

МАРЧУК АРТЕМ ІГОРОВИЧ

Допускається до захисту:

в.о. завідувача кафедри
інформаційних технологій

канд. техн. наук, доцент

_____ О. В. Зелінська

« _____ » _____ 2025р.

**РОЗРОБКА СИСТЕМИ ЗБОРУ ТА АНАЛІЗУ ДАНИХ СОЦІАЛЬНИХ
МЕРЕЖ**

Спеціальність 122 Комп'ютерні науки

Кваліфікаційна (бакалаврська) робота

Керівник:

Ю.В. Поремський,
старший викладач
кафедри інформаційних
технологій, канд. техн.
наук

Оцінка: _____ / _____ / _____
(бали/за шкалою ЕКТС/за національною шкалою)

Голова ЕК: _____

АНОТАЦІЯ

Марчук А.І. Розробка системи збору та аналізу даних з соціальних мереж. Спеціальність 122 «Комп'ютерні науки». Донецький національний університет імені Василя Стуса, Вінниця, 2025.

У кваліфікаційній (бакалаврській) роботі наведено результати розробки Viber-бота для збору та аналізу інформації з профілів конкурентів у соціальних мережах (Instagram, Facebook). Система дозволяє вставляти посилання на акаунти конкурентів і отримувати структуровані дані про їхню активність, аудиторію, частоту публікацій, популярні товари тощо. Робота включала проектування архітектури системи, реалізацію бота, створення графічного інтерфейсу, а також тестування працездатності та зручності використання.

Ключові слова: соціальні мережі, Viber-бот, збір даних, аналіз конкурентів, інтерфейс користувача, Instagram, Facebook, автоматизація.

57 с., 18 рис. 4 табл., 39джерел.

ABSTRACT

Marchuk A.I. Development of a system for collecting and analyzing data from social networks. Specialty 122 "Computer Science". Vasyl Stus Donetsk National University, Vinnytsia, 2025.

The qualifying (bachelor's) paper presents the results of developing a Viber bot for collecting and analyzing information from competitors' profiles in social networks (Instagram, Facebook). The system allows inserting links to competitors' accounts and retrieving structured data about their activity, audience, publication frequency, popular products, etc. The work included system architecture design, bot implementation, creation of a graphical interface, as well as usability and functionality testing.

Keywords: social networks, Viber bot, data collection, competitor analysis, user interface, Instagram, Facebook, automation.

57 pages, 18 figures, 4 tab., 39 references.

ЗМІСТ

ВСТУП.....	4
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ЗБОРУ ТА АНАЛІЗУ ДАННИХ З СОЦІАЛЬНИХ МЕРЕЖ	6
1.1 Поняття та значення аналізу даних у соціальних мережах	6
1.2 Методи збору даних із соціальних мереж.....	9
1.3 Особливості аналізу конкурентів за відкритими даними	16
РОЗДІЛ 2. Аналіз вимог до системи збору та аналізу даних.....	23
2.1 Огляд існуючих рішень і підходів	23
2.2 Визначення вимог до системи та її проектування	27
2.3 Вибір технологій для реалізації системи	33
РОЗДІЛ 3. Проектування та реалізація системи збору та аналізу даних	37
3.1 Загальна логіка роботи системи	37
3.2 Розробка модуля збору даних.....	40
3.3 Інтерфейс та демонстрація роботи Viber-бота	47
ВИСНОВКИ.....	53
СПИСОК ВИКОРИСТАНИХ ПОСИЛАНЬ.....	55

ВСТУП

Нові реалії цифрового світу вимагають від бізнесу, аналітиків та маркетологів миттєвого доступу до актуальної інформації про конкурентів, тренди та поведінку цільової аудиторії. Хоча соціальні мережі стали основним джерелом загальнодоступних даних, що містять корисну інформацію для прийняття рішень, ручна обробка цих даних є неефективною і забирає багато часу.

Розвиток інформаційних технологій, а також зростання соціальних платформ, таких як Instagram та Facebook, створили попит на автоматизовані інструменти збору та аналізу даних. Необхідність розробки системи, яка може швидко аналізувати загальнодоступні дані з профілів конкурентів для визначення їхньої аудиторії, структури контенту, стратегії просування та інших важливих факторів, диктує актуальність дослідження.

Метою даної бакалаврської роботи є створення системи, яка використовує Viber-бот для автоматизації процесу відстеження та обробки інформації про конкурентів з метою збору та аналізу даних з соціальних мереж. Очікується, що розроблене рішення матиме інтуїтивно зрозумілий інтерфейс користувача разом з аналітичними можливостями для маркетингових потреб.

Об'єктом дослідження є процес створення програмного забезпечення для збору та аналізу даних із соціальних мереж. Предметом дослідження є методи, алгоритми та інструменти побудови Viber-бота для збору та структурованого представлення інформації з Instagram та Facebook.

Для досягнення поставленої мети були сформульовані такі завдання:

1. Провести огляд існуючих підходів до збору даних з соціальних мереж та аналізу конкурентів.
2. Визначити технічні вимоги до системи збору та аналізу даних.
3. Розробити архітектуру та реалізувати Viber-бота для взаємодії з користувачем.

4. Реалізувати модулі збору, обробки та структурування даних з обраних соціальних мереж.
5. Провести тестування системи на функціональність, стабільність та зручність використання.
6. Надати рекомендації щодо розширення функціональності системи в майбутньому.

Стаття складається зі вступу, трьох основних розділів, висновків та списку використаних джерел. Теоретичні основи аналізу соціальних мереж та методи збору даних висвітлені в першому розділі. У другому розділі описано проектування та визначення системних вимог. Результати реалізації, тестування та приклади використання бота наведені в третьому розділі.

Розроблена система може бути корисною для бізнесу, маркетологів та експертів з маркетингу в соціальних мережах для автоматизованого аналізу конкурентів у соціальних мережах. Саме в цьому полягає практична цінність роботи.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ОСНОВИ ЗБОРУ ТА АНАЛІЗУ ДАННИХ З СОЦІАЛЬНИХ МЕРЕЖ

1.1 Поняття та значення аналізу даних у соціальних мережах

Соціальні мережі стали невіддільною частиною повсякденного життя, слугуючи платформами для спілкування, обміну думками та взаємодії між людьми. Щодня мільйони користувачів по всьому світу публікують дописи, діляться фотографіями, залишають коментарі та вподобання, беруть участь у групах і дискусіях. Уся ця активність залишає цифровий слід — дані, які можуть містити цінну інформацію для подальшого аналізу.

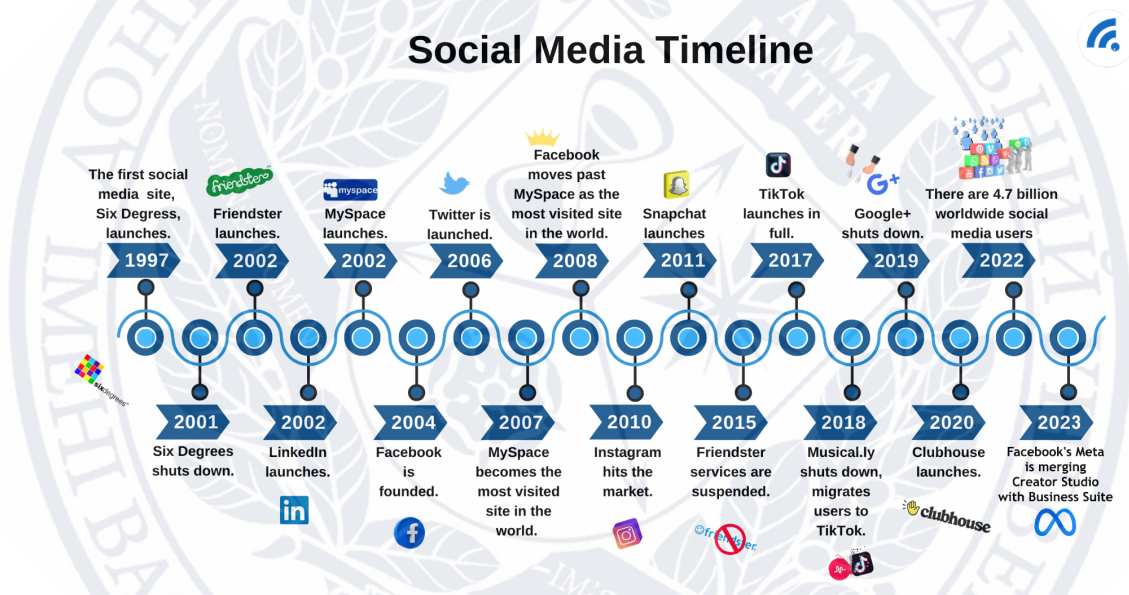


Рисунок 1.1 – Хронологія розвитку соціальних мереж

Історія соціальних мереж і формування цифрових даних:

1. Зародження соціальних мереж (1990-ті — початок 2000-х):
Перші онлайн-платформи, як-от SixDegrees (1997), LiveJournal (1999) та Friendster (2002), започаткували ідею цифрових соціальних зв'язків. Дані користувачів на цьому етапі обмежувалися профілями, текстовими записами й базовими взаємодіями.

2. Еволюція платформ і вибух даних (2004–2012):
З появою Facebook, YouTube, Twitter соціальні мережі почали масово збирати дані про поведінку користувачів. Зростає кількість форматів контенту — фото,

відео, гіперпосилання, геолокації, а також обсяг взаємодій — лайки, шери, коментарі. У цей період з'являються перші інструменти аналітики для бізнесу.

3. Соціальні мережі як джерело Big Data (після 2012): Мобільні додатки та нові платформи — Instagram, TikTok, Snapchat — додали нові типи даних: сторіс, стріми, емодзі, реакції в реальному часі. API, парсери та технології штучного інтелекту почали широко застосовуватися для збору та обробки цифрової інформації.

Роль соціальних даних у сучасному суспільстві: Соціальні мережі стали каналом формування суспільної думки, цифровою сценою для політичної боротьби, маркетинговим інструментом і джерелом соціологічного аналізу. Водночас вони поставили перед дослідниками завдання — відповідально працювати з чутливою, часто публічною, але особистісною інформацією.

Для прийняття обґрунтованих рішень у бізнесі, маркетингу, політиці, соціології та інших сферах аналіз даних соціальних мереж передбачає збір, обробку, аналіз та візуалізацію інформації з різних платформ, включаючи Facebook, Instagram, Twitter, TikTok, та інші. Це багатогранний метод дослідження поведінки користувачів, обміну інформацією та динаміки цифрової взаємодії. На відміну від традиційних опитувань чи звітів, аналіз соціальних даних дозволяє вивчати поведінку в режимі реального часу, незважаючи на величезні обсяги, різноманітність і швидкість змін інформації Big Data[1].

Соціальні мережі слугують віртуальним вікном у суспільство, відображаючи політичні переконання, сприйняття брендів, тенденції, емоції та реакцію на події[2]. Як наслідок, вивчення цих даних може дати не лише бізнес-інсайти, але й глибше соціальне, культурне та політичне розуміння.

Ідея аналізу даних соціальних мереж є особливо важливою в умовах інформаційного перевантаження. Вона встановлює методологію роботи з величезними обсягами цифрових даних, які продукують користувачі. Знання

основ цього процесу дозволяє як систематизувати дані, так і перетворювати їх на стратегічно важливі ідеї. Методичний процес збору, обробки, оцінки та візуалізації даних, які створюють користувачі на платформах соціальних мереж, відомий як аналітика соціальних мереж (Social Media Analytics, SMA). Вона дозволяє отримати корисні дані про емоційний тон постів, поведінку аудиторії, тенденції та ефективність маркетингових ініціатив. Основною метою цього підходу є отримання висновків, що сприяють прийняттю обґрунтованих рішень у сферах бізнесу, політики, медіа та досліджень.

Значення аналізу соціальних даних постійно зростає, оскільки соціальні мережі стають ключовими каналами впливу на громадську думку. Застосування таких інструментів набуває особливого значення для прогнозування трендів, оцінки впливу подій і розуміння глибинних суспільних процесів.

При цьому важливо враховувати етичні аспекти роботи з цифровими даними. Попри публічний характер багатьох даних, слід дотримуватися норм конфіденційності, етичного використання інформації та вимог законодавства, таких як GDPR в ЄС.

Ще один виклик — достовірність і чистота даних. У соціальних мережах поширюється багато фейкової інформації, бот-активності та спроб маніпуляцій. Тому критично важливим є виявлення фальшивих профілів, ідентифікація ботів, відсіювання недостовірного контенту.

Успішна реалізація аналітики соціальних мереж базується на мультидисциплінарному підході — із залученням знань з інформатики, соціології, психології, маркетингу та візуалізації. Такий підхід дозволяє глибше інтерпретувати дані в контексті реального соціального середовища.

У перспективі аналіз даних соціальних мереж розвиватиметься завдяки штучному інтелекту, машинному навчанню та технологіям обробки природної мови. Це відкриває нові можливості — автоматичне виявлення емоцій, розпізнавання тем і маніпуляцій, а також прогнозування поведінки великих груп користувачів. Таким чином, аналітика соціальних мереж стає не просто інструментом, а повноцінним джерелом стратегічної інформації.

1.2 Методи збору даних із соціальних мереж

Соціальні медіа перетворилися на величезну колекцію даних про взаємодію, поведінку, погляди та інтереси мільйонів людей по всьому світу, а також стали інструментом для комунікації в сучасному цифровому світі. Величезні обсяги даних у режимі реального часу постійно генеруються такими платформами, як Facebook, Instagram, X (раніше Twitter), TikTok, Reddit, LinkedIn та іншими. Ці дані — різноманітні за форматом, високочастотні та масштабні — відносяться до явища Big Data (великі дані). Вони охоплюють текстові пости, коментарі, фото, відео, реакції, геолокацію та метадані, і є основою для глибокої соціальної, маркетингової та поведінкової аналітики.



Рисунок 1.2 – Соцмережі як джерело Big Data

Наукові дослідження, політичний аналіз, журналістські розслідування, вивчення громадської думки, управління репутацією, моніторинг кризових ситуацій, маркетингові стратегії, створення алгоритмів машинного навчання та штучного інтелекту - все це широко використовує дані соціальних мереж. Така адаптивність і масштабність є саме тією причиною, чому потрібні ефективні, моральні та технологічно обґрунтовані методи збору даних.

Проте, збирання даних із соціальних мереж — не тривіальне завдання. Воно має низку викликів: правові обмеження, політику конфіденційності, технічні бар'єри, обмеження API, ризики блокування, а також етичні дилеми, пов'язані з приватністю користувачів. Платформи змінюють свої умови використання, змінюють структуру даних, або взагалі закривають доступ до деяких функцій. Крім того, існує ризик викривлення вибірки, адже не всі групи

користувачів однаково активні онлайн, і не всі дані однаково доступні.

Перед тим як аналізувати методи збору інформації, доцільно окреслити основні типи даних, які продукуються користувачами соціальних платформ і підлягають збору та обробці. Дані із соціальних мереж є багатовимірними й охоплюють різні аспекти цифрової активності:

- Текстові дані — пости, коментарі, теги, підписи до зображень або відео. Це основне джерело для лінгвістичного аналізу, вивчення тематики, наративів та настроїв.
- Мультимедійні дані — фотографії, відео, GIF-файли, зображення, меми. Їхня обробка передбачає використання методів комп'ютерного зору та контентного класифікування.
- Мета-дані — дата та час публікацій, геолокаційні мітки, тип пристрою, браузер. Вони дозволяють ідентифікувати часові закономірності та контекст взаємодій.
- Поведінкові дані — кількість кліків, переглядів, тривалість перегляду контенту, тип взаємодії. Важливі для аналізу рівня залученості та користувацького досвіду.
- Дані про соціальні зв'язки — підписки, лайки, шери, згадки, взаємні взаємодії, мережеві граfi. Використовуються в соціальному аналізі, побудові графів і карт спільнот.
- Аналітичні метрики — охоплення публікацій, рівень залучення, частота оновлення контенту. Це агреговані показники, які оцінюють ефективність сторінки або кампанії.
- Емоційні індикатори — реакції (лайк, обурення, сміх), використання смайлів та емодзі, тональність тексту. Вони дозволяють зрозуміти емоційне сприйняття контенту аудиторією.

У відповідь на зростаючу складність та багатовимірність даних, що генеруються у соціальних мережах, було сформовано низку підходів до їх збору та обробки. Ці методи умовно класифікуються на чотири основні категорії,

кожна з яких відповідає специфіці типів даних та цілей дослідження:

1. Використання офіційних API — метод, який дозволяє легально та структуровано отримувати доступ до публічних (а іноді й частково закритих) даних, дотримуючись політики конкретної платформи. Цей підхід є найбільш прозорим і стабільним, однак має обмеження за обсягом запитів, типами доступних даних та необхідністю реєстрації застосунків.
2. Веб-скрапінг — техніка автоматизованого зчитування контенту з вебсторінок, що часто використовується у випадках, коли API обмежений або не надає потрібної інформації. Цей метод вимагає обережності: він може порушувати умови користування платформи, а також викликає етичні питання щодо збору особистих даних без згоди.
3. Використання соціальних плагінів та трекінгових скриптів — менш очевидний, але поширений метод збору даних про взаємодію користувачів із соціальним контентом на сторонніх сайтах або в мобільних додатках. Ці інструменти дозволяють відстежувати поведінку користувача, формувати профілі, аналізувати активність у динаміці.
4. Ручний збір і краудсорсинг — більш трудомісткий, але гнучкий підхід, який включає ручний моніторинг сторінок, архівування контенту, аналіз скріншотів, а також залучення користувачів до збору даних через анкети, опитування або інтерактивні боти.

Одним із найбільш організованих і технічно зручних способів отримання даних із соціальних мереж є використання офіційних API. Вони надають доступ до даних через формалізовані запити, що дозволяє працювати з великими обсягами інформації без порушення політики платформи. Розглянемо, як функціонує цей механізм і чому він є базовим інструментом для аналітиків, розробників та дослідників.

API (Application Programming Interface) — це офіційний технічний

інструмент, що дає змогу взаємодіяти з платформою не вручну, а через чітко визначені запити. Він дозволяє отримати дані про публікації, реакції, коментарі, підписки тощо, дотримуючись політики конфіденційності.

Соціальні мережі, як-от Facebook, Instagram, Twitter (X), TikTok, YouTube, Reddit, надають власні API. Через них можна автоматично збирати інформацію з відкритих джерел: сторінок, профілів, спільнот, відео. Наприклад, Facebook Graph API дозволяє отримати дані про бізнес-акаунти, охоплення публікацій, коментарі, вподобання. Instagram API доступне лише для бізнес- і авторських акаунтів. YouTube API відкриває доступ до каналів, відео, коментарів і переглядів. TikTok API обмеженіший і переважно орієнтований на маркетинг.

Принцип роботи API такий: користувач формує запит із зазначенням потрібної інформації. Платформа повертає відповідь у форматі JSON — зручно для автоматизованої обробки та аналізу великих обсягів даних.

Переваги API — це легальність, структурованість даних, зручність фільтрації (за датою, типом контенту тощо) та дотримання конфіденційності. Проте є й обмеження: квоти на кількість запитів, платний доступ (наприклад, у Twitter), неможливість отримання особистих даних без згоди користувачів.

Для роботи з API потрібні базові технічні знання: розуміння HTTP-запитів, авторизації (OAuth), обробки JSON. Необхідно зареєструвати застосунок, отримати токен доступу та дотримуватись документації. Такий підхід забезпечує безпеку та прозорість у використанні ресурсів платформи.

Веб-скрапінг — це свого роду технічна імпровізація там, де офіційні канали, як-от API, або недоступні, або надто обмежені. Уявімо ситуацію: ви хочете дослідити, як користувачі реагують на пости в соціальній мережі, але офіційний API або не надає таких даних, або просить за них немалі гроші. В цьому випадку єдиний варіант — зайти на сторінку, "прочитати" її HTML-код і витягти звідти все, що потрібно. Але не вручну, а автоматизовано — скриптом.

По суті, веб-скрапінг — це спосіб дати машині очі: вона "заходить" на сайт, "дивиться", що там розміщено, і витягує потрібну інформацію. Це не так елегантно, як робота через API, але працює. Іноді — навіть краще, бо можна

дістатися до тих фрагментів, які офіційно не віддаються.

Щоправда, така "неофіційність" має свою ціну. По-перше, сайти активно борються з подібним доступом — додають захисти, динамічний контент, антибот-механізми. Зламав структуру сторінки — і твій скрапер уже нічого не може прочитати. По-друге, з етичного погляду — не все, що технічно можливо, допустиме юридично. Багато платформ прямо вказують у своїх умовах: не скрапити. І хоча це не завжди переслідується, ризик є.

Технічно веб-скрапінг — це свого роду конструктор: потрібно самому зрозуміти, де яка інформація на сторінці, прописати логіку, як її витягувати, і слідкувати, щоб усе не зламалося з наступним оновленням дизайну. Це складніше, ніж використання API, але й більш гнучко.

Особливо це корисно у випадках, коли потрібно зібрати нетипову інформацію, або дані розміщені у відкритому доступі, але без жодного API. Наприклад, форуми, коментарі до новин, окремі профілі користувачів. Там скрапер стає незамінним інструментом.

Наступний спосіб отримання даних із соціальних мереж — використання соціальних плагінів та трекінгових скриптів. На відміну від API чи веб-скрапінгу, цей метод працює не з платформами напряму, а з їх інтеграцією в інші сайти — через кнопки "Поділитися", "Подобається", коментарі, вбудовані пости тощо.

Кожен плагін — це скрипт, що підвантажується з серверів соцмережі та взаємодіє і з сайтом, і з користувачем. Завдяки цьому відбувається трекінг: платформи можуть ідентифікувати користувача, відстежити його дії, час перебування на сторінці, реакції на контент.

З аналітичної точки зору — це цінне джерело даних. Соцмережі використовують його для таргетингу, рекомендацій і поведінкового аналізу. Частина цих даних доступна зовнішнім користувачам — наприклад, через інструменти Facebook Pixel або LinkedIn Insights.

Дослідники також можуть частково використовувати ці механізми: встановивши плагін на власному сайті, отримують дані про джерела трафіку,

кількість взаємодій, тип поведінки. Це не повний масив користувацьких даних, але цінний сигнал для маркетингу чи соціології.

Однак цей підхід має і проблеми. Зі зростанням вимог до захисту персональних даних (особливо в ЄС — через GDPR), соцплагіни зазнали критики. Виникло питання: чи є взаємодія з плагіном під час завантаження сторінки згодою на обробку даних? У відповідь багато сайтів впровадили попередження, обмежили запуск скриптів або стали вимагати окрему згоду.

Ручний збір даних — це метод, при якому інформація з соціальних мереж фіксується безпосередньо людиною, а не автоматизованими інструментами. Це може бути копіювання текстів постів, збереження скріншотів, фіксація коментарів, посилань або іншої активності користувачів. Незважаючи на повільність, цей підхід залишається актуальним у ситуаціях, коли інші методи недоступні або малоефективні. Зокрема, це стосується випадків швидкого видалення або змінення контенту, який потрібно зафіксувати в моменті — наприклад, тимчасові сторіс або публікації на гарячі теми.

Цей підхід часто застосовується у якісних дослідженнях, де критично важливий контекст і зміст: як користувачі формулюють думки, які емоції виражають, як змінюється риторика. Автоматичні засоби не завжди здатні вловити тонкощі мови — іронію, сарказм, приховані сенси. Людина, натомість, краще розпізнає ці нюанси.

Краудсорсинг — це колективна форма ручного збору, коли до збору інформації залучається велика кількість учасників: волонтери, активісти, дослідницькі групи. Вони отримують інструкції або шаблони, за якими збирають дані, надсилають приклади, посилання або інші матеріали. Цей метод особливо ефективний у громадських ініціативах — наприклад, під час моніторингу дезінформації, мовної ворожнечі чи порушень цифрових прав.

Обидва підходи потребують ретельної координації, перевірки достовірності зібраних матеріалів, а також чітко прописаних критеріїв збору. Попри низьку масштабованість, ручний збір і краудсорсинг залишаються

важливими інструментами там, де машинні методи не забезпечують потрібної глибини або гнучкості.

Таблиця 1.1 Порівняння методів збору даних з соціальних мереж

Метод збору даних	Переваги	Недоліки	Особливості застосування
1. Офіційні API	– Легальний і прозорий доступ до даних– Структурований формат (JSON)– Підтримка з боку платформи– Можливість фільтрації даних	– Обмеження за кількістю запитів (квоти)– Необхідність технічної підготовки– Відсутність доступу до приватної інформації	Найбільш рекомендований метод для дослідників і розробників. Підходить для регулярного збору структурованих даних.
2. Веб-скрапінг	– Гнучкість у доступі до нетипових або закритих через API даних– Можливість автоматизованого збору великих обсягів інформації	– Порушення правил платформи– Юридичні й етичні ризики– Висока чутливість до змін у верстці сайтів	Застосовується, коли API не надає необхідних даних або недоступний. Необхідні технічні навички.
3. Соціальні плагіни та трекінгові скрипти	– Дані про поведінку користувачів на сторонніх ресурсах– Можливість інтеграції в аналітичні системи– Не потребує прямої взаємодії з соцмережею	– Обмеженість доступних метрик– Потреба у власному сайті/ресурсі– Конфлікт із політиками конфіденційності (GDPR)	Підходить для аналізу реакції аудиторії на контент за межами соцмереж. Часто використовується в маркетингу.
4. Ручний збір і краудсорсинг	– Висока контекстуальна чутливість– Можливість збору "вразливого" контенту (видалені пости, сторіс)– Гнучкість у форматі та тематиці	– Висока трудомісткість– Обмежена масштабованість– Ризик суб'єктивності та помилок	Використовується в якісних і точкових дослідженнях. Особливо корисний при роботі з контроверсійним чи емоційним контентом.

1.3 Особливості аналізу конкурентів за відкритими даними

У сучасних умовах ринку моніторинг конкурентів став невід'ємною частиною бізнес-аналітики. Завдяки відкритим даним, зокрема з соціальних мереж, компанії можуть отримувати важливу інформацію про активність, стратегії, реакцію аудиторії та динаміку розвитку своїх конкурентів. Такий аналіз дає змогу своєчасно реагувати на зміни в ринковому середовищі та адаптувати власну діяльність.

Відкриті дані (open data) — це інформація, яка доступна широкому загалу без обмежень щодо доступу або використання. У контексті конкурентного аналізу ці дані можуть надходити з різноманітних цифрових джерел, де компанії свідомо або несвідомо залишають сліди своєї діяльності. Нижче наведено основні категорії джерел, з яких можливо збирати дані про конкурентів.

Одним з найважливіших і найдинамічніших джерел відкритої інформації про конкурентів є соціальні мережі. Платформи на кшталт Facebook, Instagram, Twitter, LinkedIn, YouTube та TikTok слугують не лише інструментами комунікації з аудиторією, але й каналами презентації бренду, товарів та послуг. Компанії публікують в цих мережах рекламні та інформаційні дописи, взаємодіють із користувачами, запускають рекламні кампанії, а також використовують хештеги для охоплення нових сегментів аудиторії. Аналізуючи характер і частоту цих публікацій, їхню тематику, візуальний стиль та рівень залученості (engagement), можна зробити висновки про пріоритети компанії, цільову аудиторію, впроваджені маркетингові стратегії та реакцію користувачів. Наприклад, якщо компанія різко збільшує активність у певному сегменті, це може свідчити про запуск нової лінійки продуктів або масштабування бізнесу.

Офіційні вебсайти конкурентів є ще одним важливим джерелом інформації, що містить переважно формалізовані та офіційні повідомлення. Саме через сайт компанії зазвичай комунікують свої досягнення, новини, плани, а також висвітлюють корпоративну культуру. Блоги, публікації новин, описи послуг або продуктів, сторінки «Про нас», контактна інформація, доступні вакансії — усе це формує уявлення про те, в яких напрямках працює компанія, на

які цінності вона орієнтується, яку аудиторію обслуговує. Крім того, якщо компанія є публічною, на сайті можуть бути розміщені річні звіти, прес-релізи для інвесторів або результати аудиту. Такі документи дозволяють оцінити фінансовий стан компанії, її ринки збуту, а також проєкти, в які вона інвестує.

Сайти, на яких клієнти залишають свої відгуки, дозволяють зрозуміти, як користувачі сприймають продукти та послуги конкурентів. Наприклад, на Google Reviews, Trustpilot або App Store/Google Play покупці висловлюють свої враження, описують досвід взаємодії з компанією та її сервісами, зазначають переваги й недоліки. Такі платформи дають змогу виявити частотність повторюваних скарг або навпаки — сильні сторони, що цінуються користувачами. Якщо певний бренд регулярно отримує негативні відгуки на однакову тему, це може вказувати на системну проблему в обслуговуванні або у продукті. Натомість велика кількість позитивних оцінок може свідчити про успішну реалізацію клієнтоорієнтованої стратегії.

Ще одним цінним джерелом інформації є платформи для розміщення вакансій, такі як LinkedIn Jobs, Djinni, Work.ua чи Indeed. Аналізуючи, кого саме шукає конкурент, можна зробити висновки про поточні та майбутні напрямки його розвитку. Наприклад, активне наймання спеціалістів із digital-маркетингу або data science може вказувати на зміну фокусу компанії в бік цифровізації чи посилення аналітики. Крім того, опис вакансій дозволяє з'ясувати, які технології використовуються в компанії, яку корпоративну культуру вона підтримує, яким вимогам повинні відповідати працівники — усе це допомагає створити портрет внутрішньої організації конкурентної структури.

У більшості країн існують офіційні державні ресурси, які містять відкриту інформацію про юридичних осіб. В Україні, наприклад, можна скористатися сервісами YouControl чи Opendatabot, де зібрані дані з Єдиного державного реєстру підприємств, судових реєстрів, податкової звітності тощо. Такі джерела дозволяють з'ясувати, хто є фактичним власником компанії, який її фінансовий стан, чи фігурує вона у судових процесах, чи є борги, штрафи або порушення. Така інформація корисна не лише для розуміння конкурентного середовища, а й

для оцінки ризиків, пов'язаних із потенційною співпрацею.

Аналіз конкурентів у соціальних мережах — це важливий інструмент стратегічного планування, який дозволяє виявити сильні та слабкі сторони інших учасників ринку. Систематичне вивчення їхньої цифрової поведінки дає змогу своєчасно реагувати на нові тенденції, адаптувати власні стратегії просування та вдосконалювати взаємодію з цільовою аудиторією. Для цього використовуються різноманітні методи, які дозволяють досліджувати як змістовну, так і кількісну сторону комунікації конкурентів у соціальних медіа.

Один з базових підходів до аналізу конкурентної активності — це регулярний моніторинг контенту, який публікується на офіційних сторінках компаній у соціальних мережах. Цей метод передбачає фіксацію частоти публікацій, тематики та формату матеріалів (відео, зображення, текст, прямі ефіри), а також вивчення реакції аудиторії на них. Додатково оцінюється кількість підписників, загальний темп зростання аудиторії, а також ступінь залученості користувачів — кількість лайків, коментарів, репостів та збережень. Такий аналіз дозволяє оцінити ефективність контент-стратегії конкурента, виявити популярні формати подачі інформації та часові рамки найбільш активної взаємодії з аудиторією.

Ще одним важливим методом є аналіз тональності публікацій і згадок бренду, який дозволяє визначити, яким чином користувачі сприймають компанію в інформаційному просторі. Цей підхід базується на технологіях обробки природної мови (Natural Language Processing, NLP), які автоматично класифікують тексти на позитивні, негативні або нейтральні за емоційним забарвленням. Аналізуючи коментарі, дописи, відгуки, а також згадки у сторонніх публікаціях, можна зробити висновки про репутацію конкурента, фактори лояльності або невдоволення клієнтів. Особливу цінність становлять періоди різкої зміни тональності, що можуть свідчити про кризу, помилку в комунікації чи зміну стратегії.

Ретельний аналіз взаємодій користувачів із контентом конкурентів дозволяє побудувати уявлення про їхню цільову аудиторію. Наприклад,

вивчаючи, хто саме ставить вподобання, залишає коментарі або поширює дописи, можна визначити соціально-демографічні характеристики підписників (стать, вік, географія), а також зрозуміти їхні інтереси та поведінкові патерни. Така інформація допомагає порівнювати сегменти ринку, в які орієнтовані різні гравці, і виявляти незаповнені ніші. В окремих випадках можливо навіть виявити впливових користувачів (інфлюенсерів), які активно взаємодіють із конкурентом, і потенційно залучити їх до власної маркетингової кампанії.

Рекламна активність конкурентів є одним із найбільш критичних аспектів для спостереження. За допомогою відкритих інструментів, таких як Facebook Ad Library, можна отримати доступ до всіх активних реклам, які запускає компанія у Facebook та Instagram. Це дозволяє оцінити частоту появи оголошень, їх візуальний стиль, ключові меседжі, пропозиції та цільову аудиторію. Також корисним є використання сервісів типу SimilarWeb, які дозволяють проаналізувати джерела трафіку на сайт конкурента, побачити ефективність платного просування в різних каналах, а також виявити ключові запити у пошукових системах, які приносять відвідувачів. У сукупності це створює повну картину цифрової маркетингової активності конкурентів.

Останній метод полягає у створенні набору метрик, які дозволяють порівняти активність кількох компаній на однакових умовах. Наприклад, можна оцінювати кількість публікацій за певний період, середній рівень залучення на одну публікацію, темпи зростання кількості підписників, стабільність виходу контенту, частку рекламних матеріалів у загальному обсязі публікацій тощо. Порівняльний підхід дозволяє не лише об'єктивно вимірювати ефективність, але й виявляти найуспішніші практики, які можуть бути адаптовані у власну стратегію.

Попри широкі можливості, які надає аналіз відкритих даних у соціальних мережах, слід враховувати низку важливих особливостей і обмежень. Вони стосуються як технічних аспектів доступу до інформації, так і правових та етичних меж, що регулюють обробку персональних і корпоративних даних. Знання цих обмежень є критично важливим для коректної побудови системи

збору й аналізу даних, аби забезпечити не лише ефективність, а й законність такого процесу.

Не вся інформація в соціальних мережах є однаково доступною для збору та подальшого аналізу. Значна частина контенту може бути закритою налаштуваннями конфіденційності: наприклад, приватні сторінки, дописи, доступні лише друзям або певним групам користувачів. Крім того, платформи дедалі частіше впроваджують технічні бар'єри, які перешкоджають масовому збору даних через автоматизовані інструменти, зокрема веб-скрейпінг. Такі обмеження можуть виражатися у вигляді CAPTCHA, обмежень частоти запитів або навіть юридичних попереджень при порушенні політики використання API. У результаті аналітики змушені працювати з неповною картиною, що впливає на точність висновків.

Інформація в соціальних мережах є високодинамічною — дописи можуть редагуватися або видалятися, акаунти — деактивуватися, а загальна політика платформи — змінюватися в будь-який момент. Наприклад, компанія-конкурент може змінити свою контент-стратегію, очистити стрічку новин або перевести сторінку в архів. Це ускладнює побудову довготривалих аналітичних зрізів і вимагає регулярного оновлення баз даних. Також нестабільність ускладнює відтворення результатів дослідження — одна й та сама сторінка через кілька тижнів або місяців може виглядати кардинально інакше. Тому важливо проводити збір даних на постійній основі та створювати резервні копії ключової інформації.

Одним із суттєвих ризиків при аналізі соціальної активності є потенційна упередженість або штучність реакцій. Коментарі, вподобання та підписки можуть бути результатом неорганічної активності: ботів, «накруток» або платних сервісів просування. Окремі бренди можуть використовувати замовні відгуки чи коментарі з метою створення позитивного іміджу або навпаки — формування негативного враження про конкурентів. Така маніпуляція даними суттєво знижує їхню аналітичну цінність і може вводити в оману, якщо не застосовуються додаткові методи верифікації достовірності. Тому важливо

супроводжувати аналіз якісною експертизою, порівнюючи кілька джерел та оцінюючи автентичність активності.

Збір, зберігання та аналіз даних у цифровому середовищі регулюється законодавчими нормами, які потребують уважного дотримання. Зокрема, у Європейському Союзі діє Загальний регламент про захист даних (GDPR), який обмежує використання особистої інформації без згоди користувача. В Україні подібні принципи закріплені в Законі «Про захист персональних даних». Ці нормативні акти визначають, яка інформація вважається конфіденційною, як вона може бути оброблена, які права мають власники даних. Порухення таких вимог може призвести до юридичної відповідальності, тому при створенні системи збору даних слід обирати лише відкриті, публічно доступні джерела, уникати обробки персональних даних без згоди користувачів і дотримуватись етичних принципів аналітики.

У першому розділі було розглянуто теоретичні засади збору та аналізу даних із соціальних мереж, а також їхню роль у сучасному інформаційному суспільстві. Соціальні платформи трансформувалися з інструментів персонального спілкування у потужні джерела цифрової інформації, що постійно оновлюється та генерується користувачами. Аналіз таких даних дозволяє дослідникам, аналітикам та бізнесу отримувати важливі інсайти щодо настроїв, поведінки, трендів і суспільних процесів.

У процесі дослідження було встановлено, що:

- соціальні мережі є багатим джерелом Big Data, яке охоплює текст, мультимедіа, метадані, поведінкові патерни та мережеві взаємозв'язки;
- аналіз соціальних даних має прикладне значення в маркетингу, політиці, соціології, безпеці, журналістиці, прогнозуванні та штучному інтелекті;
- збирання даних можливе через різні методи: офіційні API, веб-скрапінг, трекінгові інструменти та краудсорсинг — кожен із яких має свої переваги, обмеження та етичні ризики;
- аналітика соціальних мереж базується на міждисциплінарному підході,

поєднуючи технології обробки даних, знання з поведінкових наук та етичні стандарти.

Таким чином, соціальні мережі стали не лише комунікаційними платформами, а й стратегічними ресурсами для аналізу цифрового суспільства. Успішне використання даних із соцмереж вимагає глибокого розуміння як технічних, так і соціокультурних аспектів, а також дотримання принципів конфіденційності та етики.



РОЗДІЛ 2

АНАЛІЗ ВИМОГ ДО СИСТЕМИ ЗБОРУ ТА АНАЛІЗУ ДАНИХ

2.1 Огляд існуючих рішень і підходів

Перед тим як перейти до проєктування власного рішення, варто зрозуміти, що вже існує на ринку — як у сфері комерційного моніторингу соцмереж, так і в академічному дослідженні цифрових слідів. У цьому підрозділі розглядаються основні платформи та технології, які забезпечують збір, зберігання й аналіз даних із соціальних медіа. Аналізуються їхні функціональні можливості, технічні підходи та обмеження. Окрему увагу приділено етичним аспектам і правовим обмеженням у контексті роботи з даними користувачів. Такий огляд дозволяє оцінити релевантність існуючих рішень для задач дослідження та обґрунтувати потребу у створенні власної системи.

На ринку доступні десятки платформ, що надають функції моніторингу, збору, зберігання та обробки соціальних даних. Серед них — Sprout Social, Brandwatch, Hootsuite, Meltwater, Talkwalker, а також дослідницькі інструменти на кшталт NodeXL, Netlytic або Twint. Більшість з них підтримують роботу з популярними платформами, такими як Facebook, X (Twitter), Instagram, TikTok, YouTube. Вони пропонують модулі для аналізу залученості, тональності, трендів і мережевої взаємодії.

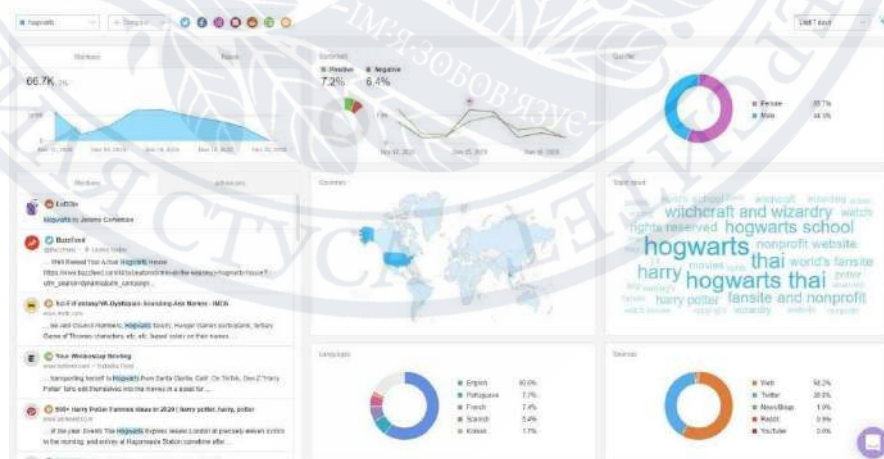


Рисунок 2.1 – Приклад платформи аналізу даних з соціальних мереж

Технічно ці системи можуть використовувати API доступ, веб-скрапінг, інтеграцію з базами даних, хмарне зберігання та візуалізацію у форматі дашбордів. Проте у кожного підходу є свої обмеження — наприклад, API можуть мати обмеження на кількість запитів або не надавати повного доступу до історичних даних.

Окремий акцент слід зробити на етичних і правових аспектах. Збір і аналіз особистих даних потребує дотримання норм конфіденційності, правил користування платформами та міжнародного законодавства, зокрема GDPR. Такі обмеження прямо впливають на можливості автоматичного збору, зберігання і публічного використання отриманих результатів.

На сучасному ринку аналітики соціальних мереж існує чимало готових рішень — як комерційних, так і відкритих інструментів для досліджень. Вони відрізняються рівнем складності, функціональністю, вартістю й орієнтацією на різні типи користувачів — від маркетологів і SMM-команд до науковців і розробників. Розглянемо найбільш відомі приклади з обох категорій.

Серед комерційних систем особливо вирізняється Brandwatch — потужна платформа, яка використовується міжнародними брендами для аналізу репутації, поведінки аудиторії та моніторингу згадок у соціальних мережах і онлайн-медіа. Сервіс дозволяє відстежувати тональність повідомлень, визначати тренди, досліджувати візуальний контент завдяки технологіям розпізнавання зображень. Brandwatch збирає дані з величезної кількості джерел, забезпечуючи користувачів гнучкими інструментами візуалізації та інтеграції з іншими системами, такими як CRM або BI-платформи. Його головна перевага — масштабованість і глибина аналітики, хоча вартість сервісу робить його доступним переважно для великих компаній.

Інший поширений інструмент — Hootsuite, який популярний серед SMM-фахівців. На відміну від Brandwatch, він більше орієнтований на щоденну роботу з контентом: дозволяє планувати публікації, керувати кількома акаунтами одночасно, відповідати на коментарі та повідомлення в єдиному інтерфейсі. Окрім цього, Hootsuite надає базову аналітику щодо залученості, охоплення та

динаміки зростання аудиторії. Завдяки простому інтерфейсу платформа підходить навіть для невеликих компаній або фрілансерів, хоча й не забезпечує глибокого аналізу даних.

Для більш глибокої аналітики варто звернути увагу на Talkwalker — ще одну масштабну платформу, що спеціалізується на аналізі брендів та управлінні репутаційними ризиками. Її сильна сторона — це не лише моніторинг текстового контенту, а й візуальний аналіз: система вміє розпізнавати логотипи, об'єкти, сцени на зображеннях і відстежувати, як і де вони згадуються в соціальних мережах. Крім того, Talkwalker активно використовує штучний інтелект для емоційного аналізу та визначення зміни настроїв у публікаціях. Платформа підтримує багатомовний аналіз, що робить її ефективною в міжнародних дослідженнях.

Окрім комерційних продуктів, існує ціла низка дослідницьких та open-source інструментів, які активно використовуються в науковому середовищі.

Наприклад, NodeXL — це надбудова для Microsoft Excel, що дозволяє аналізувати соціальні зв'язки та будувати графи взаємодії. Вона особливо корисна для дослідження мережевої структури комунікації, виявлення центральних вузлів, інфлюенсерів або інформаційних кластерів. NodeXL ідеально підходить для візуалізації даних з Twitter або YouTube і має порівняно низький поріг входу — для роботи не потрібні знання програмування.

Інший корисний дослідницький інструмент — Netlytic, який працює як онлайн-сервіс і спеціалізується на автоматичному аналізі тексту з соціальних мереж. Він дозволяє дослідникам створювати словникові хмари, графіки поширення тем, виявляти закономірності у комунікації між користувачами. Платформа добре підходить для гуманітарних і соціальних досліджень, хоча має обмеження щодо обсягів даних і підтримуваних платформ.

Для більш технічно підготовлених користувачів існує Python-бібліотека Twint — інструмент для збору публічних даних із X (Twitter) без використання офіційного API. Це особливо цінно, коли API обмежено або доступ до нього заблоковано. Twint дозволяє шукати твіти за ключовими словами, мовою,

геолокацією, а також збирати інформацію про користувачів, хештеги та взаємодії. Проте цей інструмент вимагає базових навичок програмування і не завжди стабільний через часті зміни на самій платформі Twitter.

Таким чином, доступні рішення охоплюють широкий спектр завдань: від щоденного моніторингу до глибокого наукового аналізу. Вибір інструменту залежить від цілей дослідження, обсягів даних, технічних можливостей команди та фінансових ресурсів.

Таблиця 2.1 Порівняння платформ аналізу даних з соціальних мереж

Інструмент	Тип	Призначення	Користувачі	Функціональність	Переваги	Обмеження
Brandwatch	Комерційний	Глибока аналітика соцмереж, репутація бренду	Великі компанії, аналітики	Аналіз тексту та зображень, тренди, тональність, інтеграція з CRM	Потужна аналітика, масштабованість, AI	Висока вартість, складність
Hootsuite	Комерційний	Керування контентом і базова аналітика	SMM-фахівці, малі бізнеси	Публікації, коментарі, охоплення, зростання аудиторії	Проста, підтримка багатьох платформ	Обмежена аналітика, менше гнучкості
Talkwalker	Комерційний	Аналіз бренду, емоцій, зображень	Міжнародні компанії, PR-аналітики	Моніторинг візуального і текстового контенту, багатомовна підтримка, AI	Візуальний аналіз, AI-аналітика	Висока ціна, складне налаштування
NodeXL	Open-source / дослідницький	Аналіз мереж комунікації	Дослідники, соціологи	Графи зв'язків, інфлюенсери, візуалізація	Легка інтеграція з Excel, простота	Лише Windows, обмежена функціональність
Netlytic	Дослідницький / онлайн-сервіс	Текстовий аналіз соцмереж	Гуманітарії, викладачі	Словникові хмар, поширення тем, патерни комунікації	Онлайн-доступ, зручний інтерфейс	Обмеження обсягів даних

У процесі аналізу існуючих рішень також варто звернути увагу на типову архітектуру подібних систем, яка складається з кількох ключових компонентів: модуль збору даних (через API або скрапінг), блок обробки та зберігання (зазвичай на основі баз даних або хмарних сховищ), аналітичний модуль (який може включати ML/AI-алгоритми) та інтерфейс виводу результатів (дашборди, звіти, API). Це дозволяє краще розуміти, які частини слід врахувати при проектуванні власного рішення, особливо якщо йдеться про масштабування, автоматизацію або інтеграцію з іншими системами.

Також важливо відзначити, що більшість готових платформ мають закриту структуру, що обмежує можливості кастомізації або наукової адаптації. Це є особливо критичним для дослідників, які потребують повного контролю над обробкою даних, алгоритмами та логікою аналізу. Відсутність прозорості — один із головних аргументів на користь створення власного рішення, навіть якщо воно буде менш масштабним.

Ще один фактор — обмеження за джерелами даних. Наприклад, більшість комерційних систем не працюють з Viber або обмежені в можливостях збору даних з TikTok. Також важко інтегрувати специфічні локальні джерела, форуми, коментарі на новинних сайтах тощо. Для досліджень, які потребують гнучкої багатоканальної інтеграції, готові платформи часто виявляються недостатніми.

Крім того, слід враховувати питання вартості та доступності: більшість професійних платформ працюють за моделлю підписки з високими цінами, що не завжди виправдано для освітніх або некомерційних проєктів. Саме це також стимулює розробку власного, легшого у використанні й доступного рішення, яке можна адаптувати під конкретні цілі.

Підсумовуючи, огляд існуючих рішень показує, що хоча на ринку вже є потужні інструменти, жоден з них не дає повної гнучкості, прозорості та адаптивності. Це обґрунтовує необхідність розробки власної системи збору та аналізу соціальних даних, яка буде краще відповідати поставленим дослідницьким чи прикладним завданням.

2.2 Визначення вимог до системи та її проектування

Перш ніж розпочати реалізацію системи збору даних із соціальних мереж, надзвичайно важливо було не просто окреслити загальну ідею, а чітко уявити, якою ця система має бути на практиці. Йшлося не лише про набір функцій, а про розуміння цілей, контексту використання, очікувань кінцевого користувача й умов, у яких система повинна працювати. Наприклад, якщо користувачі взаємодіятимуть із нею через Viber, потрібно було врахувати обмеження платформи, простоту повідомлень, затримки в обробці, а також те, як подати результат у зручній формі. Крім того, система мала бути не лише корисною "тут і зараз", а такою, що залишатиметься актуальною й стабільною протягом тривалого часу, навіть якщо зовнішні умови зміняться.

Ці міркування лягли в основу постановки вимог. На цьому етапі стало зрозуміло, що система повинна бути не просто здатною "забирати дані", а повинна сама розуміти, звідки їх брати, як їх структурувати, які поля витягати і як реагувати на помилки. Усі ці речі мали бути продумані наперед, щоб уникнути ситуації, коли вже на етапі реалізації виявляється, що якогось важливого функціоналу бракує або його складно додати без повного переписування коду. У цьому сенсі формування вимог — це не формальність, а інструмент, який допомагає правильно організувати роботу, уникнути надмірної складності та побудувати систему, яка справді працює.

Крім технічних аспектів, увагу також було приділено сценаріям використання. Наприклад, чи працюватиме система на одному комп'ютері чи на сервері, чи планується її використання кількома людьми одночасно, як забезпечити захист даних, і наскільки легко її буде підтримувати. Це потребувало розуміння не лише того, що робить система, а й як виглядає її життєвий цикл — від першого запуску до регулярної експлуатації, оновлень і можливого масштабування. Усе це дало змогу побудувати не просто технічне рішення, а продуману, адаптивну систему, яка відповідає конкретним цілям і може змінюватися разом із ними.

З одного боку, завдання полягало в тому, щоб забезпечити повну функціональність системи. Вона мала працювати як повноцінний інструмент збору даних із Facebook та Instagram, здатний отримувати публікації, коментарі до них, реакції користувачів і додаткову інформацію — зображення, поширення, вкладення тощо. При цьому важливо було забезпечити повну автоматизацію: користувач не повинен вручну вводити запити, копіювати посилання чи

орієнтуватися в структурі API. Усе, що від нього вимагається — це надіслати посилання на сторінку або групу, а далі система самостійно аналізує введене, визначає, з якого саме джерела брати інформацію, формує запит, обробляє отриману відповідь і структурує її у зручному для подальшого використання вигляді.

Особливу увагу довелося приділити адаптації до різних форматів даних, а також тому, як правильно обробляти великі обсяги інформації. Адже у випадку сторінок із високою активністю система може зіткнутися з десятками або навіть сотнями постів, кожен з яких має власну гілку коментарів, низку реакцій, вкладені медіафайли. Це створює додаткове навантаження і на обробку даних, і на швидкість відповіді, тому ще на етапі планування було закладено підтримку пагінації, оптимізацію запитів і обмеження на обсяг, який обробляється за один раз.

Окремий напрям у процесі постановки вимог стосувався аналітики зібраних даних. Було очевидно, що просто витягувати інформацію — недостатньо. Необхідно забезпечити можливість її подальшого аналізу, побудови узагальнень і виявлення закономірностей. Для цього система мала містити окремий модуль обробки та візуалізації, який дозволяє формувати графіки й діаграми на основі зібраних публікацій, реакцій, коментарів та іншої активності користувачів. Це відкривало можливості для швидкої оцінки змін динаміки, порівняння активності по різних постах або визначення пікових періодів залучення.

У підсумку, ключова ідея полягала в тому, щоб система не просто виконувала набір окремих функцій, а забезпечувала повноцінний, логічно завершений цикл взаємодії: від отримання запиту користувача до повернення

готового результату у форматі, з яким можна одразу працювати — без потреби в додатковому аналізі, ручному сортуванні чи доопрацюванні.

Усі попередні міркування й аналіз лягли в основу етапу постановки вимог до системи. Саме тут стало очевидно, що недостатньо просто реалізувати базовий механізм отримання даних. Система мала бути «розумною» в широкому сенсі — вона повинна вміти орієнтуватися в різних умовах: визначати, з якої саме частини платформи брати дані, розуміти структуру отриманої інформації, вибирати лише ті поля, які дійсно мають аналітичну цінність, і правильно реагувати на помилки. Наприклад, якщо сторінка недоступна або запит не дав результату — система повинна не зависати, а повернути зрозуміле повідомлення, зафіксувати проблему й далі працювати стабільно. Такий підхід дозволив уникнути типових помилок, коли вже під час реалізації доводиться переробляти великі частини коду лише тому, що деякі речі не були передбачені на старті.

Формулювання вимог розглядалося не як формальна частина проєкту, а як практичний інструмент для побудови стійкої логіки системи. Воно дало змогу заздалегідь визначити, які компоненти мають бути незалежними, які можна буде легко замінити або оновити, а також як повинна виглядати структура взаємодії між окремими модулями. Це дозволило не лише розвантажити розробку, а й створити основу для майбутнього розвитку — коли виникне потреба підключити інші платформи, додати аналітику чи вивантаження в базу даних.

Паралельно з технічними питаннями увага приділялася й практичному використанню системи. Було важливо розуміти, як і де саме вона працюватиме: чи це буде запуск на локальному комп'ютері, чи на сервері; чи використовуватиметься лише одним адміністратором, чи кількома людьми одночасно; чи буде доступ через мобільний пристрій. Усі ці сценарії прямо впливали на вибір інтерфейсу (Viber-бот), формат повернення результатів (JSON-файли або стислий звіт) і навіть на те, якою мовою писати систему, щоб її легко було підтримувати й адаптувати у майбутньому. У результаті вдалося створити не просто набір функцій, а цілісну систему, яка логічно працює, добре масштабується і залишається гнучкою до змін — як у технічному, так і в користувацькому сенсі.

У контексті розробки програмного забезпечення архітектура системи — це концептуальна модель, що описує структуру, компоненти та принципи взаємодії між ними. Архітектура визначає, з яких частин складається система, як ці частини організовані, які функції виконує кожна з них, і яким чином вони пов'язані між собою. Вона відіграє роль "скелету" проєкту, на який накладаються окремі модулі, технології та логіка виконання, забезпечуючи узгоджене функціонування всієї програми.

Архітектура системи збору та аналізу соціальних даних побудована замодульним принципом, що дозволяє легко масштабувати, оновлювати та адаптувати систему до нових джерел або змін у зовнішньому середовищі. Уся логіка роботи розподілена між п'ятьма ключовими компонентами, які пов'язані між собою у чітко визначеній послідовності (див. рис.).

1. Viber-бот — точка входу в систему, яка забезпечує взаємодію з користувачем. Бот отримує від користувача посилання на публічну сторінку чи профіль у соціальній мережі (наприклад, Instagram чи Facebook), а також надає зворотний зв'язок у вигляді текстових повідомлень, звітів або графіків. Він працює в асинхронному режимі, що дозволяє обробляти запити без затримок навіть при високому навантаженні.
2. Request-модуль — посередник, який приймає вхідні дані від бота, перевіряє їхню валідність і передає у наступний блок. У разі помилки або некоректного формату, бот повертає користувачеві відповідне повідомлення про помилку.
3. Web Scraping — модуль збору даних із соціальних платформ. За допомогою бібліотек BeautifulSoup, Selenium або Playwright відбувається автоматизоване витягування HTML-даних із відкритих сторінок. Система враховує тип вмісту (статичний або динамічний), ругує User-Agent заголовки, може використовувати проксі-сервери для уникнення блокувань, а також логуватиме помилки або зміни структури сторінки. На цьому ж етапі зібрані елементи (пости, коментарі, дати, лайки, зображення тощо) первинно структуруються.

4. Аналітичний модуль — обробляє зібрані дані, нормалізує формати, фільтрує дублі, агрегує ключові метрики (кількість реакцій, коментарів, динаміка активності). Основою аналітики виступає бібліотека Pandas, яка дозволяє легко формувати таблиці, зведення, групувати записи за часом чи типом взаємодії. Також модуль готує дані до візуалізації: обирає релевантні поля, будує масиви для графіків.
5. Модуль візуалізації (Processing & Visualization) — формує фінальний результат у вигляді графіків, діаграм або коротких текстових зведень. Готовий результат повертається назад до бота й надсилається користувачеві. У випадках, коли дані не вміщуються в одне повідомлення, реалізоване автоматичне розбиття результату на кілька частин.

Ця архітектура дозволяє ефективно реалізувати повний цикл аналізу: від введення посилання до отримання інформативного звіту. Завдяки модульності система легко адаптується до нових джерел (наприклад, TikTok чи X), а також підтримує розширення функціоналу — наприклад, додавання авторизації, збереження історії запитів, експорт у PDF або інтеграцію з базами даних.



Рисунок 2.1 – Архітектура системи збору та аналізу соціальних даних

2.3 Вибір технологій для реалізації системи

Розробка системи збору та аналізу даних із соціальних мереж вимагала ретельного й виваженого підходу до вибору технологій. Йшлося не просто про те, щоб знайти інструменти, які виконують базові функції — наприклад, надсилають HTTP-запити або зберігають інформацію. Важливо було обрати такі рішення, які дозволять не лише ефективно виконувати поточні завдання, а й дадуть змогу масштабувати систему у разі зростання обсягів даних або потреби в обробці нових джерел. Враховувалася також здатність адаптуватися до змін у зовнішньому середовищі — зокрема, до частих оновлень соціальних мереж, змін у політиці приватності, структурі відповідей або обмеженнях доступу.

Ще на етапі планування було очевидно, що система повинна бути достатньо стійкою до змін — особливо враховуючи те, наскільки динамічними є великі платформи. Якщо рішення буде «жорстко прив'язане» до поточної реалізації чи певного формату даних, будь-яке навіть незначне оновлення на стороні платформи може паралізувати її роботу. Тому основною вимогою стала гнучкість: система мала не тільки працювати «тут і зараз», а й легко перебудовуватись під нові умови — з мінімальними зусиллями з боку розробника. Це також включало простоту супроводу, можливість швидко вносити зміни в код, замінювати окремі модулі або додавати нові.

Окремим критерієм було питання інтеграції. Із самого початку було зрозуміло, що система повинна мати зручний і швидкий спосіб взаємодії з користувачем. У нашому випадку це — Viber-бот, який виступає простим інтерфейсом для запуску запитів та отримання результатів. Тож усі обрані технології мали бути сумісними не лише між собою, а й із зовнішніми сервісами.

У процесі вибору мови програмування, на якій будуватиметься уся система, було розглянуто кілька варіантів, однак рішення на користь Python виявилось найбільш обґрунтованим. Python поєднує в собі зручний, зрозумілий синтаксис і широку функціональність, що робить його ідеальним інструментом як для швидкої розробки прототипу, так і для створення повноцінної, масштабованої системи. Його популярність у сфері обробки даних, веб-

=скрапінгу, автоматизації запитів і машинного навчання створила навколо мови велику екосистему готових рішень, документації та активної спільноти.

Особливу увагу було приділено бібліотекам, які підтримують роботу з HTTP-запитами, JSON-даними та HTML-документами. Зокрема, requests — одна з найзручніших і найстабільніших бібліотек для роботи з API, дозволяє швидко налаштовувати запити, передавати параметри, обробляти помилки та отримувати структуровані відповіді. Для обробки тексту, парсингу HTML та роботи з регулярними виразами також є широкий вибір інструментів — таких як BeautifulSoup, re, json, що дозволяє гнучко працювати з контентом будь-якої складності.

Ще однією причиною вибору Python стала його сумісність із ViberAPI. Це дає змогу легко реалізувати бот, який виступає інтерфейсом між користувачем і системою збору даних. Бот, реалізований на Python (наприклад, за допомогою бібліотеки python-viber-bot), не потребує складного розгортання, миттєво реагує на запити користувача й дозволяє взаємодіяти з системою у звичному месенджері без зайвих бар'єрів. Такий підхід забезпечує простоту використання для кінцевого користувача, а для розробника — гнучкість і швидкість реалізації.

Таблиця 2.2 Порівняння мов програмування для розробки систем збору соціальних даних

Критерій	Python	JavaScript (Node.js)	Java
Простота синтаксису	Висока – ідеально для швидкого старту	Висока, але менш зручна для обробки даних	Середня – потребує більше коду
Екосистема для збору даних	Розвинена (requests, BeautifulSoup, Twint)	Середня (axios, puppeteer)	Середня (HttpClient, JSoup)
Можливості візуалізації	Хороші (matplotlib, seaborn, Plotly)	Середні (D3.js для вебу)	Слабкі без сторонніх бібліотек
Обробка великих обсягів даних	Добра, з підтримкою багатьох бібліотек	Добра для асинхронної обробки	Висока ефективність, але складніше

Попри функціональність і гнучкість обраного рішення, важливо враховувати й ризики, пов'язані з обхідними методами авторизації. У поточній реалізації система отримує доступ до даних через cookie-файл активної сесії користувача, з якої витягується токен доступу шляхом парсингу HTML-сторінки Facebook. Такий підхід дозволяє обійти складну процедуру авторизації через офіційний механізм OAuth, однак він залишається нестабільним за своєю природою. Соціальні платформи часто оновлюють структуру своїх сторінок, змінюють назви елементів, політику безпеки або вводять додаткові механізми захисту. У результаті навіть незначні технічні зміни можуть призвести до того, що система перестане працювати або буде заблокована.

Ще один критичний аспект — це юридичні та етичні межі. Оскільки система працює з персональними даними, навіть якщо мова йде про публічні пости чи коментарі, важливо дотримуватись чинного законодавства у сфері захисту приватності. У першу чергу — це Загальний регламент захисту даних ЄС (GDPR), який встановлює чіткі вимоги до обробки, зберігання й передачі особистої інформації. Порушення цих правил може мати серйозні наслідки як на рівні відповідальності, так і з точки зору репутації розробника або організації, що використовує систему.

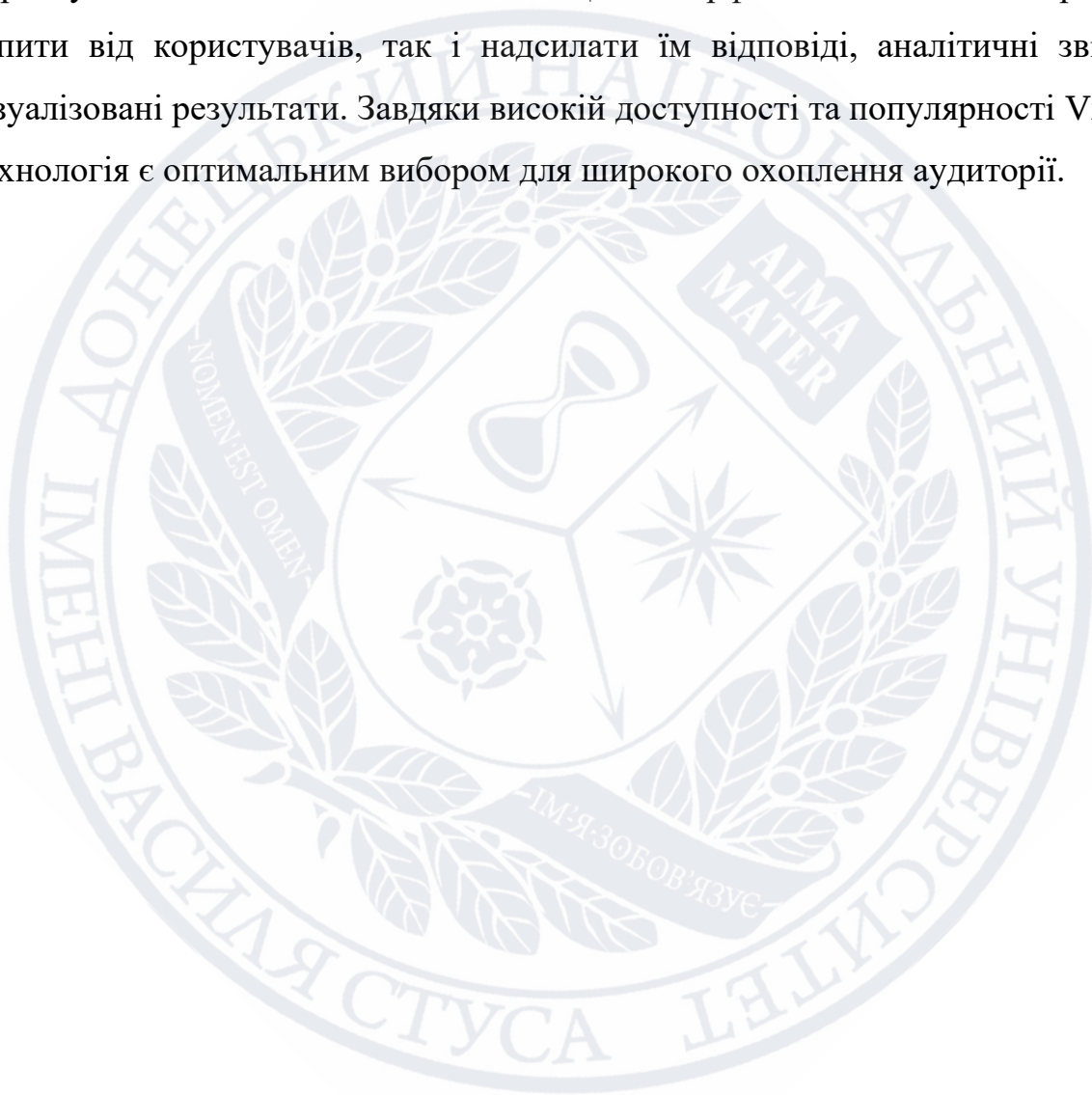
З урахуванням цих факторів, систему свідомо обмежено у сфері її застосування — вона працює виключно з публічно доступною інформацією, яка не вимагає додаткової авторизації чи отримання згоди користувачів. Уся взаємодія з платформою відбувається в межах того, що офіційно відкрито для загального перегляду. Такий підхід дозволяє залишатися в правовому полі, мінімізує ризики блокування або правових претензій і водночас дає змогу ефективно збирати й аналізувати великі масиви соціальних даних.

Загалом, обрані технології — Python, Graph API, Viber API — забезпечують надійну основу для реалізації функціональної, гнучкої та адаптивної системи збору соціальних даних. Кожна з цих складових відіграє ключову роль у забезпеченні стабільності, розширюваності та інтеграційних

можливостей системи, що є особливо важливим у контексті сучасної цифрової трансформації.

Python як основна мова програмування забезпечує широкий набір бібліотек та інструментів для обробки даних, автоматизації процесів і побудови логіки взаємодії з зовнішніми сервісами. Завдяки своїй гнучкості та зрозумілому синтаксису Python дозволяє швидко реалізовувати нові функціональні модулі, адаптуючи систему до нових завдань або змін у вимогах.

Viber API забезпечує зручний канал для взаємодії з кінцевими користувачами за допомогою ботів. Цей інтерфейс дозволяє як отримувати запити від користувачів, так і надсилати їм відповіді, аналітичні звіти або візуалізовані результати. Завдяки високій доступності та популярності Viber, ця технологія є оптимальним вибором для широкого охоплення аудиторії.



РОЗДІЛ 3

ПРОЕКТУВАННЯ ТА РЕАЛІЗАЦІЯ СИСТЕМИ ЗБОРУ ТА АНАЛІЗУ ДАНИХ

3.1. Загальна логіка роботи системи

Функціонування системи збору та аналізу соціальних даних базується на принципі інтерактивної взаємодії користувача з Viber-ботом — спеціально розробленим інтерфейсом, що виконує роль зв'язувальної ланки між користувачем і ядром системи. Бот реалізує логіку "запит–відповідь", забезпечуючи зручність, оперативність та доступність використання на будь-якому пристрої.

Першим етапом функціонування системи є ініціалізація сесії взаємодії з користувачем через інтерфейс бота. Процес розпочинається з активації бота шляхом надсилання команди /start, що ініціює сценарій стартової комунікації між користувачем та програмною системою.

Після того як користувач вводить команду /start, бот не лише надсилає вітальне повідомлення, а й одразу переходить до наступного етапу — запитує у користувача посилання на сторінку або групу конкурента у Facebook. Це запит формулюється максимально просто і зрозуміло: без технічної термінології, із прямим формулюванням завдання. Такий підхід мінімізує плутанину, дає змогу користувачу швидко зрозуміти, що саме від нього очікується, і не потребує жодних попередніх налаштувань.

Другим етапом є надсилання користувачем запиту у вигляді публічного посилання на об'єкт для аналізу — сторінку, групу або окремий пост у соціальних мережах Facebook або Instagram. Цей запит слугує точкою входу в механізм збору даних і є ключовим тригером для запуску внутрішньої логіки обробки.

Viber-бот автоматично фіксує отримане повідомлення, здійснює його попередню верифікацію, зокрема:

- перевіряє, чи є в повідомленні валідне гіперпосилання;
- визначає тип соціальної платформи;
- зчитує структуру посилання для подальшого розбору.

У разі, якщо посилання є некоректним (наприклад, веде на приватний

профіль, має порушену структуру або не підтримується системою), бот формує відповідь з уточненням і проханням надіслати правильне публічне посилання. Це дозволяє запобігти обробці недійсних або не підтримуваних об'єктів, підвищуючи надійність системи.

Якщо запит проходить перевірку успішно, дані передаються до відповідного модуля обробки (залежно від типу запиту), і користувач отримує повідомлення про старт аналізу. Такий підхід забезпечує чітку, контрольовану логіку взаємодії, з мінімальним ризиком помилок на етапі введення даних.

На етапі обробки запиту система автоматично ідентифікує, до якої саме соціальної платформи — Facebook чи Instagram — належить отримане посилання, після чого передає його до відповідного модуля збору даних. У залежності від типу джерела для витягування інформації використовується або офіційний API, або метод веб-скрейпінгу. Після встановлення з'єднання з ресурсом, система або завантажує HTML-структуру сторінки, або отримує структуровану відповідь у форматі JSON. На основі отриманої інформації зчитуються ключові характеристики облікового запису, зокрема назва сторінки, кількість підписників, опис профілю та його тематика.

Після цього система переходить до збору останніх публікацій, разом із супровідними даними: кількістю реакцій, коментарів, поширень, а також датами публікацій. Усі отримані дані проходять первинну очистку та структурування. Далі бот переходить до етапу попереднього аналізу, в межах якого здійснюється автоматичний збір базової публічної інформації з відповідної соціальної мережі — Instagram або Facebook. Метою цього етапу є формування короткого зведеного звіту, який надається користувачеві практично миттєво. Такий підхід виконує одразу кілька функцій: підтверджує працездатність системи, дозволяє переконатися в коректності посилання та релевантності обраної сторінки, а також надає початкове уявлення про активність і зміст профілю ще до глибшого аналізу.

Після завершення попереднього аналізу користувач отримує змогу самостійно обрати напрям подальшого дослідження, натиснувши одну з

інтерактивних кнопок: «Аналіз контенту», «Взаємодія з аудиторією», «Вивчення хештегів» або «Аналіз активності». Такий підхід забезпечує гнучкість і дозволяє зосередитися саме на тій інформації, яка є найбільш актуальною для досягнення конкретної цілі — будь то аналіз стилю комунікації, ефективності контенту чи поведінки підписників.

Функція «Аналіз контенту» надає розгорнутий огляд візуального й текстового оформлення сторінки конкурента. Бот аналізує останні 10 публікацій, оцінюючи стиль фотографій, тематику постів, характер підписів і рівень взаємодії з контентом. Результати подаються у структурованому форматі з короткими аналітичними висновками й практичними рекомендаціями щодо можливого застосування вдалих прикладів у власній стратегії.

Функція «Взаємодія з аудиторією» дозволяє оцінити, як ефективно акаунт будує комунікацію зі своїми підписниками. Бот аналізує активність у коментарях, використання інтерактивних елементів у сторіс, заклики до дії в постах та участь у конкурсах. На основі цього формується блок рекомендацій, які допоможуть підвищити залучення у власному акаунті та зробити взаємодію більш живою й персоналізованою.

Функція «Вивчення хештегів» забезпечує глибинний аналіз тегів, які активно використовуються конкурентом. Viber-бот визначає найпоширеніші хештеги, обчислює середню кількість реакцій на пости з ними та оцінює їхню популярність і ефективність. На основі цього користувач отримує список найрезультативніших тегів разом із порадами щодо їхнього потенційного використання у власному контенті для збільшення охоплення.

Функція «Аналіз активності» дозволяє дослідити поведінку підписників на сторінці конкурента протягом останніх 14 днів. Бот збирає дані про публікації, лайки, коментарі, найактивніші дні та найпопулярніші формати контенту. Результати виводяться у вигляді графіків і таблиць, а також супроводжуються рекомендаціями щодо оптимального часу для публікацій і пріоритетних типів контенту, що дозволяє ефективно адаптувати власну стратегію просування.



Рисунок 3.1 – Діаграма логіки роботи Viber -бота

3.1 Розробка модуля збору даних

Модуль збору даних є базовим компонентом системи, оскільки саме він забезпечує витяг первинної інформації зі сторінок соціальних мереж. У цьому проєкті акцент зроблено на веб-скрейпінгу як основному методі отримання даних, оскільки API-підхід має суттєві обмеження — зокрема, потребу авторизації, обмеження за швидкістю запитів і урізаний доступ до публічного контенту. Скрейпінг дозволяє отримувати дані напряму з HTML-коду сторінок, що відкриває більше можливостей, але водночас вимагає дотримання обережності, врахування змін у верстці, можливого використання JavaScript, а також засобів захисту від ботів.

Модуль реалізовано на Python з використанням бібліотек requests, BeautifulSoup, Selenium (для динамічних сторінок) та lxml. Логіка роботи: завантаження сторінки → витяг елементів → нормалізація → збереження (наприклад, у JSON). Передбачено окремі парсери для різних типів сторінок (стрічка новин, пости, коментарі), які адаптуються до змін DOM-структури.

Задіяні механізми обходу захисту: ротація User-Agent, затримки між запитамі, проксі-сервери. При виявленні CAPTCHA або JavaScript-завантаження бот автоматично використовує Selenium для імітації дій користувача.

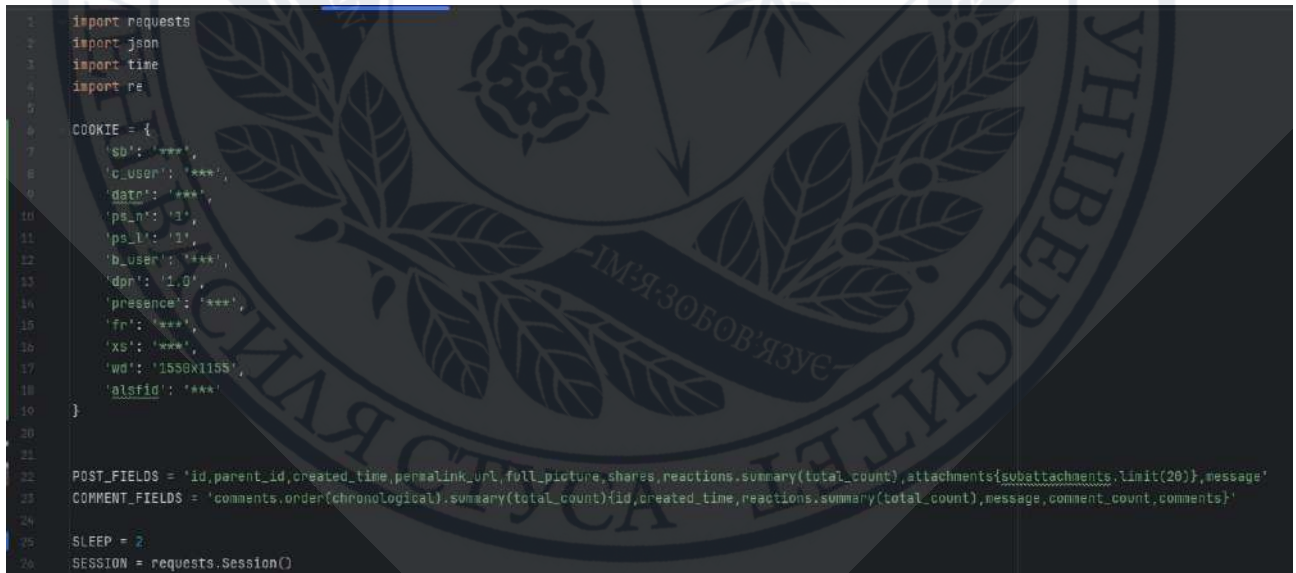
Модуль витягує текст, авторів, реакції, коментарі, дату, вкладення тощо. Дані очищуються від HTML-тегів, нормалізуються (час, емодзі), що спрощує подальший аналіз.

Для Facebook передбачено два режими:

- Graph API — для структурованих запитів;
- Авторизація через cookie — як альтернатива OAuth 2.0, що дозволяє взаємодіяти як реальний користувач без додаткової інтеграції.

Сесія підтримується через `requests.Session()`, що зберігає cookies і заголовки. Визначаються ключові параметри запитів (`POST_FIELDS`, `COMMENT_FIELDS`) та встановлюється інтервал `SLEEP = 2` для зменшення навантаження і запобігання блокуванню.

Модуль протестовано на різних сторінках (від малих груп до активних профілів). Забезпечено стабільність, обробку помилок та масштабованість при великій кількості запитів.



```

1 import requests
2 import json
3 import time
4 import re
5
6 COOKIE = {
7     'sb': '***',
8     'c_user': '***',
9     'data': '***',
10    'ps_n': '1',
11    'ps_l': '1',
12    'b_user': '***',
13    'dpr': '1.0',
14    'presence': '***',
15    'fn': '***',
16    'xs': '***',
17    'wd': '1558x1155',
18    'alsfid': '***'
19 }
20
21 POST_FIELDS = 'id,parent_id,created_time,permalink,url,full_picture,shares,reactions.summary(total_count),attachments{subattachments.limit(20)},message'
22 COMMENT_FIELDS = 'comments.order(chronological).summary(total_count){id,created_time,reactions.summary(total_count),message,comment_count,comments}'
23
24 SLEEP = 2
25 SESSION = requests.Session()
  
```

Рисунок 3.2 – Ініціалізація сесії, авторизаційних cookies і полів запиту до Graph API

Наступним етапом реалізовано функцію `get_node_id()`, яка визначає унікальний ідентифікатор сторінки або групи Facebook на основі введеного користувачем посилання. Цей ID є необхідним для формування запитів до Graph

API, оскільки саме він використовується як ключ для отримання пов'язаної інформації — публікацій, коментарів, реакцій тощо.

Функція спочатку намагається отримати ID напряму, виділяючи останні частини URL за допомогою `.split('/')[ -2:]`. Для груп, наприклад у посиланні `https://www.facebook.com/groups/1234567890`, числовий компонент визначається як ID. Якщо ж URL вказує на сторінку з символьною назвою (наприклад, `/examplepage`), вона також підходить для використання з Graph API.

У разі, якщо ID неможливо отримати безпосередньо з URL, використовується регулярний вираз `r'(?<=/groups/)(\d+)'` для пошуку ID у тексті посилання. Якщо і це не дає результату, функція надсилає GET-запит на відповідну сторінку і виконує пошук ID у HTML-кодї відповіді. Такий підхід дозволяє обробляти посилання з параметрами або нестандартною структурою.

Функція враховує різні формати посилань, адаптуючись до реальних сценаріїв, коли URL може бути скороченим, містити додаткові параметри або вводиться вручну (наприклад, через Viber-бота). У разі невдачі `get_node_id()` повертає `None`, що дозволяє основній логіці парсера зупинити процес і повідомити про помилку. Така реалізація забезпечує стабільність і універсальність у роботі з різними джерелами.

```

27 def get_node_id(competitor_link): 1 usage: Quan Dang
28     node_type, node_name = competitor_link.split('/')[ -2:]
29     if node_type != 'groups':
30         return node_name
31
32     id_in_url = re.search( pattern: r'(?<=/groups/)(\d+)'($|(?=\/)|(?=8))', competitor_link)
33     if id_in_url and id_in_url.group(1):
34         return id_in_url.group(1)
35
36     response = SESSION.get(competitor_link)
37     search_group_id = re.search( pattern: '(?<=/group/?id=)(\d+)', response.text)
38     if search_group_id and search_group_id.group(1):
39         return search_group_id.group(1)
40
41     return None

```

Рисунок 3.3 – Функція отримання ID сторінки або групи з URL-адреси

Третій функціональний блок відповідає за отримання access token — ключа, що дозволяє взаємодіяти з Facebook Graph API. Замість повної авторизації через OAuth 2.0, у розробленому модулі використовується спрощений підхід: токен отримується на основі наявних cookies авторизованого користувача.

Функція формує заголовок запиту з cookies і надсилає HTTP-запит до внутрішньої сторінки Facebook, де у HTML-відповіді міститься токен доступу. За допомогою регулярного виразу токен витягується з відповіді. Якщо він знайдений — зберігається для подальшого використання; інакше — повертається None.

Цей підхід дозволяє обійти складну процедуру реєстрації додатку та OAuth-авторизації, проте має ризики: залежність від структури сторінки та можливі зміни інтерфейсу Facebook. Незважаючи на це, метод ефективний для внутрішніх інструментів і прототипів, що працюють із публічною інформацією.

```

44 def get_access_token():
45     cookie_string = ''.join([f'{key}={value}' for key, value in COOKIE.items()])
46     response = SESSION.get(url='https://business.facebook.com/business_locations', headers={'cookie': cookie_string})
47
48     if response.status_code == 200:
49         search_token = re.search(pattern='(EAAG\\w+)', response.text)
50         if search_token and search_token.group(1):
51             return search_token.group(1)
52
53     return None

```

Рисунок 3.4 – Витяг access token із HTML-сторінки через парсинг

Наступний функціональний блок відповідає за формування запиту до Facebook Graph API та обробку пагінації — поетапного отримання великих обсягів даних.

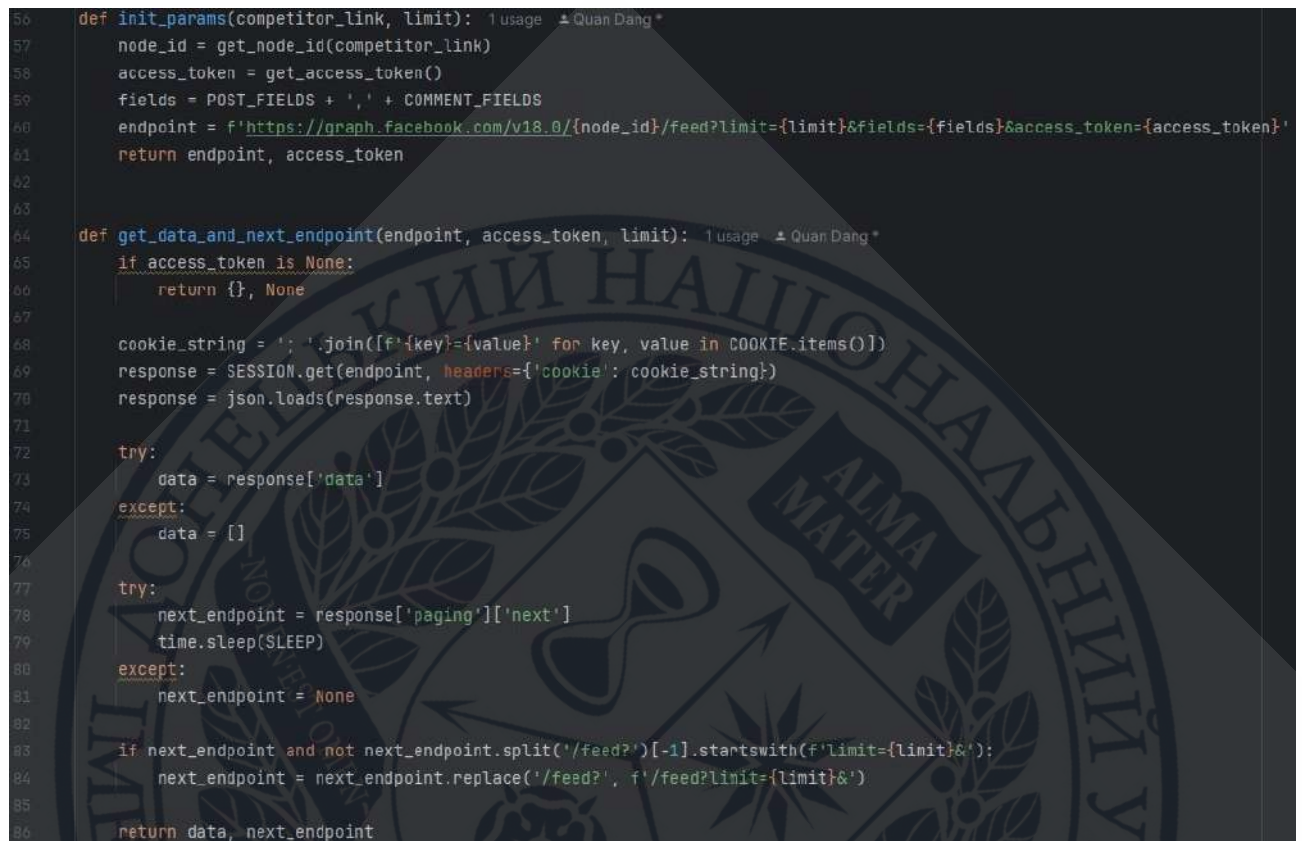
На першому етапі створюється запит, який містить: ID сторінки/групи, access token та перелік потрібних полів (ідентифікатор поста, дата, текст, реакції, поширення, вкладення, коментарі тощо). Це дозволяє одразу отримати повні дані без додаткових запитів.

Оскільки Facebook обмежує обсяг відповіді, система обробляє пагінацію: якщо відповідь містить посилання на наступну сторінку, виконується новий запит. Цикл триває, доки не буде зібрано всі дані або досягнуто заданого ліміту.

Передбачена затримка між запитами (SLEEP) для запобігання

блокуванням з боку Facebook.

У разі помилок (відсутні поля, неправильний токен або ID) модуль продовжує роботу, повертаючи порожні результати без повної зупинки процесу.



```

56 def init_params(competitor_link, limit): !usage - Quan Dang *
57     node_id = get_node_id(competitor_link)
58     access_token = get_access_token()
59     fields = POST_FIELDS + ',' + COMMENT_FIELDS
60     endpoint = f'https://graph.facebook.com/v18.0/{node_id}/feed?limit={limit}&fields={fields}&access_token={access_token}'
61     return endpoint, access_token
62
63
64 def get_data_and_next_endpoint(endpoint, access_token, limit): !usage - Quan Dang *
65     if access_token is None:
66         return {}, None
67
68     cookie_string = '; '.join([f'{key}={value}' for key, value in COOKIE.items()])
69     response = SESSION.get(endpoint, headers={'cookie': cookie_string})
70     response = json.loads(response.text)
71
72     try:
73         data = response['data']
74     except:
75         data = []
76
77     try:
78         next_endpoint = response['paging']['next']
79         time.sleep(SLEEP)
80     except:
81         next_endpoint = None
82
83     if next_endpoint and not next_endpoint.split('/feed?')[-1].startswith(f'limit={limit}&'):
84         next_endpoint = next_endpoint.replace('/feed?', f'/feed?limit={limit}&')
85
86     return data, next_endpoint

```

Рисунок 3.5 – Формування API-запиту та обробка відповіді з підтримкою пагінації

П'ятий блок коду відповідає за обробку окремого поста після того, як він був отриманий у відповідь на запит до Facebook Graph API. Основне його завдання — витягнути з великої кількості сирих даних тільки те, що справді потрібно для подальшого аналізу. Цей підхід не просто полегшує роботу з даними, а й забезпечує стабільність та однорідність структури виводу.

Кожен пост, який надходить у форматі JSON, містить десятки вкладених полів, однак не всі з них мають значення з точки зору дослідження або аналітики. Тому функція проходиться лише по ключових елементах: ідентифікатор публікації, дата її створення, текст повідомлення, кількість реакцій та поширень, а також пряме посилання на публікацію. Саме ці параметри дозволяють оцінити активність, популярність і характер контенту.

Окремий акцент зроблено на роботі із зображеннями. Facebook часто зберігає медіафайли у вкладених структурах, які можуть бути не тільки в основному полі поста, але й у вкладеннях або навіть у підвкладеннях.

Функція також враховує, що деякі пости можуть бути "неповними" — наприклад, без реакцій або з порожнім повідомленням. У таких випадках застосовується обережна логіка обробки: система перевіряє, чи взагалі існує те чи інше поле, і тільки тоді намагається його зчитати. Це дозволяє уникнути помилок при обробці та гарантує стабільну роботу навіть при нестандартних або "мінімалістичних" публікаціях.

У результаті кожен пост перетворюється на компактний словник зі стандартною структурою, який легко зберігати, аналізувати або надсилати через інтерфейс. Такий формат зручно передавати до Viber-бота, виводити на екран чи додавати до звітів.

```

97 def process_post(post): 1usage: new *
98     images = []
99     shares = post.get('shares', {}).get('count', 0)
100     reactions_count = post.get('reactions', {}).get('summary', {}).get('total_count', 0)
101
102     if 'attachments' in post:
103         attachments = post['attachments']['data']
104         for attachment in attachments:
105             if 'media' in attachment and 'image' in attachment['media']:
106                 images.append(attachment['media']['image']['src'])
107                 subattachments = attachment.get('subattachments', {}).get('data', [])
108                 for subattachment in subattachments:
109                     if 'media' in subattachment and 'image' in subattachment['media']:
110                         images.append(subattachment['media']['image']['src'])
111
112     if 'full_picture' in post and post['full_picture']:
113         images.append(post['full_picture'])
114
115     post_data = {
116         'id': post['id'],
117         'created_time': post['created_time'],
118         'message': post.get('message', ''),
119         'permalink_url': post['permalink_url'],
120         'images': images,
121         'shares': shares,
122         'reactions': reactions_count
123     }
124
125     return post_data

```

Рисунок 3.6 – Обробка одного поста: виділення ключових параметрів

Фінальним елементом модуля збору даних є головна функція parse(), яка виконує роль координаційного центру всієї системи. Вона послідовно викликає

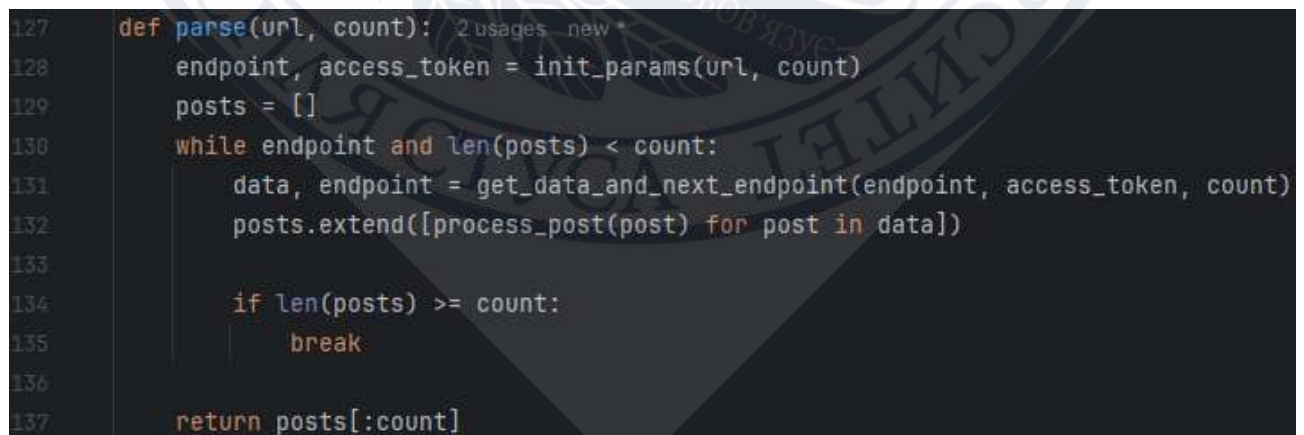
всі допоміжні функції, об'єднуючи їх у єдиний логічний процес: від отримання ID ресурсу до повернення готового набору публікацій.

На початку виконання `parse()` ініціалізує ключові компоненти: визначає унікальний ідентифікатор (ID) сторінки або групи, отримує access token, формує API-запит із зазначеним набором полів. Далі активується цикл, у якому функція: надсилає запит до Facebook Graph API; зчитує «сторінку» результатів (публікації); передає кожен пост на обробку до функції `process_post()`; зберігає результат у фінальний список даних.

Особлива увага приділена пагінації: якщо у відповіді присутнє посилання на наступну порцію даних, функція автоматично продовжує завантаження. Цикл триває до досягнення заданої кількості публікацій або до завершення доступних даних. Для запобігання перевищенню ліміту запитів реалізовано паузу між зверненнями до API (2 секунди), що знижує ризик блокування зі сторони Facebook.

Також реалізовано контроль за максимальним числом постів, яке дозволено зібрати — це дозволяє використовувати функцію як для тестового збору (на кілька записів), так і для повноцінного завантаження великих вибірок.

Результатом роботи `parse()` є стандартизований список словників, кожен із яких містить: текст публікації, дату створення, пряме посилання, кількість реакцій і поширень, зображення (за наявності).



```

127 def parse(url, count):
128     endpoint, access_token = init_params(url, count)
129     posts = []
130     while endpoint and len(posts) < count:
131         data, endpoint = get_data_and_next_endpoint(endpoint, access_token, count)
132         posts.extend([process_post(post) for post in data])
133
134     if len(posts) >= count:
135         break
136
137     return posts[:count]

```

Рисунок 3.7 – Основна функція збору постів `parse()`

3.2 Інтерфейс та демонстрація роботи Viber-бота

У цьому розділі розглядається практична демонстрація функціонування розробленого Viber-бота, який здійснює аналіз сторінок конкурентів у соціальних мережах (Instagram, Facebook). Користувач взаємодіє з ботом у режимі реального часу, надсилаючи посилання на сторінки конкурентів, після чого бот автоматично виконує аналіз контенту, хештегів, активності аудиторії та залучення користувачів.



Рисунок 3.8 – Аватар Viber-бота InsightBot

На ілюстрації (рис 3.8) зображено аватарку бота, яка використовується в інтерфейсі системи. Образ робота з лупою символізує основне призначення бота — допомагати користувачам у зборі та аналізі даних із соціальних мереж. Такий візуальний стиль створює впізнаваність і дружню атмосферу взаємодії з ботом.

Далі поетапно розглянемо усі екрани, що демонструють взаємодію користувача з ботом та результати аналізу сторінок у соціальних мережах.

Крок 1. Надсилання посилання на сторінку конкурента

Користувач надсилає боту посилання на сторінку конкурента у Facebook або Instagram. Бот перевіряє коректність посилання та переходить до етапу збору та обробки інформації.

На цьому етапі важливо, що бот одразу переходить до дії й не залишає користувача у підвішеному стані. Повідомлення побудоване так, що спонукає до наступного кроку — надсилання посилання. Це не просто зручність, а частина

загальної логіки взаємодії, яка має бути швидкою, прямолінійною та зрозумілою.

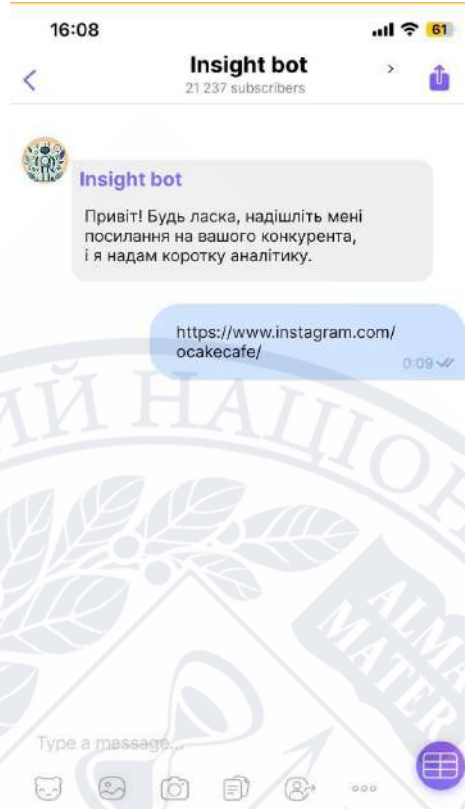


Рисунок 3.9 – Очікування посилання від користувача

Крок 2. Аналіз сторінки: загальна інформація

Для демонстрації можливостей системи аналізу було обрано сторінку мережі кав'ярень O'CAKE у соціальній мережі Instagram — [@ocakecafe](https://www.instagram.com/ocakecafe/). Цей акаунт є офіційним представництвом бренду в Instagram і активно використовується для взаємодії з аудиторією. Він регулярно оновлюється, має сталу візуальну айдентику та містить типові приклади контенту, що характерні для закладів у сфері харчування: фото десертів, напоїв, інтер'єру, сезонних акцій і новин.

Бот здійснює попередній аналіз, отримуючи загальні дані про сторінку, зокрема ім'я профілю, кількість підписників, загальну активність та інші базові метрики. Це дозволяє оцінити загальний масштаб присутності конкурента в соціальній мережі.

У наведеному прикладі (рис. 3.10) видно типовий результат такого попереднього аналізу: бот дякує за посилання, підтверджує його обробку й надає

основну інформацію про обліковий запис. У відповідь користувач отримує такі дані, як нікнейм сторінки, кількість підписників, кількість підписок, загальна кількість публікацій, опис профілю та навіть години роботи закладу (якщо ця інформація вказана в профілі). Ці дані дають змогу швидко оцінити масштаб сторінки, тип контенту та її потенційну активність.

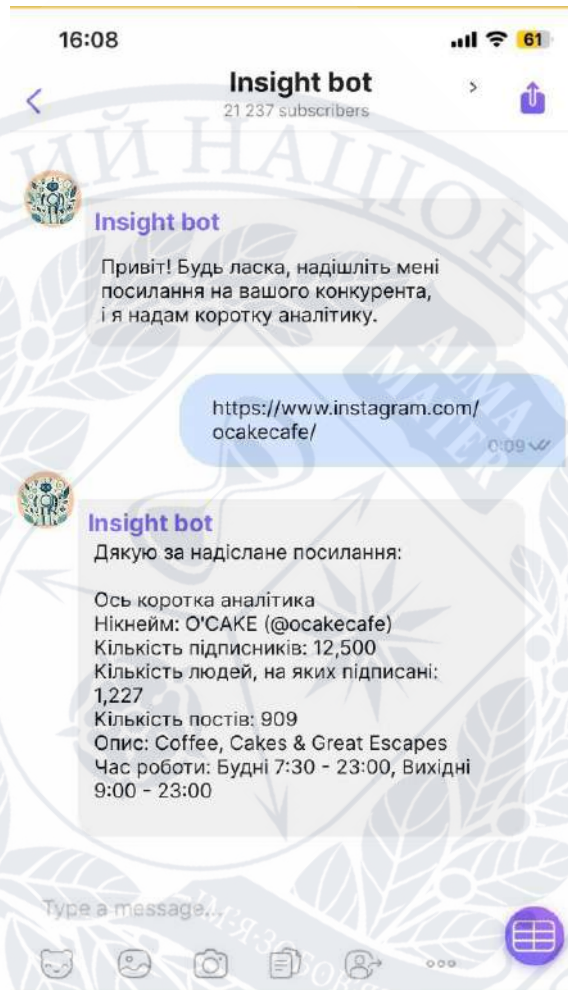


Рисунок 3.10 – Попередній аналіз та вибір напрямку дослідження

Крок 3. Аналіз контенту (пости, візуальне оформлення)

На цьому етапі бот аналізує вміст останніх публікацій, включаючи типи контенту (фото, відео, каруселі), частоту публікацій, візуальну айдентику та стиль оформлення. Це дає змогу зрозуміти контентну стратегію конкурента.



Insight bot

📊 Аналіз контенту конкурента (@osakesafe)

Ми проаналізували останні 10 публікацій Instagram-сторінки @osakesafe — мережі кав'ярень у Вінниці. Ось коротке зведення:

- ♦ Візуальний стиль: Яскраві, естетичні фото десертів, кави та інтер'єру. Увага до деталей, чисті композиції.
- ♦ Тематика постів: Тістечка, торти, кава, атмосфера закладу. Контент візуально однорідний.
- ♦ Підписи: Лаконічні, з емодзі та релевантними хештегами. Створюють легкий і «смачний» тон спілкування.
- ♦ Активність: Лайки та коментарі стабільні. Аудиторія залучена, але без різких стрибків охоплення.

📌 Що можна взяти для себе:

- Використовувати якісні, світлі фото продукції.
- Писати короткі, дружні підписи з акцентом на смак і емоції.
- Впровадити інтерактив — опитування, конкурси або реакції в сторіс.
- Акцентувати сезонні новинки (наприклад, «весняні десерти», «кава з гарбузом» тощо). Цей підхід допомагає формувати впізнаваний стиль і стабільно тримати увагу аудиторії.

Рисунок 3.11 – Результат виконання функції «Аналіз контенту»

Крок 4. Аналіз взаємодії з аудиторією

Бот обчислює рівень залучення аудиторії, включаючи середню кількість лайків, коментарів, а також динаміку зростання підписників. Це дає розуміння ефективності взаємодії конкурента зі своєю аудиторією.



Insight bot

👉 Взаємодія з аудиторією (@osakesafe)

Ось як @osakesafe підтримує контакт зі своєю аудиторією:

- ♦ Коментарі: Регулярні відповіді на запитання щодо меню, годин роботи тощо.
- ♦ Сторіс: Часте використання інтерактиву: опитування, питання, реакції.
- ♦ Заклики до дії: У постах ставлять запитання до підписників — це стимулює коментарі.
- ♦ Акції: Розіграші й конкурси з умовами «підпишись і теги друга».

📌 Що варто запозичити:

- Систематично відповідати на коментарі та згадки.
- Додати інтерактив у сторіс — тести, реакції, питання.
- Стимулювати діалог через пости.
- Запустити прості розіграші для охоплення.

Рисунок 3.12 – Аналіз взаємодії конкурентів з аудиторією

Крок 5. Аналіз використаних хештегів

Бот виділяє хештеги, які найчастіше використовуються в публікаціях конкурента. Це допомагає ідентифікувати ключові теми, які висвітлює конкурент, та дає можливість адаптувати стратегію для власного бренду.

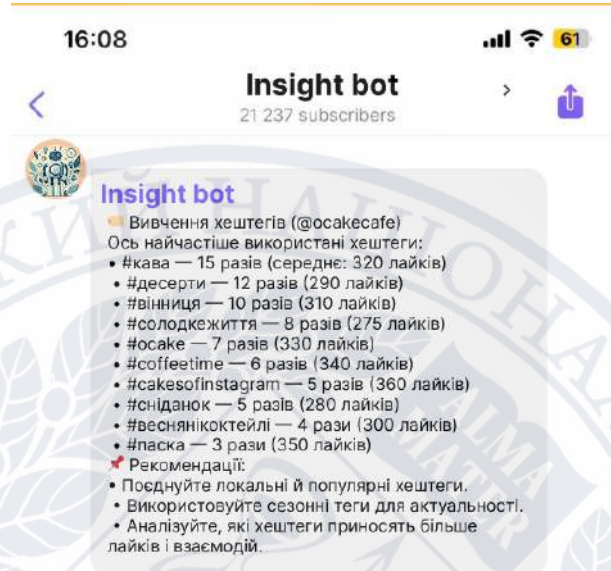


Рисунок 3.13 – Вивчення хештегів

Крок 6. Аналіз активності

Функція «Аналіз активності» в Viber-боті дозволяє дослідити поведінку аудиторії Instagram-акаунту конкурента за останні 14 днів. Після введення імені профілю бот автоматично збирає дані про публікації за вказаний період і обчислює основні показники, зокрема середню кількість лайків і коментарів, визначає найактивніші дні тижня та найефективніші типи контенту.

Додатково формується графік змін активності, що відображає динаміку взаємодій по днях. На основі цих даних бот генерує практичні рекомендації щодо оптимального часу для публікацій, вибору контенту, що найбільше залучає аудиторію, а також дозволяє адаптувати стратегію просування відповідно до змін у поведінці підписників. Такий інструмент є особливо корисним для користувачів, які прагнуть підвищити ефективність власного контенту, орієнтуючись на конкурентів і працюючи з конкретною аналітикою, а не на інтуїтивних припущеннях.

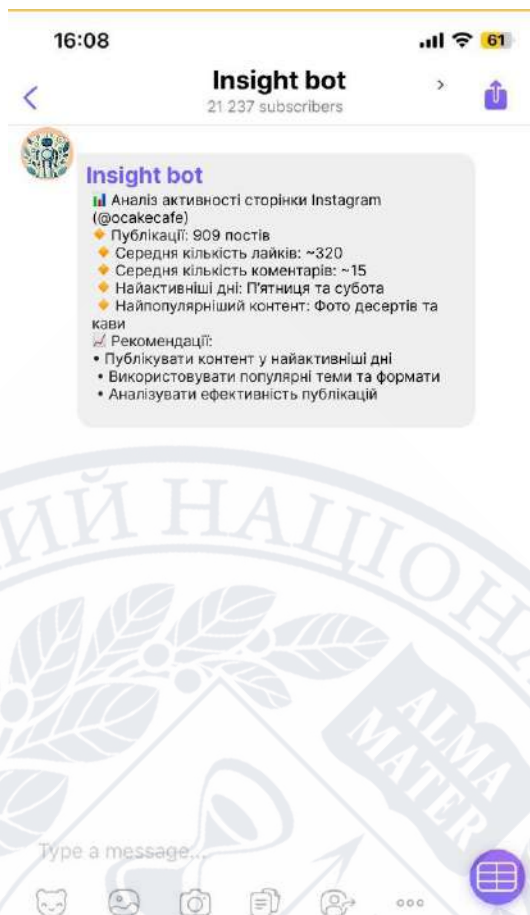


Рисунок 3.14 – Аналіз активності сторінки конкурента

Розділ 3.3 наочно демонструє ефективність розробленого бота як інструменту конкурентного аналізу в соціальних мережах. Покроковий огляд інтерфейсу та функціоналу показує зручність взаємодії для кінцевого користувача — від старту діалогу до отримання глибокої аналітики сторінки конкурента. Бот забезпечує повний цикл збору та аналізу даних: загальна інформація про акаунт, контентний аналіз, взаємодія з підписниками, використані хештеги та рівень активності аудиторії.

Особливу цінність становить функція «Аналіз активності», яка дозволяє приймати стратегічні рішення на основі реальної динаміки поведінки підписників. Завдяки поєднанню автоматизації, простоти інтерфейсу та глибини аналітики бот може бути використаний як інструмент маркетингової аналітики не лише для власників бізнесу, а й для SMM-фахівців, маркетологів та дослідників цифрової поведінки споживачів.

ВИСНОВКИ

У результаті виконання дипломної роботи було створено функціональну систему збору та аналізу даних із соціальних мереж, реалізовану у вигляді Viber-бота. Основною метою стало забезпечення автоматизованого доступу до публічної інформації з профілів конкурентів у мережах Instagram і Facebook, що дозволяє оперативно здійснювати маркетинговий аналіз без використання складних програмних рішень або платних сервісів.

Розроблена система враховує актуальні потреби малого та середнього бізнесу, який потребує швидкого та зручного інструменту для моніторингу цифрової активності конкурентів. Завдяки інтеграції інструментів аналізу контенту, взаємодій з аудиторією та хештегів, бот надає користувачеві зрозумілу та корисну інформацію у зручному форматі. Це дозволяє приймати обґрунтовані рішення щодо стратегії контенту, часу публікацій та формату взаємодії з власною аудиторією.

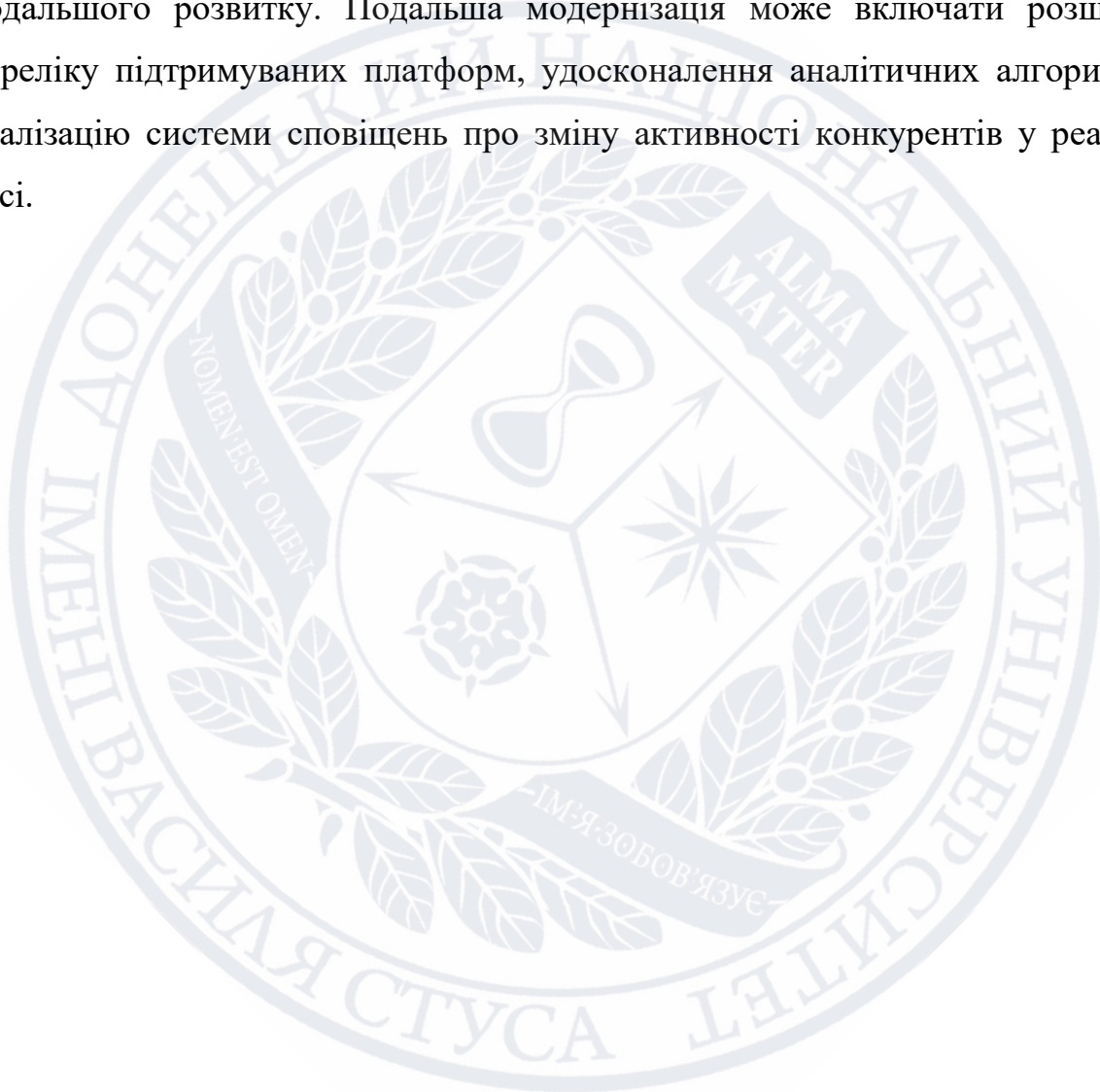
Під час розробки було враховано технічні обмеження платформ, зокрема закритість API, що обумовило використання веб-скрейпінгу як основного методу збору даних. Система показала стабільну роботу навіть при обробці великих обсягів інформації, демонструючи хорошу масштабованість і гнучкість у використанні. Бот автоматично розпізнає тип ресурсу, структурує отримані дані та виводить їх у зрозумілому вигляді, що значно знижує вхідний поріг для користувачів.

Аналітичний модуль бота дозволяє не лише збирати інформацію, але й проводити попередній аналіз динаміки активності, рівня залучення аудиторії та інших показників. Таким чином, система виконує функції не просто технічного інструменту, а повноцінного аналітичного середовища, яке можна використовувати у щоденній роботі маркетолога або аналітика. Візуалізація даних сприяє швидкому ухваленню управлінських рішень на основі поточних цифрових трендів.

Практична цінність розробленої системи полягає у її доступності, простоті впровадження та можливості адаптації під конкретні потреби користувача. Бот не потребує окремого програмного забезпечення, його можна запускати з будь-

якого пристрою, що робить його універсальним інструментом у сфері цифрової аналітики. Такий підхід особливо важливий для організацій, які прагнуть зменшити витрати на ІТ-рішення, не втрачаючи в ефективності.

Підсумовуючи результати дослідження, можна зробити висновок, що поставлена мета була досягнута. Розроблена система є релевантною поточним потребам ринку, функціонально завершеною та перспективною в контексті подальшого розвитку. Подальша модернізація може включати розширення переліку підтримуваних платформ, удосконалення аналітичних алгоритмів та реалізацію системи сповіщень про зміну активності конкурентів у реальному часі.



СПИСОК ВИКОРИСТАНИХ ПОСИЛАНЬ

1. Kyivstar Business Hub. Що таке Big Data: визначення, приклади та можливості URL. Kyivstar Business Hub. – 2024. hub.kyivstar.ua.
2. Кукіна Н.В., Савчук Я.О., Лялюк А.М. Аналіз ролі соціальних мереж у формуванні брендової лояльності. *Актуальні проблеми економіки*. 2024. – № 8(278). С. 116–128.
3. Netpeak Ukraine. Від автоматизації до повторних продажів: як чат-боти допомагають бізнесу рости URL. Netpeak. 2024. netpeak.net.
4. Нановська В. Арсенал соціальних мереж: інструменти, які необхідні кожному SMM-фахівцю URL. *Mediamaker*. 2023. netpeak.net.
5. Proxy-Seller. Веб-скрейпінг проти веб-краулінгу: переваги та недоліки URL. Proxy-Seller. 2024. 27 бер. 2024. proxy-seller.com.
6. Полухін А.В. Автоматизація процесу збору та систематизації публічно доступної інформації з соціальних мереж URL. : кваліфікаційна робота. Київ: КМАН, 2021. – 29 с.
7. Решетнікова О.В., Боровик Т.В., Сенько І.М. Маркетинг у соціальних мережах як інструмент просування бренду та підвищення обсягів збуту. *Економіка та суспільство*. – 2024. – № 65. – С. 876–884.
8. Шестакова А.В. *SMM та аналітика ринку: навч. посіб.* URL. Житомир: Державний ун-т “Житомирська політехніка”, 2021. 215 с. eztuir.ztu.edu.ua.
9. Сімагін Д. Як створити Viber-бот на Python за 30 хвилин. *Highload Tech*. – 2024.
10. Супруненко С.М., Чорновол А., Гаврилюк В. Використання аналітики даних для управління фінансовими процесами в цифровому середовищі України. *Економіка та суспільство*. – 2024. – № 62. – С. 51–57.
11. TopDigital UA. Як виконати конкурентний аналіз – класичні та сучасні

методики URL. TopDigital. – 2023.

12. Cukier K., Mayer-Schönberger V. Big Data: A Revolution That Will Transform How We Live, Work, and Think. – Boston: Houghton Mifflin Harcourt, 2013.
13. Krug S. Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability. – Berkeley: New Riders, 2014.
14. Liu B. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. 2nd ed. – New York: Springer, 2007.
15. Sponder M. Social Media Analytics: Effective Tools for Building, Interpreting, and Delivering Business Insights from Social Media. – New York: McGraw-Hill, 2011.
16. Russell M.A. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More. 3rd ed. – Sebastopol, CA: O'Reilly, 2018.
17. Shevat A. Designing Bots: Creating Conversational Experiences. – Sebastopol, CA: O'Reilly, 2017.
18. Sweigart A. Automate the Boring Stuff with Python: Practical Programming for Total Beginners. 2nd ed. – San Francisco: No Starch Press, 2019.
19. Tidwell J. Designing Interfaces: Patterns for Effective Interaction Design. 3rd ed. – Sebastopol, CA: O'Reilly, 2019.
20. Zafarani R., Abbasi A., Liu H. Social Media Mining: An Introduction. – Cambridge: Cambridge University Press, 2014.
21. Norman D. The Design of Everyday Things. Rev. ed. – New York: Basic Books, 2013.
22. Laudon K.C., Laudon J.P. Management Information Systems: Managing the Digital Firm. 15th ed. – London: Pearson, 2018.
23. Wasserman S., Faust K. Social Network Analysis: Methods and Applications. – Cambridge: Cambridge University Press, 1994.
24. Borgatti S.P., Everett M.G., Johnson J.C. Analyzing Social Networks. – Thousand Oaks: SAGE Publications, 2013.
25. Viber Bot API / Viber. – 2025

26. Instagram Graph API / Meta for Developers. – 2024
27. Facebook Graph API / Meta for Developers. – 2025
28. Іванюк Т. Telegram-боти для бізнесу: переваги, кейси, обмеження. AIN.UA. – 2024. Режим доступу: <https://ain.ua>
29. Міністерство цифрової трансформації України. Цифрова трансформація та штучний інтелект в Україні. – 2023. – Режим доступу: <https://thedigital.gov.ua>
30. Дудар А.О. Застосування Python для аналізу соціальних медіа в маркетингових дослідженнях / Маркетинг і цифрові технології. – 2023. – № 2(14). – С. 45–52.
31. Стеценко І.Ю. Застосування чат-ботів у системах електронного урядування / Інформаційні технології і засоби навчання. – 2023. – № 4(92). – С. 111–119.
32. Яворська І.В. Інтернет-маркетинг: сучасні інструменти просування в соціальних мережах / Економіка та держава. – 2022. – № 7. – С. 52–55.
33. Олійник Д.С. Боти в месенджерах як засіб комунікації бренду з клієнтами / Маркетинг і менеджмент інновацій. – 2023. – № 3. – С. 98–105.
34. Гринчук О.О. Веб-скрейпінг як метод збирання даних для соціологічних досліджень / Вісник Київського національного університету імені Тараса Шевченка. Соціологія. 2023. № 1. – С. 37–42.
35. Журавель М.І. Системи аналізу даних у бізнес-аналітиці. Бізнес Інформ. 2023. № 12. – С. 41–46.
36. Ковальчук О.М. Аналіз користувацької активності в Instagram: методика та інструменти. Маркетинг та цифрові технології. 2022. – № 1(12). С. 24–30.
37. Чорна А.М. Telegram-боти у навчальному процесі: досвід і перспективи. Освітній простір України. 2023. № 2. – С. 67–73.
38. Гудзь М.П. Розробка та використання систем для моніторингу соціальних медіа. Інформаційні технології в освіті. 2022. № 50. – С. 33–40.