer

ІС – інформаційна система

ІІІ – штучний інтелект

БД – база даних

IDEF0 – Icam DEFinition for Function Modeling

IDEF3 – Integrated DEFinition for Process Description Capture Method

ER-діаграма – Entity-Relationship diagram

REST – Representational State Transfer

HTML – HyperText Markup Language

PDF – Portable Document Format

JSON – JavaScript Object Notation

TTLCache – Time To Live Cache

ВСТУП

У сучасному світі, що характеризується стрімким розвитком цифрових технологій та великим обсягом доступної інформації, питання її достовірності та надійності набуває першочергового значення. Поява та швидке поширення дезінформації, фейкових новин, а також потенційно маніпулятивних тверджень створюють нові виклики для суспільства, впливаючи на формування громадської думки, освітній процес, проведення наукових досліджень, журналістські розслідування та прийняття рішень у різних сферах. У відповідь на ці виклики спостерігається зростання критичної потреби у розробці ефективних інструментів та методик для фактчекінгу – процесу перевірки фактів, який дозволяв би оперативно та об'єктивно оцінювати правдивість текстової інформації.

Традиційні підходи до верифікації фактів часто є трудомісткими, вимагають значних людських ресурсів і глибокої експертизи, що обмежує їх масштабованість та швидкість реагування на потоки нової інформації. З огляду на це, актуальною стає розробка спеціалізованих інформаційних систем, здатних автоматизувати та оптимізувати процес крос-перевірки фактів. Такі системи покликані використовувати можливості сучасних інформаційних технологій для комплексного збору, аналізу та зіставлення інформації з різноманітних джерел, що дозволяє суттєво підвищити ефективність та швидкість верифікації даних.

Ключовим елементом для реалізації таких систем є пошукові машини, які відіграють центральну роль у забезпеченні доступу до веб-даних. Ці програмно-апаратні комплекси відповідають за виявлення, індексацію, збереження та подальше надання користувачам релевантної інформації, перетворюючи неструктуровані дані у зручний та доступний формат.

Проте, для забезпечення повноцінної та всебічної крос-перевірки фактів недостатньо покладатися виключно на можливості однієї пошукової системи або традиційні методи. Необхідною є інтеграція різноманітних підходів та сучасних технологій. У цьому контексті наукова новизна полягає саме у поєднанні механізмів веб-скрапінгу (для автоматизованого збору даних з веб-джерел) та передових

мовних моделей, таких як OpenAI GPT-4, для семантичного аналізу тексту та виокремлення потенційно сумнівних тверджень.

Предметом дослідження кваліфікаційної роботи є методи, засоби та інструменти, що використовуються у процесі проєктування, моделювання та реалізації інформаційної системи крос-перевірки фактів.

Об'єктом дослідження є процес застосування цих методів, засобів та інструментів для створення ефективного програмного рішення інформаційної системи крос-перевірки фактів.

Метою дослідження є розробка інформаційної системи для крос-перевірки фактів на основі даних з відкритих джерел, яка дасть можливість автоматизувати та оптимізувати процес перевірки достовірності текстової інформації через пошук і аналіз тверджень за допомогою SERP API та мовних моделей. Для досягнення поставленої мети визначено такі завдання дослідження:

- аналіз особливостей предметної області, ролі та інформаційних потреб інформаційної системи крос-перевірки фактів;
- моделювання та проєктування структури інформаційної системи крос-перевірки фактів;
- програмна реалізація інформаційної системи крос-перевірки фактів, включаючи розробку її ключових модулів та інтерфейсу;

Розробка такої інформаційної системи передбачає реалізацію двох основних сценаріїв використання. У першому сценарії користувач має можливість здійснювати пошук заданого твердження одночасно в п'яти пошукових системах (Google, Bing, Yahoo, DuckDuckGo, Yandex), що дозволяє отримати широкий спектр результатів з різноманітних джерел та зіставити знайдену інформацію для попередньої верифікації. У другому сценарії система здійснює аналіз наданого користувачем тексту за допомогою мовної моделі OpenAI GPT-4, яка виконує семантичний аналіз тексту, ідентифікуючи та виокремлюючи потенційно сумнівні твердження для подальшої перевірки.

Результати дослідження апробовано на двох конференціях, а саме [1-2]:

1. Міжфакультетська науково-практична інтернет-конференція здобувачів вищої освіти і молодих вчених «Безпека, технології, інновації: нові горизонти» (Поліський національний університет, Житомир, 12 листопада 2024 року).

2. Всеукраїнська науково-практична конференція «Інформаційні технології та моделювання систем» (Житомир, 15 травня 2025 року).

Розділ 1. АНАЛІЗ ОСОБЛИВОСТЕЙ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Роль, інформаційні потреби та визначення інформаційної системи крос-перевірки фактів

У сучасному інформаційному суспільстві, яке характеризується високою динамікою генерації та розповсюдження даних, достовірність інформації набуває критично важливого значення [3-6]. Зростання обсягів цифрового контенту, доступність засобів його масового поширення та відсутність належного контролю з боку джерел формують сприятливе середовище для виникнення та поширення дезінформації і маніпулятивних тверджень [7-10]. Ці явища становлять загрозу не лише для об'єктивного інформування суспільства, а й для процесів прийняття рішень у науці, освіті, державному управлінні, медіа та інших сферах. У відповідь на ці виклики зростає потреба в ефективних інструментах пошуку та аналізу інформації [11-13].

Механізм функціонування пошукової системи охоплює кілька послідовних етапів: на початковій стадії веб-краулери, або «павуки», – автоматизовані програми – обходять доступні їм в мережі гіперпосилання, завантажують вміст веб-сторінок і передають його до індексувального модуля. Суть процесу індексації полягає у аналізі контенту: система виділяє ключові слова, визначає їхню семантику, частотність та контекст використання, формуючи на основі отриманої інформації спеціальний індекс. В подальшому, при виконанні пошукового запиту, система звертається до цього побудованого індексу та ранжує знайдені документи відповідно до низки алгоритмів, включаючи алгоритми на основі зважених обчислень, семантичного аналізу та машинного навчання.

Однією із груп цільової аудиторії, зацікавлених у системному підході та автоматизації процесу крос-перевірки фактів є представники медіа-сфери – журналісти, редактори, фактчекери, контент-аналітики та медіаменеджери, для яких своєчасна, точна й незалежно перевірена інформація становить професійну цінність. Також система може бути корисною медіаосвітнім ініціативам,

платформам незалежної журналістики, громадським організаціям, які здійснюють моніторинг дезінформації або публічних заяв [14-23].

Важливим є забезпечення можливості нейтралізувати ризики, пов'язані з обмеженнями традиційного пошуку. Адже результати видачі в різних пошукових системах можуть істотно відрізнятися через відмінності в алгоритмах індексації, політиці регіонального фільтрування, частоті оновлення даних і цензурних обмеженнях. Наприклад, одна і та ж фраза, введена у Google, Bing і Yandex, може повертати як абсолютно різні першоджерела, так і різну інтерпретацію події. Це пов'язано з тим, що кожна система має власну інфраструктуру, правила ранжування, фільтри для чутливих тем і стратегії персоналізації. До того ж, жодна з них не має повного охоплення всього Інтернету [24-31].

Таким чином, лише багатоджерельний підхід, який забезпечує крос-перевірку тверджень у різних екосистемах, дозволяє отримати повноцінне уявлення про ступінь достовірності інформації [32-36]. У цьому полягає перевага запропонованого підходу до крос-перевірки фактів – вона не замінює користувача у процесі мислення, але надає йому інструмент орієнтації в інформаційному середовищі та виводить на перший план факти, які варті перевірки. Саме такий підхід пропонується покласти в основу інформаційної системи факт-чекінгу даних для задоволення потреб визначеної вище цільової аудиторії.

1.2 Моделювання інформаційної системи крос-перевірки фактів

Для вирішення проблем та інформаційних потреб у рамках організації крос-перевірки фактів для задоволення потреб представників мас-медіа пропонується програмний інструмент, що автоматизує процес верифікації тверджень шляхом пошуку, аналізу та порівняння результатів з різних незалежних джерел. Її принцип роботи базується на паралельному зверненні до кількох пошукових систем (Google, Bing, Yahoo, DuckDuckGo, Yandex). У поєднанні з семантичним аналізом тексту за допомогою мовної моделі GPT-4 це дозволяє виявити потенційно сумнівні твердження.

Вхідними параметрами для системи є запит користувача, який може містити тезу для пошуку, посилання на статтю або її текст для аналізу (рис. 1.1). Механізми реалізації охоплюють веб-інтерфейс для взаємодії з користувачем та модулі, що реалізують інтеграцію пошукових систем та мовної моделі. Для коректної реалізації модулів пошуку та аналізу необхідним є ознайомлення із документаційними матеріалами залученими в проект зовнішніх інтерфейсів прикладного програмування: SOAX SERP та Open AI. Останній, в свою чергу, потребує застосування знань з промпт інжинірингу.

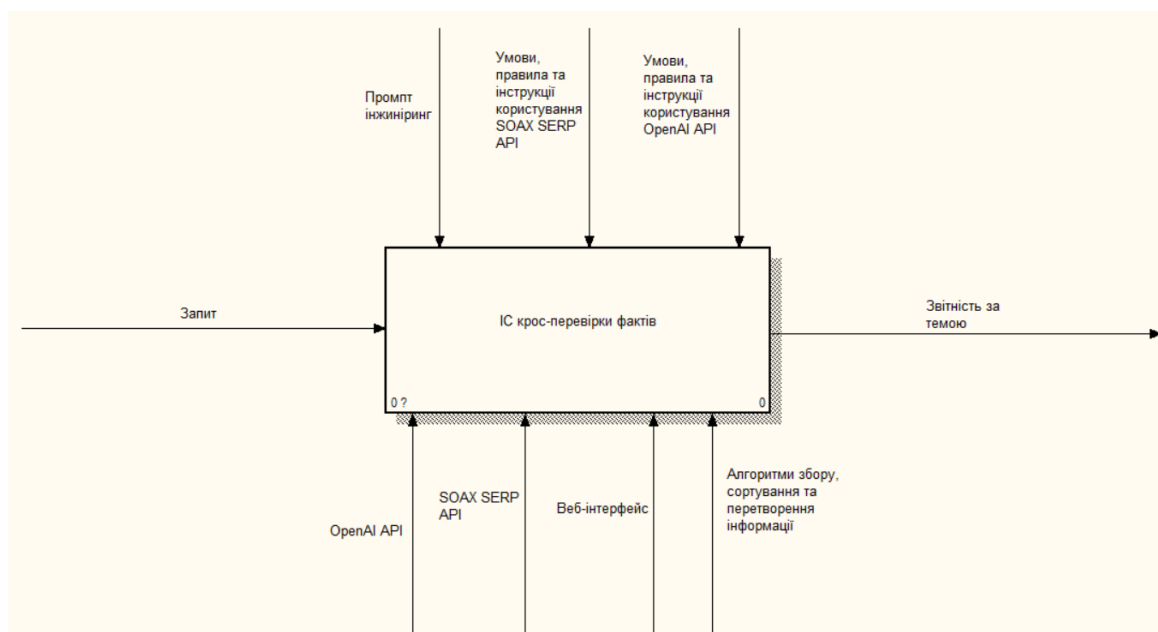


Рис. 1.1 – Контекстна IDEF0-діаграма ІС крос-перевірки фактів

Вихідними результатами в залежності від сценарію використання є перелік посилань із зазначених пошукових систем, або результат аналізу текстових даних.

IDEF0-діаграма роботи системи крос-перевірки фактів представлена на рисунку 1.2, що відображає послідовність функціональних блоків та їх взаємозв'язки. Первинна обробка запиту передбачає взаємодію користувача з ІС через веб-інтерфейс, де формується запит для подальшої обробки. В залежності від типу запиту, застосовуються блоки виконання пошуку в зазначених пошукових системах або аналіз текстових даних. Формування відповіді користувачу є заключним етапом роботи системи і передбачає відображення звітності за допомогою веб-інтерфейсу.

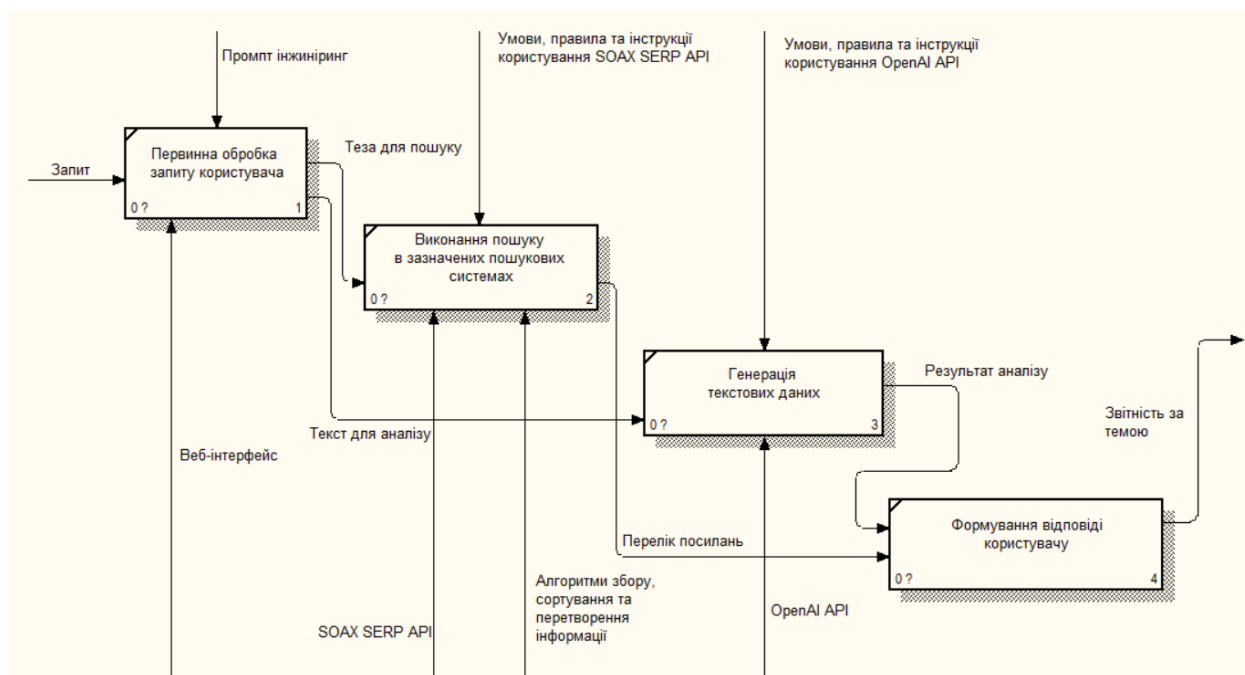


Рис. 1.2 – IDEF0-діаграма роботи ІС крос-перевірки фактів

Порядок, у якому проходить одна ітерація пошуку або аналізу даних, можна описати наступними чином: після отримання запиту система визначає тип запиту (рис. 1.3). Якщо запит містить посилання на статтю, задіюється веб-скрапер. Наступним етапом є аналіз отриманої за допомогою веб-скрапера статті через ВММ. Запит до мовної моделі містить текст статті та промпт-інструкцію (Додаток А).

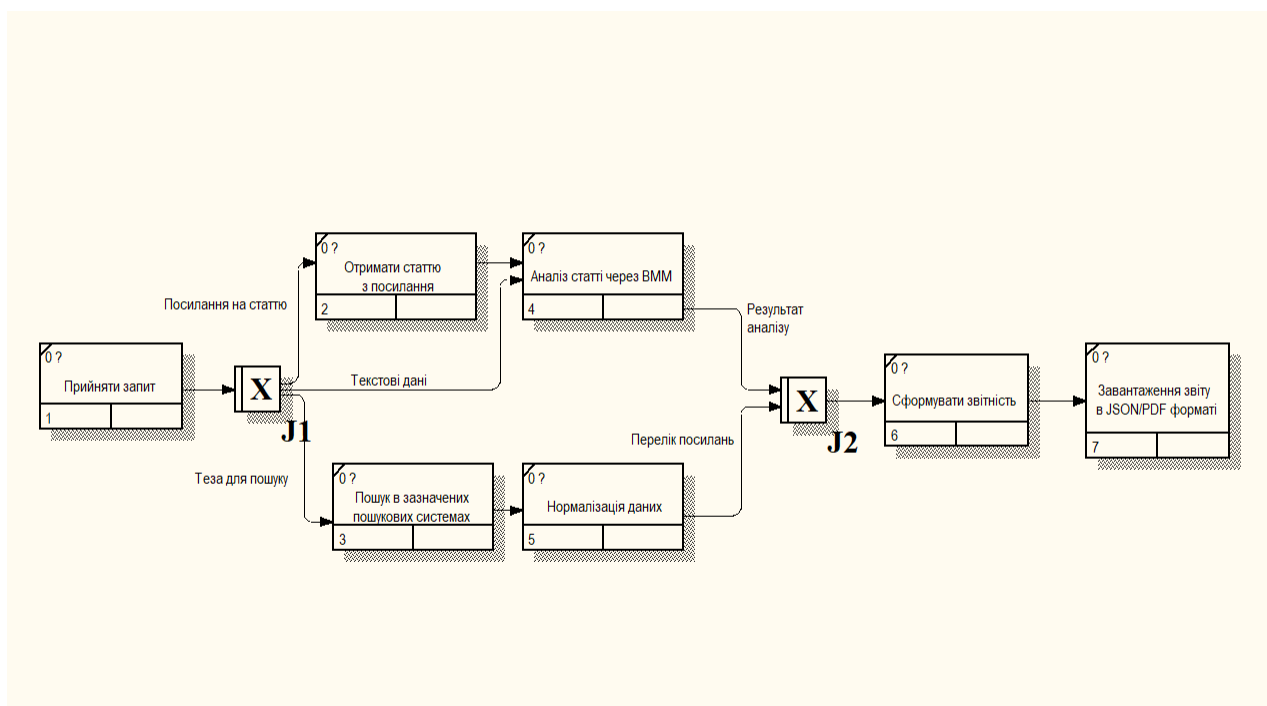


Рис 1.3 – IDEF3-діаграма ІС крос-перевірки фактів

У випадку, якщо запит є тезою для пошуку, здійснюється паралельний пошук у п'яти пошукових системах (Google, Bing, Yahoo, DuckDuckGo, Yandex). Результати пошуку нормалізуються і формують основу для звіту. Окрім відображення звітності за допомогою веб-інтерфейсу, передбачено завантаження звітності у PDF і JSON форматах.

Висновки до розділу 1

У розділі 1 проаналізовано особливості предметної області крос-перевірки фактів, зокрема роль та інформаційні потреби системи у контексті сучасних викликів дезінформації. Розроблено ключові характеристики системи, включаючи алгоритм паралельного пошуку в п'яти пошукових системах (Google, Bing, Yahoo, DuckDuckGo, Yandex) та використання мовної моделі GPT-4 для семантичного аналізу текстів. Побудовано IDEF0-діаграму, яка відображає функціональну структуру системи, зокрема блоки обробки запитів, пошуку, аналізу даних та формування звітів. Також розроблено IDEF3-діаграму, що деталізує послідовність виконання процесів, таких як обробка запитів, веб-скрапінг, аналіз тексту та генерація результатів. Визначено цільову аудиторію системи, до якої входять журналісти, науковці, освітні установи та інші користувачі, що потребують оперативної перевірки інформації.

Розділ 2. ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ КРОС-ПЕРЕВІРКИ ФАКТІВ

2.1 Структура інформаційної системи крос-перевірки фактів

Представлена інформаційна система крос-перевірки фактів функціонує на основі багаторівневої архітектури, яка включає в себе веб-інтерфейс, серверний застосунок та рівень API, побудованої на принципах REST [37]. Веб-інтерфейс є клієнтською частиною системи та точкою входу для користувача. Він складається з розділів «Головна», «Пошук» та «Аналіз» (рис. 2.1). Використання функцій ініціює запити до серверного застосунку для виконання відповідних операцій з обробки даних.

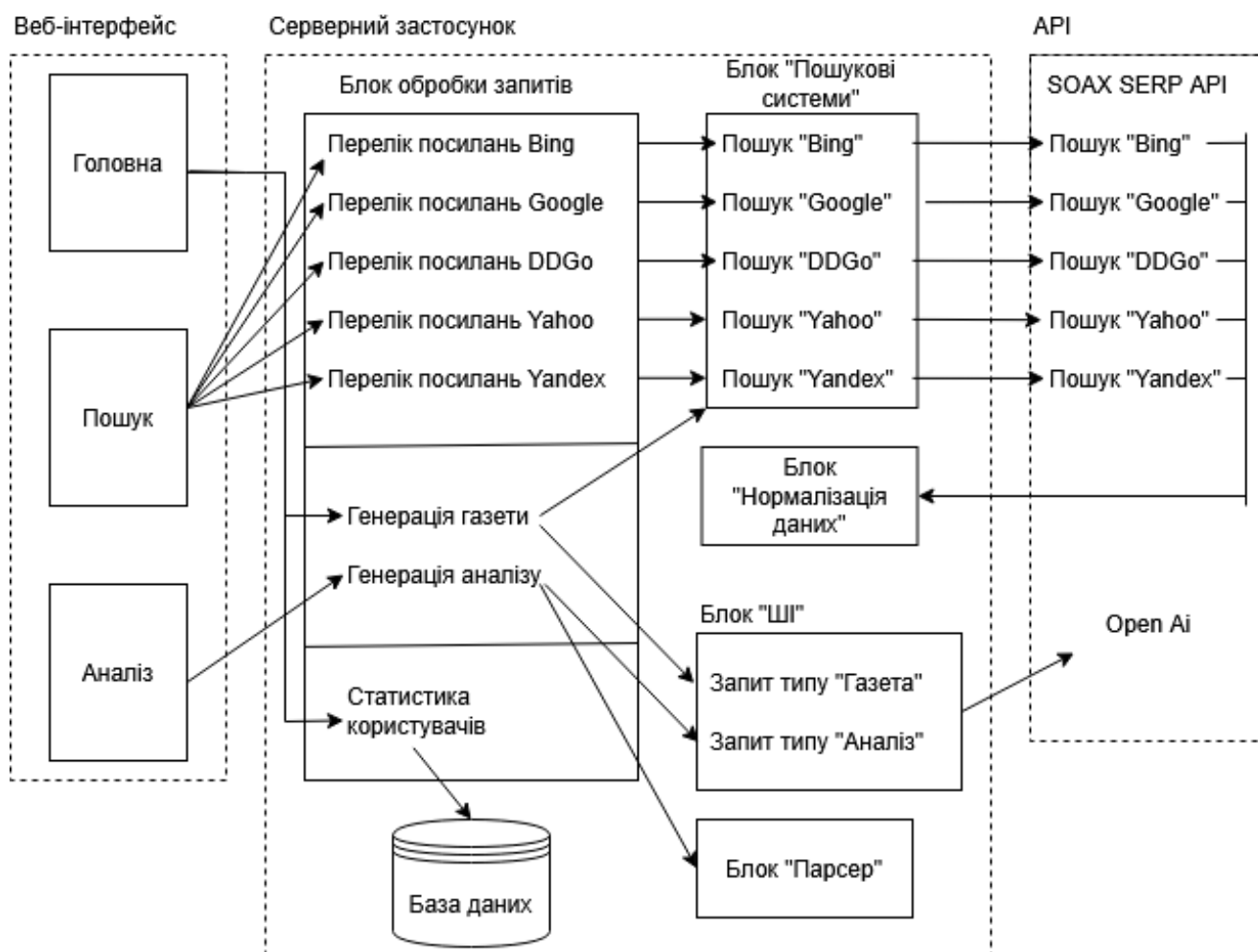


Рис 2.1 – Структура ІС

Серверний застосунок є ядром системи, відповідальним за обробку запитів. Обробка запиту починається з визначення його типу. Блок прийому запитів

визначає, які ресурси потрібні для виконання запиту, та залучає їх. У випадку із запитом на перелік посилань з пошукової системи реалізується блок «Пошукові системи», який виконує функцію підготовки та виконання запиту до SERP API. Дані, що містить відповідь від SERP API з різних пошукових систем має різну структуру і тому потребує нормалізації. За приведення даних до однієї структури відповідає блок «Нормалізація даних».

Запити щодо аналізу текстових даних залучають блок «ШІ», а у випадку з поданням як аргумента запиту посилання на статтю – блок «Парсер». Блок «ШІ», в свою чергу, виконує роль підготовки, виконання та пост обробки запиту до мовної моделі.

Головна сторінка веб-інтерфейсу слугує точкою входу до інформаційної системи та виконує функцію первинної навігації користувача. Вона забезпечує ознайомлення з ключовими можливостями системи, структурує доступ до основних функцій. Головна сторінка містить блок зі статистикою користувачів та газетою – блоком, що містить останні новини в Україні та світі. Блок статистики користувачів отримує дані з БД, яка містить історію користувацьких запитів. Газета активує модуль «Пошукові системи» для пошуку посилань на тему актуальних новин, після чого ці посилання проходять пост обробку модулем «ШІ», з використанням промпта-інструкції, що поданий в додатках (Додаток Б).

Для зберігання даних користувачів та історії їх запитів розроблено базу даних (рис. 2.2). Таблиця «users» містить дані авторизації користувачів. Таблиці «ai_requests» та «ss_requests» містять статистику щодо запитів до модулів, які виконують аналіз текстових даних та пошук посилань, відповідно. Таблиця «region» містить перелік регіонів України, які можна обрати при реєстрації користувача.

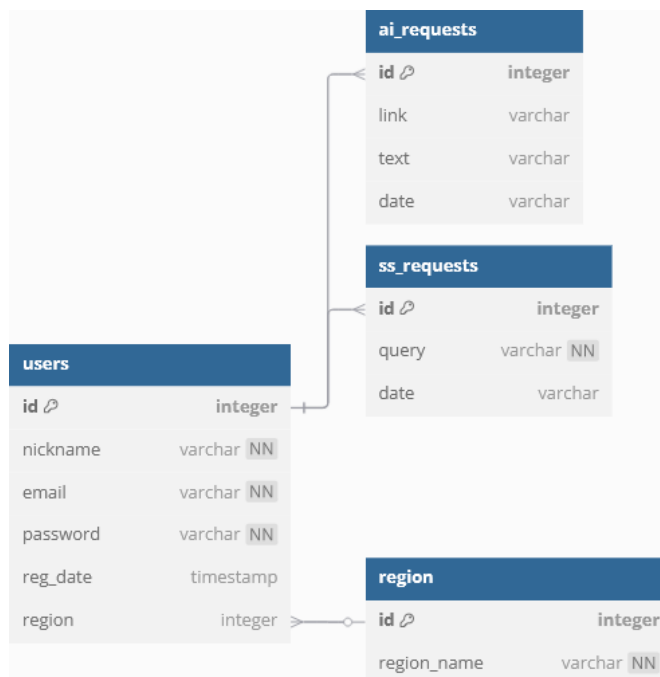


Рис 2.2 – ER-діаграма для БД

2.2 Моделювання інформаційної систем крос-перевірки фактів

Основними сутностями, що беруть участь у функціонуванні ІС, є користувач та серверний застосунок, який під час обробки запитів здійснює аналіз текстових даних за допомогою мовної моделі та їх пошук в пошукових системах, а також зберігає дані щодо запитів у БД. Загальну схему функціонування ІС зображено на діаграмі прецедентів (рис. 2.3).



Рис 2.3 – Діаграма прецедентів

Вхід до серверного застосунку можна описати як постійний цикл перевірки наявності запитів. Після надходження запиту визначається його тип, зокрема запит на отримання контенту для головної сторінки (газета), авторизації або щодо основного функціоналу (рис 2.4). Запит авторизації може стосуватися реєстрації користувача, або входу в уже існуючий кабінет. Незалежно від успіху або помилки авторизації, користувачу (клієнту) необхідно відобразити результат (формування відповіді). Запит щодо доступу до основного функціоналу реалізовує два основні сценарії використання ІС – запит до пошукових систем і повернення переліку посилань, або аналіз статті за допомогою мовної моделі. Запит на отримання контенту для газети, що розміщена на головній сторінці, послідовно задіює модулі пошуку та ВММ, для пошуку посилань на останні новини в Україні та світі та формування з них релевантного переліку новин.

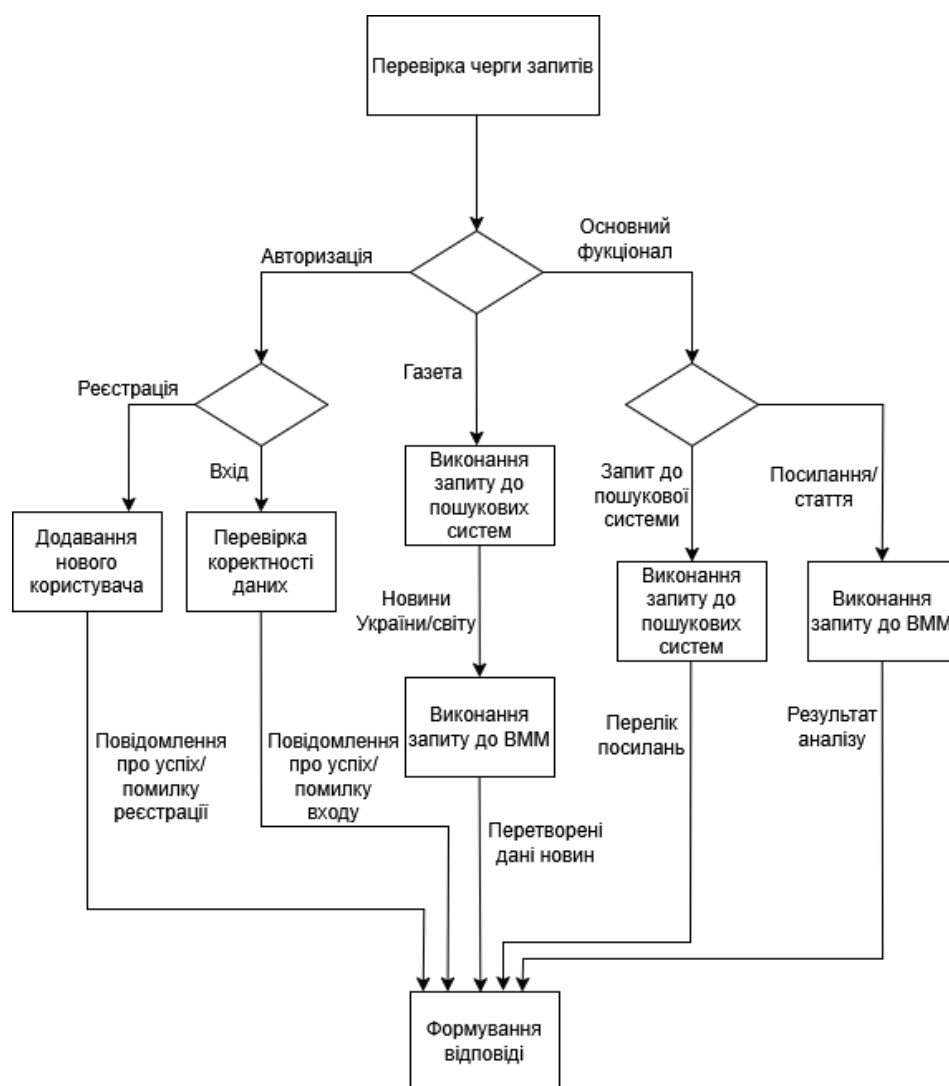


Рис 2.4 – Діаграма активності

Процес отримання переліку посилань ініціюється при надсиланні користувачем відповідного запиту через веб-інтерфейс (рис. 2.5). Сервер ініціює запит до SOAX SERP API для отримання результатів пошуку з підключених пошукових систем. Отримані дані, що включають перелік посилань та відповідні метадані, підлягають нормалізації для уніфікації різнорідних структур від різних пошукових систем. Після обробки сервер формує звіт, а дані про запит зберігаються в БД.

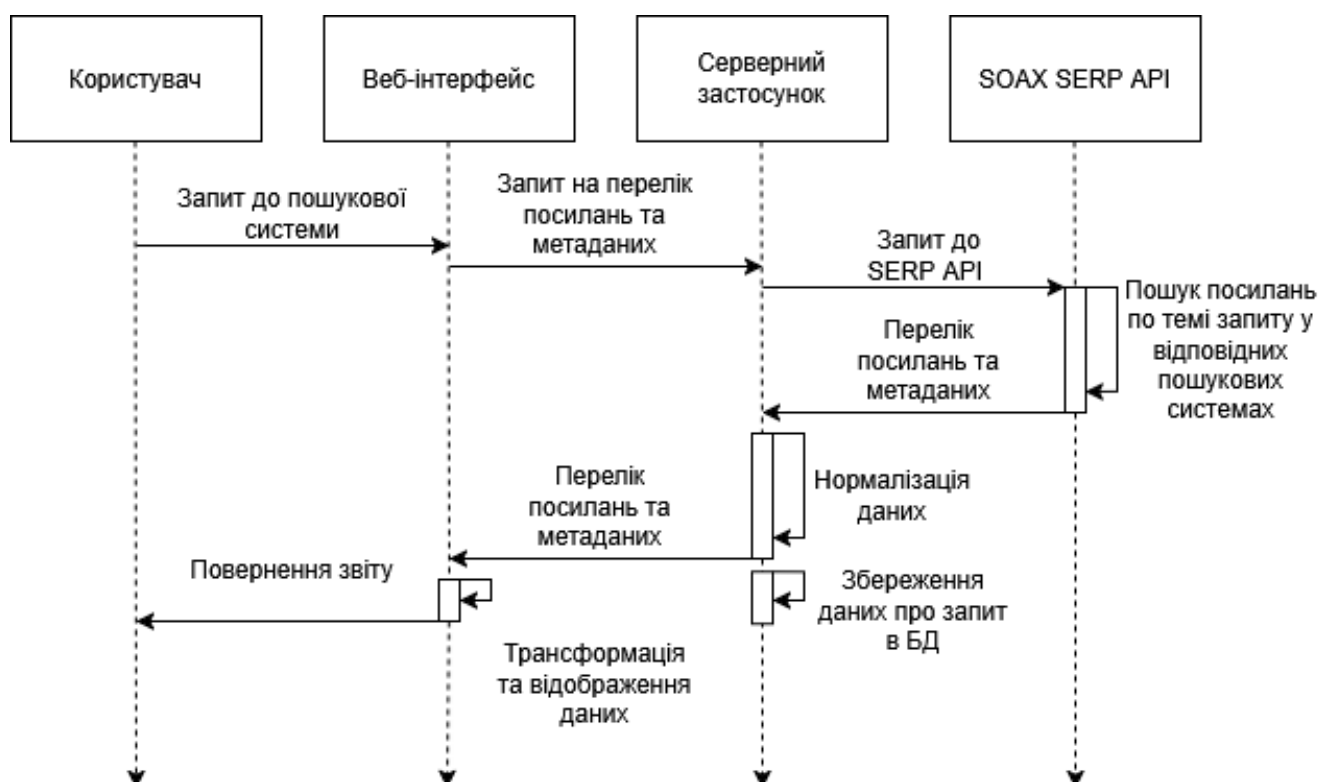


Рис 2.5 – Діаграма послідовності процесу отримання переліку посилань

Алгоритм аналізу текстових даних активується при надсиланні користувачем текстового контенту або посилання на статтю (рис. 2.6). Серверний застосунок отримує запит та ініціює процес парсингу статті, виконуючи витягнення текстового вмісту з наданого джерела. Отриманий текст аналізується через мовну модель GPT-4, яка передбачає здійснення семантичної обробки контенту з виявленням ключових тверджень та потенційно маніпулятивних елементів.

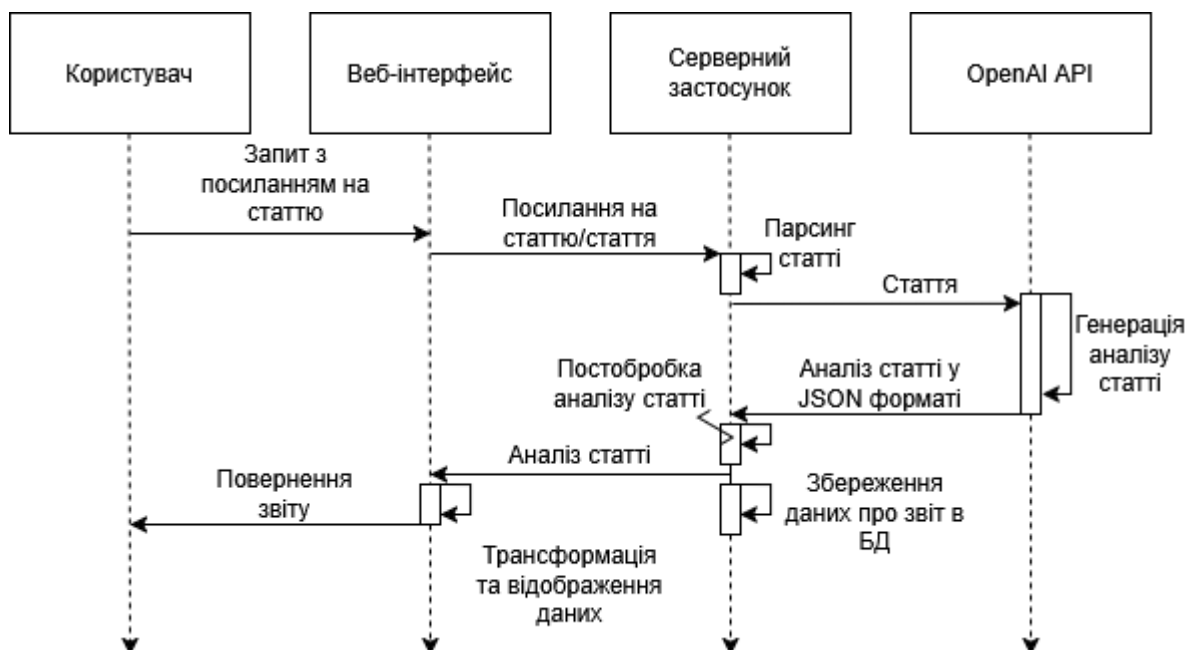


Рис 2.6 – Діаграма послідовності процесу аналізу текстових даних

Процес авторизації реалізований через стандартний механізм перевірки облікових даних (рис. 2.7).

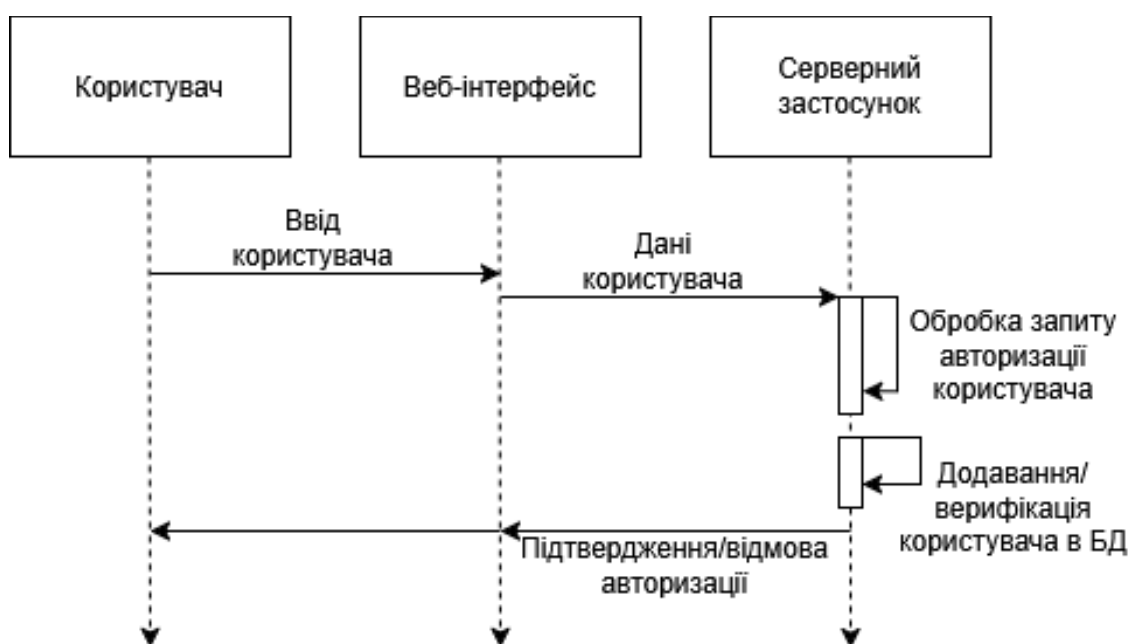


Рис 2.7 – Діаграма послідовності процесу авторизації

Серверний застосунок отримує введені дані та ініціює процес обробки запиту авторизації. Система виконує перевірку наданих даних, зіставляючи їх з інформацією, що зберігається в базі даних. В результаті перевірки система приймає рішення про підтвердження або відмову в авторизації.

Висновки до розділу 2

У розділі 2 було розроблено структуру ІС крос-перевірки фактів. Визначено багаторівневу архітектуру, що включає веб-інтерфейс, серверний застосунок та рівень API. Систему розділено на функціональні модулі: пошуковий (інтеграція з п'ятьма пошуковими системами через SERP API), аналітичний (робота з GPT-4 для семантичного аналізу), парсер (видобуток тексту зі статей), нормалізація даних (уніфікація різнорідних результатів) та база даних (зберігання історії запитів).

Для візуалізації логіки системи створено діаграми прецедентів, активності та послідовності, які описують процеси пошуку, аналізу текстів та авторизації. Розроблено ER-діаграму бази даних PostgreSQL з таблицями для користувачів, запитів до ШІ, пошукових систем. Описано інтерфейс системи, що включає три основні сторінки: головну (статистика та новини), пошук (форма запиту та результати) та аналіз (завантаження тексту/посилань та звіти). Для експорту даних реалізовано функціонал збереження результатів у PDF та JSON форматах.

Розділ 3. РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ КРОС-ПЕРЕВІРКИ ФАКТІВ

3.1 Програмна реалізація інформаційної системи крос-перевірки фактів

Серверна частина інформаційної системи реалізована на фреймворку Flask [38]. Модуль з веб-скрапером використовує для отримання html-розміток бібліотеку Requests (Додаток В) [39]. Кешування даних здійснюється за допомогою cachetools, а саме класу TTLCache (Time To Live), що зберігає дані у кеші заданий проміжок часу [40]. В якості БД використано PostgreSQL [41]. Схема, що описує архітектуру застосунку представлено на рисунку 3.1.

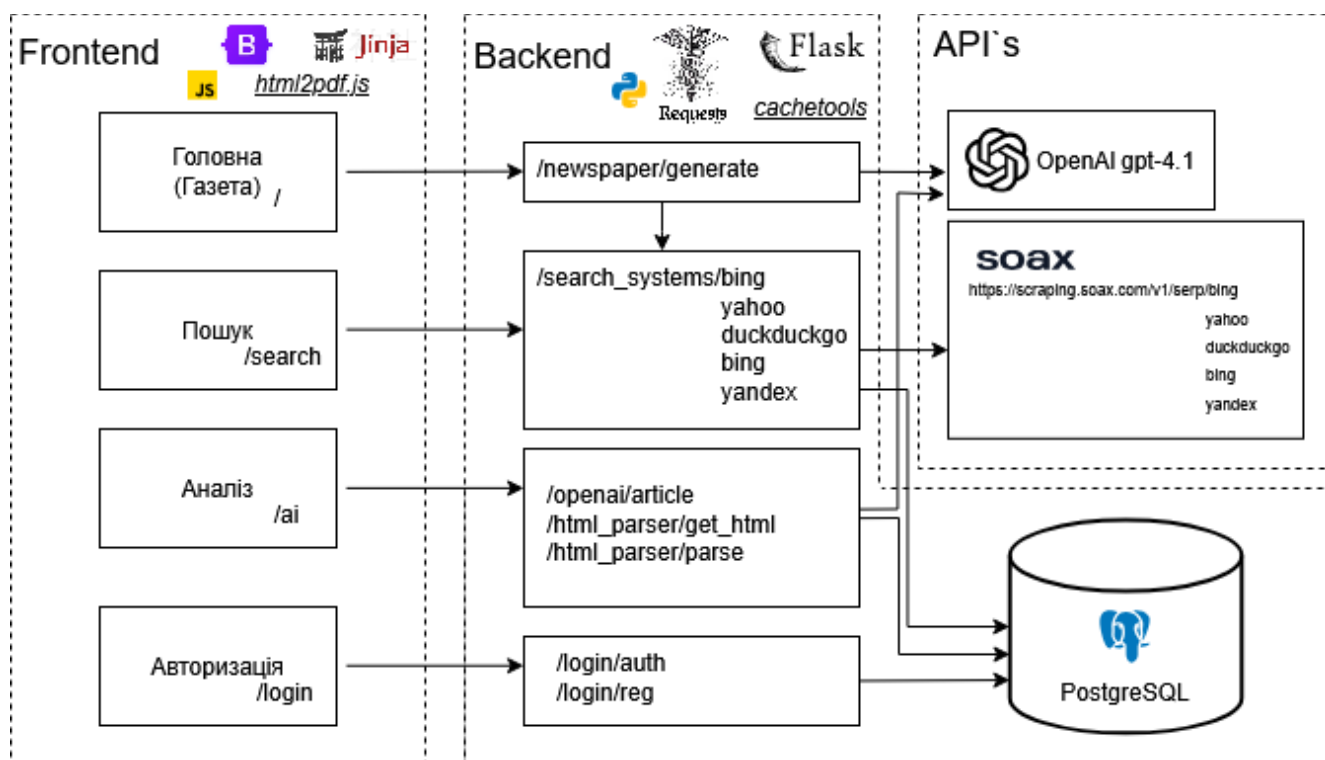


Рис. 3.1 – Архітектура застосунку з точки зору модулів та технологічного стеку

Основна логіка системи включає два ключові модулі:

1. Пошуковий модуль – інтегрує SOAX SERP API для паралельного запиту до п'яти пошукових систем (Google, Bing, Yahoo, DuckDuckGo, Yandex) [42]. Отримані результати нормалізуються для уніфікації різнорідних структур даних. Типовий приклад обробки запиту та нормалізації даних наведено в додатках Г та Д.
2. Аналітичний модуль – використовує OpenAI GPT-4 для семантичного аналізу текстів [43]. Виокремлює ключові твердження, дату публікації, авторів,

емоційний тон та цитати з вираженою емоційністю (код модуля представлено у Додатку Е).

Клієнтська частина побудована на основі шаблонізатора Jinja [44]. Для перетворення результатів у PDF використовується бібліотека html2pdf.js. Інтерфейс включає три основні сторінки:

1. Головна – відображає статистику запитів користувачів та актуальні новини (згенеровані через пошуковий модуль та GPT-4).
2. Пошук – форма для введення запиту та відображення результатів у вигляді списку посилань.
3. Аналіз – інструмент для завантаження тексту або посилання, з подальшим виведенням структурованого звіту.

Для спрощення процесу розгортання використано контейнеризатор Docker [45]. Розгорнутий застосунок являє собою два контейнери, а саме: Flask-застосунок та БД PostgreSQL (додаток Ж).

3.2 Інтерфейс інформаційної системи крос-перевірки фактів

Інтерфейс головної сторінки представлений багатосекційною структурою, призначеною для агрегації та категоризації інформації (рис 3.2). Навігаційна панель містить заголовок з логотипом «FACTCHECK», перехід до розділів «Пошук» та «Аналіз», а також піктограми, що, переспрямовують до інструкції користувачу та профілю користувача.

Ліва панель містить статистичні дані користувачів, розподілені за типами запитів «Пошук» та «Аналіз». Нижче розміщено секцію «Користувачі запитують», де відображено популярні запити. Центральні блоки інтерфейсу (газета), що позначений назвою «Україна», акумулює національні новини та події. Праворуч розташований блок «Мейнстрім», який висвітлює ключові події загального характеру.



Рис. 3.2 – Інтерфейс головної сторінки

Сторінка пошуку містить блок з полем для запиту. Під пошуковою панеллю розташовано ряд піктограм, що представляють фільтри за пошуковою системою та кнопки для завантаження звітів у форматах PDF і JSON (рис 3.3).

Основну частину сторінки займає область результатів пошуку, представлена у вигляді вертикального списку. Кожен елемент списку містить наступні атрибути: домен веб-сайту-джерела, заголовок інформаційного матеріалу та дату публікації.

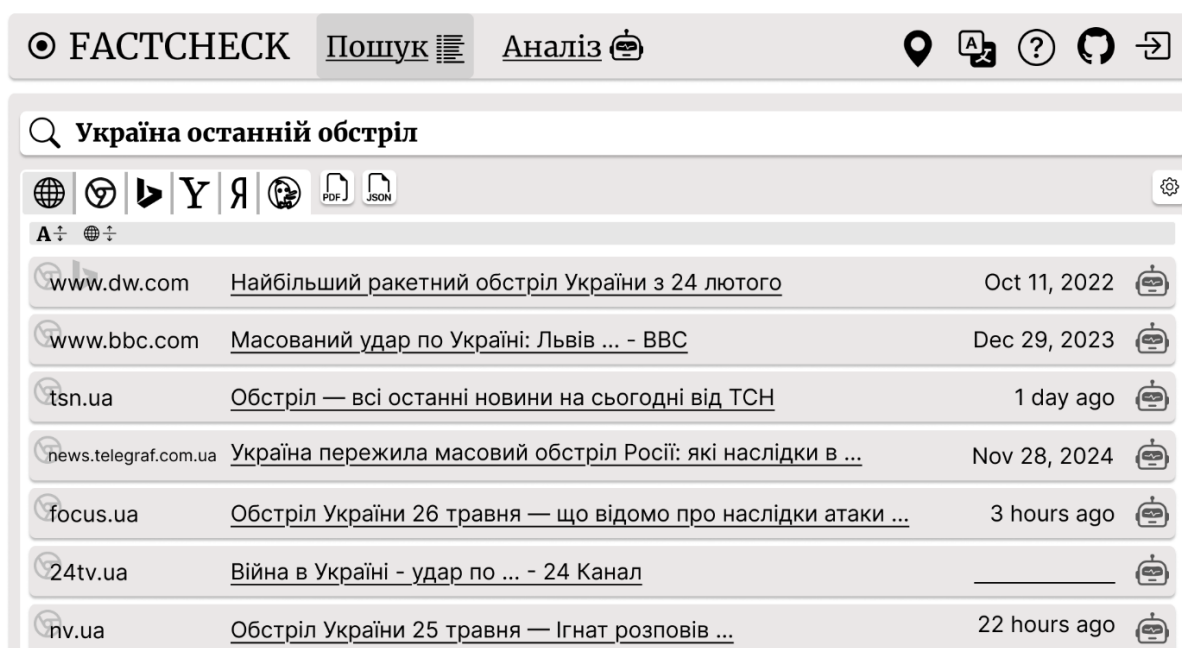


Рис. 3.3 – Інтерфейс сторінки «Пошук»

Ліва панель інтерфейсу слугує для ідентифікації та попереднього перегляду вихідного матеріалу, що аналізується (рис 3.4). Вона містить поле вводу для посилання на джерело та текстовий блок, який відображає текст, що вдалось видобути з посилання, шляхом веб-скрапінгу та очистки текстових даних від HTML-елементів. Центральний блок інтерфейсу призначений для представлення результатів аналізу і відображає теги, що були повернуті у відповіді сервера після обробки запиту. Приклад звіту в PDF форматі наведено в Додатку ІІ.



Рис. 3.4 – Інтерфейс сторінки «Аналіз»

3.3 Інструкція користувачу інформаційної системи крос-перевірки фактів

Інформаційна система крос-перевірки фактів реалізує два основні сценарії використання: паралельний пошук у п'яти пошукових системах та аналіз тексту за допомогою мовної моделі. Система передбачає можливість створення облікового запису. Неавторизований користувач може користуватись функціями пошуку і аналізу, але історія запитів вестись не буде.

Доступ до системи здійснюється через веб-інтерфейс, який складається з трьох основних сторінок:

- **Головна** – містить статистику запитів, блок новин («газета»), актуальні події («мейнстрім») та переходи до основних функцій;
- **Пошук** – дозволяє ввести твердження для перевірки, здійснити паралельний запит до Google, Bing, Yahoo, DuckDuckGo, Yandex, переглянути результати та завантажити звіт у форматах PDF або JSON;
- **Аналіз** – передбачає вставлення тексту або посилання на статтю для обробки GPT-4, результатом є структурований звіт, що містить тему, авторів, дату публікації, ключові тези, емоційний тон та цитати.

Приклад використання: користувач виявляє у соціальних мережах поширене твердження: *«Уряд України повністю скасував фінансування вищої освіти»*.

Для перевірки достовірності цього повідомлення він переходить до розділу «Пошук» веб-інтерфейсу системи, вводить ключову фразу та ініціює паралельний запит до п'яти пошукових систем (Google, Bing, Yahoo, DuckDuckGo, Yandex). У відповідь система формує список релевантних джерел, з яких можна зробити висновок, що інформація стосується часткових змін у бюджетному фінансуванні, але не підтверджує повне скасування підтримки вищої освіти.

Розділ «Аналіз» доцільно використовувати у випадках, коли необхідно швидко сформулювати загальне уявлення про зміст великого текстового матеріалу без повного прочитання. Такий підхід дозволяє зекономити час, автоматизувати попередню оцінку тексту та оперативно отримати структурований звіт зі стислим викладом теми, ключовими тезами, емоційним тоном та іншими характеристиками. Це особливо корисно при роботі з великими масивами статей або новинних повідомлень.

Висновки до розділу 3

У розділі 3 було описано реалізацію інформаційної системи крос-перевірки фактів. Система інтегрує два ключові функціональні модулі: пошуковий, який використовує SOAX SERP API для паралельного запиту до п'яти пошукових систем (Google, Bing, Yahoo, DuckDuckGo, Yandex) та аналітичний модуль, що застосовує

потужності OpenAI GPT-4 для семантичного аналізу текстів, виокремлення ключових тверджень, ідентифікації авторів, дат публікації та аналізу емоційного тону контенту.

Для оптимізації продуктивності системи реалізовано механізм кешування даних з використанням бібліотеки `cachetools`, зокрема класу `TTLCache`. Користувацький інтерфейс системи організований у вигляді трьох основних сторінок: головної (відображає статистику запитів та актуальні новинні матеріали), сторінки пошуку (забезпечує введення запитів та візуалізацію результатів у формі структурованого списку посилань) та сторінки аналізу (надає інструментарій для завантаження текстів чи посилань з подальшим генеруванням детальних звітів).

Додатково реалізовано функціонал експорту отриманих результатів у зручних для користувача форматах PDF та JSON. Процес розгортання системи значно спрощено завдяки використанню технології контейнеризації Docker.

Практичні аспекти роботи системи, зокрема результати аналізу текстів за допомогою GPT-4 (виділення ключових тез, атрибутів публікацій та емоційного забарвлення) наведено у Додатку К, а приклади результатів паралельного пошуку в різних пошукових системах представлено у Додатку Л.

ВИСНОВКИ

У сучасних умовах стрімкого поширення інформації особливої важливості набуває швидка та об'єктивна перевірка фактів. Зростання кількості маніпулятивних повідомлень у медіа та соціальних мережах зумовлює потребу у розробці інструментів, здатних автоматизувати процес верифікації даних з відкритих джерел.

У рамках дослідження було створено інформаційну систему, яка поєднує паралельний пошук у п'яти пошукових системах із семантичним аналізом текстів за допомогою GPT-4. Система реалізована на базі Flask-фреймворку, використовує PostgreSQL для зберігання даних та Docker для розгортання. Ключові модулі – пошуковий (інтеграція з SERP API) та аналітичний (робота з GPT-4) – забезпечують автоматизовану перевірку фактів, виокремлення ключових тез та оцінку емоційного тону. Система призначена для журналістів, науковців, освітніх установ та інших користувачів, які потребують оперативної перевірки інформації. Її ключові переваги – автоматизація трудомістких процесів, зменшення впливу суб'єктивних оцінок та можливість аналізу великих масивів даних.

Подальші дослідження можуть бути спрямовані на розширення функціоналу системи, зокрема на інтеграцію додаткових джерел даних (інших пошукових систем) та впровадження алгоритмів аналізу мультимедійного контенту (зображень, відео). Також варто розглянути можливість багатомовної підтримки та вдосконалення промпт-інжинірингу для підвищення точності семантичного аналізу. Це дозволить адаптувати систему до нових викликів у сфері цифрової безпеки.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Данильчук Н. В. Інструменти пошуку й аналізу текстових даних / Н. В. Данильчук // Безпека, технології, інновації: нові горизонти : зб. праць учасників Міжфакультетської наук.-практ. інтернет-конф. здобувачів вищої освіти і молодих вчених (Житомир, 12 лист. 2024 р.). – Житомир : Поліський національний університет, 2024. – С. 73-75.
2. Данильчук Н. В. Інструменти пошуку й аналізу текстових даних / Н. В. Данильчук // Інформаційні технології та моделювання систем : зб. праць учасників Всеукр. наук.-практ. конф. здобувачів вищої освіти і молодих вчених (Житомир, 15 трав. 2025 р.). – Житомир, 2025. – С. 15-17.
3. Vosoughi S., Roy D., Aral S. The spread of true and false news online [Поширення правдивих і фейкових новин в інтернеті]. *Science*. 2018. Vol. 359, № 6380. P. 1146–1151.
4. Lazer D. M. J., Baum M. A., Benkler Y. et al. The science of fake news [Наука про фейкові новини]. *Science*. 2018. Vol. 359, № 6380. P. 1094-1096. DOI: 10.1126/science.aao2998.
5. Allcott H., Gentzkow M. Social media and fake news in the 2016 election [Соціальні медіа та фейкові новини під час виборів 2016 року]. *Journal of Economic Perspectives*. 2017. Vol. 31, № 2. P. 211-236. DOI: 10.1257/jep.31.2.211.
6. Guess A., Nagler J., Tucker J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook [Менше, ніж ви думаєте: поширення та предиктори розповсюдження фейкових новин у Facebook]. *Science Advances*. 2019. Vol. 5, № 1. P. eaau4586. DOI: 10.1126/sciadv.aau4586.
7. Shu K., Sliva A., Wang S. et al. Fake News Detection on Social Media: A Data Mining Perspective [Виявлення фейкових новин у соціальних мережах: підхід з використанням методів інтелектуального аналізу даних]. *ACM SIGKDD Explorations*. 2017. Vol. 19, № 1. P. 22–36.
8. Wang W. Y. "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection ["Брехун, брехун, штани в огні": новий набір даних для виявлення

- фейкових новин]. *Proceedings of ACL*. 2017. P. 422-426. DOI: 10.18653/v1/P17-2067.
9. Conroy N. K., Rubin V. L., Chen Y. Automatic deception detection: Methods for finding fake news [Автоматичне виявлення обману: методи пошуку фейкових новин]. *Proceedings of the Association for Information Science and Technology*. 2015. Vol. 52, № 1. P. 1-4. DOI: 10.1002/pra2.2015.145052010082.
 10. Shu K., Mahudeswaran D., Wang S. et al. FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media [FakeNewsNet: репозиторій даних з контентом новин, соціальним контекстом та просторово-часовою інформацією для дослідження фейкових новин у соцмережах]. *Big Data*. 2019. Vol. 8, № 3. P. 171-188. DOI: 10.1089/big.2020.0062.
 11. Thorne J., Vlachos A. Automated Fact-Checking: Progress and Challenges [Автоматизована перевірка фактів: досягнення та виклики]. *Annual Review of Linguistics*. 2021. Vol. 7. P. 201–220.
 12. Hassan N., Arslan F., Li C. et al. ClaimBuster: The first-ever end-to-end fact-checking system [ClaimBuster: перша повноцінна система перевірки фактів]. *Proceedings of the VLDB Endowment*. 2017. Vol. 10, № 12. P. 1945-1948. DOI: 10.14778/3137765.3137815.
 13. Zeng L., Starbird K., Spiro E. S. Fact-checking meets fake news: Emerging research and open problems [Перевірка фактів та фейкові новини: нові дослідження та відкриті проблеми]. *ACM Computing Surveys*. 2022. Vol. 55, № 1. P. 1-36. DOI: 10.1145/3487890.
 14. Pennycook G., Rand D. G. The psychology of fake news [Психологія фейкових новин]. *Trends in Cognitive Sciences*. 2021. Vol. 25, № 5. P. 388-402. DOI: 10.1016/j.tics.2021.02.007.
 15. Zhou X., Zafarani R. A survey of fake news: Fundamental theories, detection methods, and opportunities [Огляд фейкових новин: основні теорії, методи виявлення та перспективи]. *ACM Computing Surveys*. 2020. Vol. 53, № 5. P. 1-40. DOI: 10.1145/3395046.

16. Potthast M., Kiesel J., Reinartz K. et al. A stylometric inquiry into hyperpartisan and fake news [Стилометричне дослідження гіперпартійних та фейкових новин]. *Proceedings of ACL*. 2018. P. 231-240. DOI: 10.18653/v1/P18-1022.
17. Rubin V. L., Chen Y., Conroy N. K. Deception detection for news: Three types of fakes [Виявлення обману в новинах: три типи фейків]. *Proceedings of the Association for Information Science and Technology*. 2015. Vol. 52, № 1. P. 1-4. DOI: 10.1002/pra2.2015.145052010083.
18. Baly R., Karadzhov G., Alexandrov D. et al. Predicting factuality of reporting and bias of news media sources [Прогнозування достовірності та упередженості новинних джерел]. *Proceedings of EMNLP*. 2020. P. 20-30. DOI: 10.18653/v1/2020.emnlp-main.2.
19. Graves D., Cherubini F. The rise of fact-checking sites in Europe [Розвиток сайтів для перевірки фактів в Європі]. *Reuters Institute for the Study of Journalism*. 2016. 28 p. URL: <https://reutersinstitute.politics.ox.ac.uk/our-research/rise-fact-checking-sites-europe> (дата звернення: 15 трав. 2025).
20. Nguyen A. T., Kharosekar A., Krishnan S. et al. Believe it or not: Designing a human-AI partnership for mixed-initiative fact-checking [Вірте чи ні: проектування партнерства людини та ШІ для змішаної перевірки фактів]. *Proceedings of CHI*. 2018. Paper 374. P. 1-12. DOI: 10.1145/3173574.3173948.
21. Hassan N., Zhang G., Arslan F. et al. ClaimRank: Detecting check-worthy claims in Arabic and English [ClaimRank: виявлення тверджень, що потребують перевірки, арабською та англійською мовами]. *Proceedings of NAACL*. 2017. P. 26-35. DOI: 10.18653/v1/N17-1003.
22. Guo H., Cao J., Zhang Y. et al. Rumor detection with hierarchical social attention network [Виявлення чуток за допомогою ієрархічної соціальної мережі уваги]. *Proceedings of CIKM*. 2018. P. 637-646. DOI: 10.1145/3269206.3271700.
23. Popat K., Mukherjee S., Strötgen J. et al. CredEye: A credibility lens for analyzing and explaining misinformation [CredEye: інструмент аналізу та пояснення дезінформації]. *Proceedings of WWW*. 2019. P. 187-200. DOI: 10.1145/3308558.3313654.

- 24.Vo N., Lee K. The rise of guardians: Fact-checking URL recommendation to combat fake news [Поява захисників: рекомендація URL для боротьби з фейками]. *Proceedings of SIGIR*. 2020. P. 275-284. DOI: 10.1145/3397271.3401031.
- 25.Tacchini E., Ballarin G., Della Vedova M. L. et al. Some like it hoax: Automated fake news detection in social networks [Дехто любить фейки: автоматичне виявлення неправдивих новин у соцмережах]. *Proceedings of ACM Web Science*. 2017. P. 1–8. DOI: 10.1145/3091478.3091523.
- 26.Ruchansky N., Seo S., Liu Y. CSI: A hybrid deep model for fake news detection [CSI: гібридна модель глибокого навчання для виявлення фейкових новин]. *Proceedings of CIKM*. 2017. P. 797–806. DOI: 10.1145/3132847.3132877.
- 27.Pérez-Rosas V., Kleinberg B., Lefevre A. et al. Automatic detection of fake news [Автоматичне виявлення фейкових новин]. *Proceedings of COLING*. 2018. P. 3391–3401. URL: <https://aclanthology.org/C18-1287> (дата звернення: 15 трав. 2025).
- 28.Golbeck J., Mauriello M., Auxier B. et al. Fake news vs satire: A dataset and analysis [Фейкові новини vs сатира: набір даних та аналіз]. *Proceedings of ACM Web Science*. 2018. P. 17–21. DOI: 10.1145/3201064.3201100.
- 29.Oshikawa R., Qian J., Wang W. Y. A survey on natural language processing for fake news detection [Огляд методів обробки природної мови для виявлення фейкових новин]. *Proceedings of ACL*. 2020. P. 91–102. DOI: 10.18653/v1/2020.acl-main.10.
- 30.Bondielli A., Marcelloni F. A survey on fake news and rumour detection techniques [Огляд методів виявлення фейкових новин і чуток]. *Information Sciences*. 2019. Vol. 497. P. 38–55. DOI: 10.1016/j.ins.2019.05.035.
- 31.Shu K., Wang S., Liu H. Beyond news contents: The role of social context for fake news detection [За межами контенту: роль соціального контексту у виявленні фейків]. *Proceedings of WSDM*. 2019. P. 312–320. DOI: 10.1145/3289600.3290994.

- 32.Monti F., Frasca F., Eynard D. et al. Fake news detection on social media using geometric deep learning [Виявлення фейків у соцмережах за допомогою геометричного глибокого навчання]. *Proceedings of ICWSM*. 2019. P. 328–335. URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/3224> (дата звернення: 15 трав. 2025).
- 33.Kaliyar R. K., Goswami A., Narang P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach [FakeBERT: виявлення фейків у соцмережах за допомогою BERT]. *Multimedia Tools and Applications*. 2021. Vol. 80. P. 11765–11788. DOI: 10.1007/s11042-020-10183-2.
- 34.Zhang X., Ghorbani A. A. An overview of online fake news: Characterization, detection, and discussion [Огляд онлайн-фейків: характеристика, виявлення та обговорення]. *Information Processing & Management*. 2020. Vol. 57, № 2. P. 102025. DOI: 10.1016/j.ipm.2019.03.004.
- 35.Singhal S., Shah R. R., Chakraborty T. et al. SpotFake: A multi-modal framework for fake news detection [SpotFake: мультимодальна система виявлення фейків]. *Proceedings of IEEE BigMM*. 2019. P. 39–47. DOI: 10.1109/BigMM.2019.00-24.
- 36.Jwa H., Oh D., Park K. et al. exBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers [exBAKE: автоматична модель виявлення фейків на основі BERT]. *Proceedings of ACL*. 2019. P. 1–6
- 37.Kotstein S., Bogner J. REST API Design Rules: A Data-Driven Categorization and Analysis of REST Design Guidelines [Правила проектування REST API: класифікація та аналіз на основі даних] // *Proceedings of the 25th European Conference on Pattern Languages of Programs (EuroPLoP '21)*. July 2021. Article 7. P. 1–13. DOI: 10.1145/3489449.3489971.
- 38.Flask Documentation [Електронний ресурс]. URL: <https://flask.palletsprojects.com> (дата звернення: 15 трав. 2025).
- 39.Requests: HTTP for Humans [Електронний ресурс]. URL: <https://requests.readthedocs.io> (дата звернення: 15 трав. 2025).

- 40.Cachetools Documentation [Электронный ресурс].
URL: <https://cachetools.readthedocs.io> (дата звернення: 15 трав. 2025).
- 41.PostgreSQL Documentation [Электронный ресурс].
URL: <https://www.postgresql.org/docs> (дата звернення: 15 трав. 2025).
- 42.SOAX SERP API Documentation [Электронный ресурс].
URL: <https://docs.soax.com> (дата звернення: 15 трав. 2025).
- 43.OpenAI API Reference [Электронный ресурс].
URL: <https://platform.openai.com/docs> (дата звернення: 15 трав. 2025).
- 44.Jinja Documentation [Электронный ресурс].
URL: <https://jinja.palletsprojects.com> (дата звернення: 15 трав. 2025).
- 45.Docker Documentation [Электронный ресурс]. URL: <https://docs.docker.com>
(дата звернення: 15 трав. 2025).

ДОДАТКИ

ДОДАТОК А

ПРОМПТ-ІНСТРУКЦІЯ ДЛЯ АНАЛІЗУ СТАТТІ

На вхід отримуй текст статті (наукової, новинної або будь-якого іншого текстового матеріалу). Проаналізуй його та поверни структуровану відповідь у форматі JSON з наступними полями:

1. **"topic"** - основна тема статті (1-2 речення)
2. **"authors"** - список авторів (якщо вони вказані)
3. **"publication_date"** - дата публікації (якщо вона вказана)
4. **"emotional_tone"** - емоційний характер тексту (наприклад: нейтральний, позитивний, негативний, суміш емоцій тощо)
5. **"emotional_quotes"** - список цитат (1-3), які найкраще демонструють емоційний характер
6. **"key_points"** - список основних тез (1-3 цитати), які коротко передають суть тексту

Якщо якась інформація відсутня (наприклад, автор або дата), повертай null для відповідного поля.

Надай відповідь виключно у форматі JSON, без додаткових пояснень чи тексту перед/після.

Приклад відповіді:

```
{  
  "topic": "...",  
  "authors": [...],  
  "publication_date": "...",  
  "emotional_tone": "...",  
  "emotional_quotes": [...],  
  "key_points": [...],  
}
```

ДОДАТОК Б

ПРОМПТ-ІНСТРУКЦІЯ ДЛЯ ОБРОБКИ ПОСИЛАНЬ ЩОДО НОВИН

На вхід отримуй фрагменти новинних результатів (заголовок, короткий опис, дата, посилання). Проаналізуй їх і поверни структуровану відповідь у форматі JSON. Для кожної новини сформувай об'єкт з такими полями:

1. "title" — короткий заголовок новини
2. "snippet" — стислий виклад змісту новини своїми словами
3. "date" — дата публікації (у форматі YYYY-MM-DD, якщо вона є, інакше null)
4. "link" — масив з одного або кількох релевантних посилань
5. "ss" — масив назв пошукових систем (bing, yahoo, google, ddgo, yandex), в яких була ця новина, в порядку відповідно до масиву link
6. "emoji" — смайлик, що описує характер або емоційний тон новини (наприклад: 🗣️ — конфлікт, 📈 — зростання, 🗳️ — дипломатія, 😭 — трагедія, 🌺 — свято, 🗳️ — політика, 📖 — наука тощо)

Якщо якась інформація відсутня (наприклад, автор або дата), повертай null для відповідного поля. Надай відповідь виключно у форматі JSON, без додаткових пояснень чи тексту перед/після.

Приклад відповіді:

```
[
  {
    "title": "Example Title 1",
    "snippet": "Short description here.",
    "date": "2025-06-05",
    "link": ["https://example.com/news1"],
    "ss": ["yahoo", "yandex"],
    "emoji": "🗣️"
  },
  {
    "title": "Example Title 2",
    "snippet": "Another description.",
    "date": null,
    "link": ["https://example.com/news2", "https://example.com/related-article"],
    "ss": ["google"],
    "emoji": "📈"
  }
]
```

ДОДАТОК В

```
from lxml import html

import requests

WORDS_PER_LINE = 2

def extract_text(html_content):
    tree = html.fromstring(html_content)

    for element in tree.xpath('//script | //style'):
        element.getparent().remove(element)

    text = tree.text_content().strip()

    lines = list()

    for line in text.splitlines():
        if line.strip():
            line = line.strip()
            if line and len(line.split()) > WORDS_PER_LINE:
                lines.append(line)

    return '\n'.join(lines)

def get_html(url: str):
    html_content = None
    headers = {
        'User-Agent': 'Mozilla/5.0'
    }

    response = requests.get(url, headers=headers)

    if response.status_code == 200:
        html_content = response.text

    return html_content
```

Рисунок В.1 – Функції отримання HTML-розмітки («get_html») та очищення від елементів розмітки («extract_text»)

ДОДАТОК Г

```
@bp.route('/bing', methods=['POST'])
def bing():
    kwargs = request.form.to_dict()

    if not kwargs:
        kwargs = request.get_json()

    status_code = ok
    response = {'data': None,
                'error': None}

    try:
        data = _bing.search(**kwargs)
        data = bing_data_transform(data)
        response['data'] = data

    except TypeError as error:
        response['error'] = str(error)
        status_code = bad_request

    except Exception as error:
        response['error'] = str(error)
        status_code = server_error

    return json.dumps(response), status_code
```

Рисунок Г.1 – Типовий приклад обробки запиту щодо отримання результатів пошукової видачі

ДОДАТОК Д

```
▼ def transform(data: dict):
    transformed = {
        'videos': [],
        'organic_results': [],
    }
    empty = True

    if 'inline_videos' in data and 'items' in data['inline_videos']:
        transformed['videos'] = [
            {
                'link': video.get('link'),
                'duration': video.get('duration'),
                'title': video.get('title'),
                'date': video.get('date')
            }
            for video in data['inline_videos']['items']
        ]

    if 'organic_results' in data:
        transformed['organic_results'] = [
            {
                'title': result.get('title'),
                'link': result.get('link'),
                'snippet': result.get('snippet'),
                'date': result.get('date')
            }
            for result in data['organic_results']
        ]
        empty = False

    if empty:
        return None

    return transformed
```

Рисунок Д.1 – Типовий приклад функції нормалізації даних

ДОДАТОК Е

```
from openai import OpenAI
from loguru import logger
from cachetools import TTLCache
from cachetools import cached
import os

BROWSERS_CACHE_REMAINS = os.getenv('BROWSERS_CACHE_REMAINS')

cache = TTLCache(maxsize=100, ttl=float(BROWSERS_CACHE_REMAINS))

client = OpenAI()

instructions = '''На вхід отримуй текст статті (наукової, новинної або будь-якого іншого текстового матеріа

1. "topic" - основна тема статті (1-2 речення)
2. "authors" - список авторів (якщо вони вказані)
3. "publication_date" - дата публікації (якщо вона вказана)
4. "emotional_tone" - емоційний характер тексту (наприклад: нейтральний, позитивний, негативний, суміш емоц
5. "emotional_quotes" - список цитат (1-3), які найкраще демонструють емоційний характер
6. "key_points" - список основних тез (1-3 цитати), які коротко передають суть тексту

Якщо якась інформація відсутня (наприклад, автор або дата), повертай null для відповідного поля.

Надай відповідь виключно у форматі JSON, без додаткових пояснень чи тексту перед/після.

Приклад відповіді:
{
  "topic": "...",
  "authors": [...],
  "publication_date": "...",
  "emotional_tone": "...",
  "emotional_quotes": [...],
  "key_points": [...],
}'''

model = "gpt-4.1"

@cached(cache)
def output(query: str):
    response = client.responses.create(model=model, instructions=instructions, input=query)

    return response.output[0].content[0].text
```

Рисунок Е.1 – Програмна реалізація модуля запитів до API мовної моделі

ДОДАТОК Ж

```
services:

  api:
    build:
      dockerfile: Dockerfile
    container_name: app
    volumes:
      - ./:/app
    env_file:
      - .env
    ports:
      - "5000:5000"
    networks:
      - api
    depends_on:
      - db

  db:
    image: postgres:14.5
    container_name: db
    env_file:
      - .env
    ports:
      - "5432:5432"
    networks:
      - api

networks:
  api:
    driver: bridge
```

Рисунок Ж.1 – Конфігурація контейнерів під час розробки («docker-compose.yml»)

ДОДАТОК И

Аналіз статті

Основна тема

Стаття розповідає про операцію «Павутина», масштабну спецоперацію СБУ із залученням морських дронів в Криму, що передувала ураженню значної кількості стратегічної авіації Росії на її аеродромах.

Автори

null

Дата публікації

15 червня 2025

Емоційний тон

нейтральний із елементами гордості та важливості

Цитати з емоційним вмістом

"Цей дрон успішно виконав кілька бойових місій, сумарно пройшовши понад 4000 кілометрів Чорним морем і щоразу безпечно повертаючись на базу."

"Військові експерти вже назвали цю операцію «російським Перл-Харбором»."

"Речник Повітряних сил ЗСУ Юрій Ігнат заявив, що внаслідок знищення значної частини стратегічної авіації РФ, Росія на тривалий час втратила здатність до масованих ракетних ударів по території України."

Ключові тези

- Ключові елементи операції «Павутина», під час якої було уражено близько 34% усієї стратегічної авіації Росії, були відпрацьовані ще у 2023 році під час атак у Криму.
- На Sea Baby випробували нові технології, які потім використали в операції «Павутина».
- За попередніми оцінками, вартість авіаційної техніки, враженої під час спецоперації Служби безпеки України під назвою «Павутина», перевищує 7 мільярдів доларів США. Загальні втрати російської авіації внаслідок удару становлять 41 одиницю.

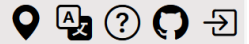
Рисунок И.1 – Приклад звітності щодо аналізу статті

ДОДАТОК К

FACTSHECK

Пошук

Аналіз



ah-10673615.html?utm_source=ukmet_news

Операція Павутина - СБУ провела аналог операції з дронами в Криму - Главред
Новини Погляди Аналітика Інтерв'ю
Популярне Weekend Ексклюзиви Погляди Аналітика Інтерв'юЧати
ДослідженняВідеоФотоПопулярне Україна Фронт Світ Війна Політика Сноупик Наука Спорт Техно та IT ЕкономікаWeekend Stars
Рецепти Лайфхаки Життя Кино та ТБ Сад і город Мода та краса Гороскоп Світа Культура Авто Зірки ПрикметиПро насРекламі
політикаКонтактиСонмережіРусЧитати російськоюНовини «ВійнаСБУ провела аналог "Павутини" в Криму: ЗМІ дізналися про операцію на "мінімалах"»
15 червня 2025, 18:31 594
На морському дроні Sea Baby тестували нові технології, що лягли в основу операції "Павутина".
https://glavred.net/war/sbu-provela-analog-pavutiny-v-krimu-smi-uznali-ob-ogratsii-na-minimalakh-10673615.htmlПосилання скопійованеНа Sea Baby тестували нові технології, що лягли в основу операції "Павутина" / Колаж: Главред, фото: скріншот з відео СБУКоротко:Ключові елементи операції "Павутина" відпрацьовувалися у 2023 роціМорський дрон Sea Baby знищив російські гелікоптери в КримуНа Sea Baby випробували нові технології, які потім використали в операції "Павутина"Ключові елементи операції "Павутина", під час якої було уражено близько 34% усієї стратегічної авіації Росії, були відпрацьовані ще в 2023 році під час атак в

{ } Стаття розповідає про операцію «Павутина», масштабну спецоперацію СБУ із залученням морських дронів в Криму, що передувала ураженню значної кількості стратегічної авіації Росії на її аеродромах.

PDF JSON 15 червня 2025

Емоційний тон: нейтральний із елементами гордості та важливості
Цей дрон успішно виконав кілька бойових місій, сумарно пройшовши понад 4000 кілометрів Чорним морем і щоразу безпечно повертаючись на базу.

Військові експерти вже назвали цю операцію «російським Перл-Харбором».

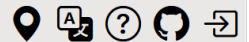
Речник Повітряних сил ЗСУ Юрій Ігнат заявив, що внаслідок знищення значної частини стратегічної авіації РФ, Росія на тривалий час втратила здатність до масованих ракетних ударів по території України.

Рисунок К.1 – Приклад результату аналізу

FACTSHECK

Пошук

Аналіз



ra-vid-vizytu-cherez-pytannya-tranzytu-gazu/

Пеллегріні планує приїхати до України після того, як відмовився від візиту через питання транзиту газу – Рубрика
Укр Рус Eng
Освіта і наука
спецпроектМІНІ: просто про безпеку спецпроектРазом діємо. Разом ми — Європа спецпроектГолоси українців за кордоном
15 Чер 2025Пеллегріні планує приїхати до України після того, як відмовився від візиту через питання транзиту газу
Фото: AP/Darko Bandic
Уляна ГузьПрезидент Словаччини Петер Пеллегріні знову планує відвідати українську столицю, хоча ще в січні скасував візит через припинення транзиту російського газу.Про це він сказав в етері STVR, передає Pravda, інформує Рубрика.За словами політика, він уже повідомив українського колегу Володимира Зеленського про початок підготовки до візиту."Я попросив українську сторону надати певний зміст, щоб це не був просто візит ввічливості. Щоб ми могли знайти якийсь позитивний зміст для обох сторін, щоб ми послали позитивний сигнал", — сказав словацький президент.За його словами, потрібно дивитися в майбутнє після війни, коли Україна почне відбудовуватися та розвиватися."Словаччина, яка розташована на кордоні з Україною, що межує з Європейським Союзом, може отримати від цього користь, і це дійсно може стати величезним поштовхом для усунення регіональних відмінностей усередині Словаччини", — додав Пеллегріні.Зазначимо, що наприкінці 2024 року президент Словаччини

{ } Президент Словаччини Петер Пеллегріні планує відвідати Україну після скасування попереднього візиту через питання транзиту російського газу і політичних суперечностей щодо санкцій проти Росії.

PDF JSON Уляна Гузь 15 Чер 2025

Емоційний тон: нейтральний з елементами стриманого оптимізму
"Я попросив українську сторону надати певний зміст, щоб це не був просто візит ввічливості. Щоб ми могли знайти якийсь позитивний зміст для обох сторін, щоб ми послали позитивний сигнал", — сказав словацький президент.

"Словаччина, яка розташована на кордоні з Україною, що межує з Європейським Союзом, може отримати від цього користь, і це дійсно може стати величезним поштовхом для усунення регіональних відмінностей усередині Словаччини", — додав Пеллегріні.

"Протягом останніх років з початку війни ми дуже тісно співпрацювали

Рисунок К.2 – Приклад результату аналізу

ДОДАТОК Л

🔍 операція павутина



🌐 Операція «Павутина» – Wikipedia
uk.wikipedia.org
1 червня 2025 року Служба безпеки України здійснила завершальний етап операції «Павутина» з ураження російської стратегічної авіації в ході повномасштабної російсько –

📰 Операція СБУ "Павутина": скільки бомбардувальників –
bbc.com
"Скидати Київ з рахунків не варто". Зухвала операція СБУ - важливий сигнал Росії і Заходу 2 червня 2025
Jun 2, 2025

📰 Операція Павутина - що відомо про операцію СБУ і ураження ...
fakty.com.ua
Операція Павутина - що відомо про унікальну операцію СБУ під керівництвом Василя Малека, атаки дронами по російських аеродромах, ураження понад 40 літаків і ...
Jun 1, 2025

📰 Операція Павутина - свіжі новини - фото та відео наслідків ...
obozrevatel.com
Операція Павутина - читайте останні новини. Дізнайтеся, хто підготував і як провели операцію з атаки на російські аеродроми та стратегічну авіацію.
Jun 1, 2025

Що це за операція "Павутина"
espresso.tv
Попри заяви Дональда Трампа, що в Україні " немає карт" у війні, операція СБУ "Павутина" довела протилежне. У виданні кажуть , що операція, гідна шпигунського ...
Jun 2, 2025

📰 Операція СБУ Павутина – що про неї відомо і реакція Заходу ...
tsn.ua
За даними західних ЗМІ, таємна операція України з використанням безпілотників, під час якої в РФ уражено 41 ворожий літак стратегічної авіації, ...
Jun 2, 2025

СБУ розкрила нові деталі підготовки "Павутини": що відомо ...
suspilne.media
Відомство опублікувало нове відео, у якому детально розповідає про те, як готувалася масштабна операція "Павутина", яка стала відповіддю на російські масовані ...
4 days ago

🔍 операція павутина



🌐 Операція «Павутина»
uk.wikipedia.org
Цілі цієї операції стали 5 аеродромів на території РФ: «Олень» в Мурманській, «Біла» в Іркутській, «Дегілево» в Рязанській, «Іваново» в Іванівській та «Україна» в ...

📰 Все, що відомо про операцію «Павутина» від СБУ ...
youtube.com
Операція СБУ під назвою «Павутина», яку здійснили на чотирьох аеродромах РФ – наймасштабніша українська операція щодо знищення російської ...

📰 Операція Павутина: невідвірна атака на авіабазу РФ!
youtube.com
Операція "Павутина": невідвірна атака на авіабазу росії Оберіть зручний спосіб підтримки проєкта 🇺🇦: 🌐 https://uadd.me/GTBT_UA 1 червня Служба 6 ...

Історична операція СБУ «Павутина» крізь призму ...
ukrinform.ua
1 червня 2025 року Україна вразила світ, провівши операцію "Павутина" – атаку на п'ять російських авіабаз, знищивши за сотні кілометрів від сво ...
Jun 2, 2025

📰 Операція Павутина - Кремль пригрозило відповіддю – ТСН
tsn.ua
У Кремлі пригрозили Україні відповіддю на операцію "Павутина", під час якої було уражено чотири російські військові аеродроми. Про це заявив ...
Jun 4, 2025

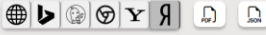
📰 У ЦІ ХВИЛИНИ! Палають понад 40 ЛІТАКІВ РФ! РОЗКРИТО ...
youtube.com
Масштабна спецоперація СБУ "Павутина". Це операція зі знищення стратегічної бомбардувальної авіації в тилу РФ. Дрони СБУ відпрацьовують по ...

📰 Операція СБУ "Павутина": скільки бомбардувальників ...
bbc.com
Внаслідок ударів українських дронів по аеродромах російської дальньої та стратегічної авіації 1 червня знищили або пошкодили понад 11 літаків, ...
Jun 2, 2025

📰 Операція "ПАВУТИНА" матиме наслідки для СЕИТУ ...
youtube.com
Операція «Павутина» – це не просто дрони. Це початок нової епохи війни! Генерал Омельченко розкривав шокуєчі деталі: як українські ...

📰 Операція "Павутина": як вплине на перебіг війни і миру ...

🔍 операція павутина



🌐 Operation Spider's Web - Wikipedia
en.m.wikipedia.org
Operation Spider's Web (Ukrainian: Операція «Павутина», romanized: Operatsiia "Pavutyna") was a covert drone attack carried out by the Security Service of...

📰 Ukraine's Pavutyna operation: 18 months in planning, 41...
yahoo.com
The Security Service of Ukraine (SSU) conducted a special operation codenamed Pavutyna (Web) on 1 June 2025, striking 41 Russian strategic aircraft...Missing: онепація

📰 Операція Павутина або коментар для рускі / Operation...
youtube.com

📰 Ukraine's Operation 'Pavutyna' strikes major blow to Russian...
euro-sd.com
Footage from a Ukrainian FPV drone during Operation 'Pavutyna' as it overflies an already-targeted Russian Tu-95. The tyres that the footage showed placed...Missing: онепація

📰 SBU Operation Pavutyna: How Ukraine Dealt a Devastating...
en.defence-usa.com
June 1, 2025 – This date will be remembered as a pivotal moment in the russian-ukrainian war, when Ukraine's Security Service (SBU) executed a bold and...Missing: онепація

📰 Operation Spider Web, Objectives, Significance and Future...
studyiq.com
It was codenamed "Operation Spider Web" (Ukrainian: Операція «Павутина», romanized: Operatsiia "Pavutyna") by the Security Service of Ukraine (SBU).

📰 More than 40 Russian aircraft were hit as... - AERONAUT.media
aeronaut.media
Спецоперація "Павутина". On June 1, Ukraine's Security Service carried out a special operation known as "Pavutyna" ("Web").Missing: онепація

📰 Inside Ukraine's Secret Deep Strike Against... | Galaxy.ai
galaxy.ai
Ukraine executed a covert operation, dubbed Operation Pavutyna, deploying drones to strike Russian air bases, resulting in significant damage to Russia's military...Missing: онепація

Рисунок Л.1 – Приклад результатів пошуку з різних пошукових систем