

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

УДК 004.4:004.738

«ПОГОДЖЕНО»

Декан факультету
інформаційних технологій

Болбот І. М., д.п.н., професор

«ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ»

Завідувач кафедри комп'ютерних наук

Голуб Б.Л., к.т.н., доцент

_____ 202_ р.

_____ 202_ р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему “Програмне забезпечення системи аналізу популярності відео в інтернеті”

Спеціальність _121 – Інженерія програмного забезпечення

(код і назва)

Освітня програма Програмне забезпечення інформаційних систем

(назва)

Орієнтація освітньої програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Гарант освітньої програми

професор, д.т.н.

(науковий ступінь та вчене звання)

Семко В. В.

(підпис)

(ПІБ)

Керівник магістерської кваліфікаційної роботи

старший викладач

(науковий ступінь та вчене звання)

Міловідов Ю. О.

(підпис)

(ПІБ)

Виконав

Масюк Д. В.

(підпис)

(ПІБ студента)

ЗМІСТ

ЗМІСТ	2
ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	4
ВСТУП	5
1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	7
1.1 Опис предметної області.....	7
1.2 Аналіз існуючих рішень та їх проблем.....	7
1.3 Постановка завдання для магістерської роботи	8
2 МОДЕЛЮВАННЯ СИСТЕМИ.....	9
2.1 Діаграма прецедентів.....	9
2.2 Діаграма послідовності	10
2.3 Діаграма активностей	10
2.4 Діаграма компонентів.....	11
2.5 Абстракції предметної області	12
3 РОЗРОБКА СИСТЕМИ	13
3.1 Структура сховища даних.....	13
3.2 Структура LLM	14
3.3 Діаграма пакетів.....	16
3.4 Технології	16
4 ПОДАЛЬШИЙ РОЗВИТОК І ВДОСКОНАЛЕННЯ.....	18
4.1 Neo4j.....	18
4.2 Microsoft GraphRAG	20
4.3 Аналіз субтитрів.....	21

4.4	Vision	22
5	РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ	24
5.1	Результати.....	24
5.2	Апаратні та програмні вимоги.....	32
	ВИСНОВОК.....	33
	СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	35

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

СД – сховище даних.

LLM – велика мовна модель.

ВСТУП

Актуальність. Аналіз популярності відео є важливим завданням для створювачів контенту, маркетологів, рекламодавців та інших зацікавлених сторін. Розуміння того, які відео є найбільш популярними та чому, допомагає приймати обґрунтовані рішення щодо створення та просування контенту. Актуальність даного дослідження полягає у розробці інструменту для аналізу популярності відео в інтернеті.

Об'єкт. Об'єктом дослідження є процес аналізу популярності відео в інтернеті. Предметом дослідження є методи та алгоритми для збору, обробки та візуалізації даних про популярність відео.

Мета дослідження. Мета даного дослідження є розробка програмного забезпечення для ефективного аналізу популярності відео в інтернеті. Система повинна збирати, обробляти та візуалізувати дані про перегляди, лайки, коментарі.

Завдання. Метою роботи є розробка програмного забезпечення для аналізу популярності відео в інтернеті. Для досягнення мети передбачається виконання таких завдань:

Провести системний аналіз існуючих підходів до аналізу популярності відео в інтернеті.

Сформулювати вимоги до програмного забезпечення системи аналізу популярності відео.

Побудувати моделі предметної області, що описують зв'язки між каналами, відео.

Розробити архітектуру системи дослідження, яка забезпечує ефективний збір, обробку та аналіз даних.

Розробити програмне забезпечення із застосуванням сучасних технологій обробки даних.

Провести тестування програмного забезпечення та оцінити його ефективність.

Сформулювати висновки та результати дослідження.

Методи дослідження. У роботі застосовуються такі методи та технології: OLAP-технології – для аналізу даних та розрахунку ключових показників популярності відео.

LLM – для дискретизацій назв відео.

Наукова новизна.

Запропоновано сукупність інформаційних технологій таких як СД та LLM для аналізу популярності відео в інтернеті.

Апробація результатів дослідження.

XV МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ МОЛОДИХ ВЧЕНИХ «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ: ЕКОНОМІКА, ТЕХНІКА, ОСВІТА»

XIV МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ МОЛОДИХ ВЧЕНИХ «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ: ЕКОНОМІКА, ТЕХНІКА, ОСВІТА»

Структура роботи.

Документ має 35 сторінок, 8 використаних джерел, 5 розділів:

Системний аналіз предметної області – аналіз сучасних підходів і інструментів для оцінки популярності відео та визначення вимог до системи.

Моделювання системи – розробка архітектури, UML-діаграм та потоків даних для побудови системи аналізу.

Розробка системи – реалізація програмного забезпечення для збору, обробки та візуалізації даних про популярність відео.

Подальший розвиток і вдосконалення – рекомендації щодо масштабування, оптимізації та впровадження нових функцій.

Результати дослідження – аналіз отриманих даних.

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Опис предметної області

Популярність відео в інтернеті є однією з ключових метрик успішності контенту на платформах відеохостингу, таких як YouTube. Вона визначається за допомогою показників, таких як кількість переглядів, лайків та інші параметри, що характеризують взаємодію аудиторії з відео. Аналіз цих показників дозволяє виявляти тренди, оцінювати ефективність контенту та створювати рекомендації для його оптимізації.

Предметна область включає:

- Джерела даних: метрики відео, характеристики каналів.
- Процеси аналізу: збір, фільтрація, обробка даних.
- Кінцевий результат: рекомендації для створення та просування контенту.

1.2 Аналіз існуючих рішень та їх проблем

На сьогодні існує низка інструментів для аналізу популярності відео, включаючи вбудовані аналітичні сервіси платформ (наприклад, YouTube Analytics) та сторонні програми (наприклад, Social Blade, TubeBuddy, Vidooly).

Переваги існуючих рішень:

- Автоматизований збір даних з платформ.
- Наявність візуалізації метрик для користувачів.
- Простота у використанні для базового аналізу.

Основні проблеми:

- Обмеженість функціоналу: більшість сервісів не враховують взаємозв'язок між факторами популярності (час публікації, теми відео).
- Відсутність глибокого аналізу: недостатня інтеграція алгоритмів LLM.

1.3 Постановка завдання для магістерської роботи

Метою магістерської роботи є розробка програмного забезпечення для аналізу популярності відео в інтернеті, яке дозволить усунути недоліки існуючих рішень.

Для досягнення цієї мети необхідно виконати такі завдання:

- Провести аналіз існуючих підходів до аналізу популярності відео та виявити ключові фактори впливу.
- Розробити модель предметної області, яка враховує основні показники та взаємозв'язки між ними.
- Створити архітектуру системи, яка забезпечує збір та обробці даних.
- Реалізувати програмне забезпечення із застосуванням сучасних технологій аналізу даних (OLAP, LLM).
- Забезпечити для користувача візуалізації результатів аналізу.

2 МОДЕЛЮВАННЯ СИСТЕМИ

2.1 Діаграма прецедентів

На рис. 1 зображено діаграму прецедентів в якій є Актор Менеджер та Use case Аналіз даних, Завантаження даних, Збереження в СД, Визначення теми відео.

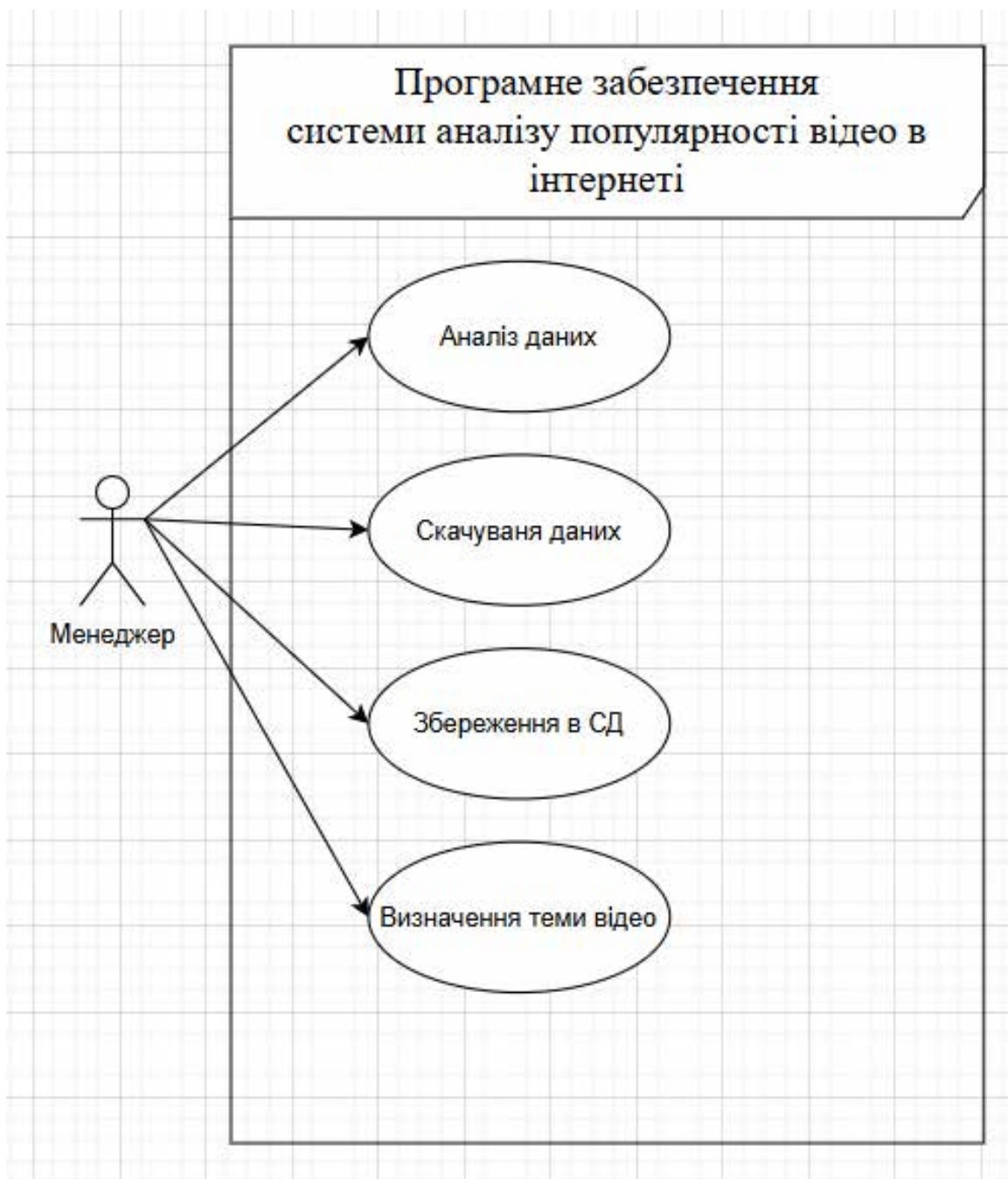


Рис. 1 Діаграма прецедентів

2.2 Діаграма послідовності

На рис. 2 зображено діаграму послідовностей.

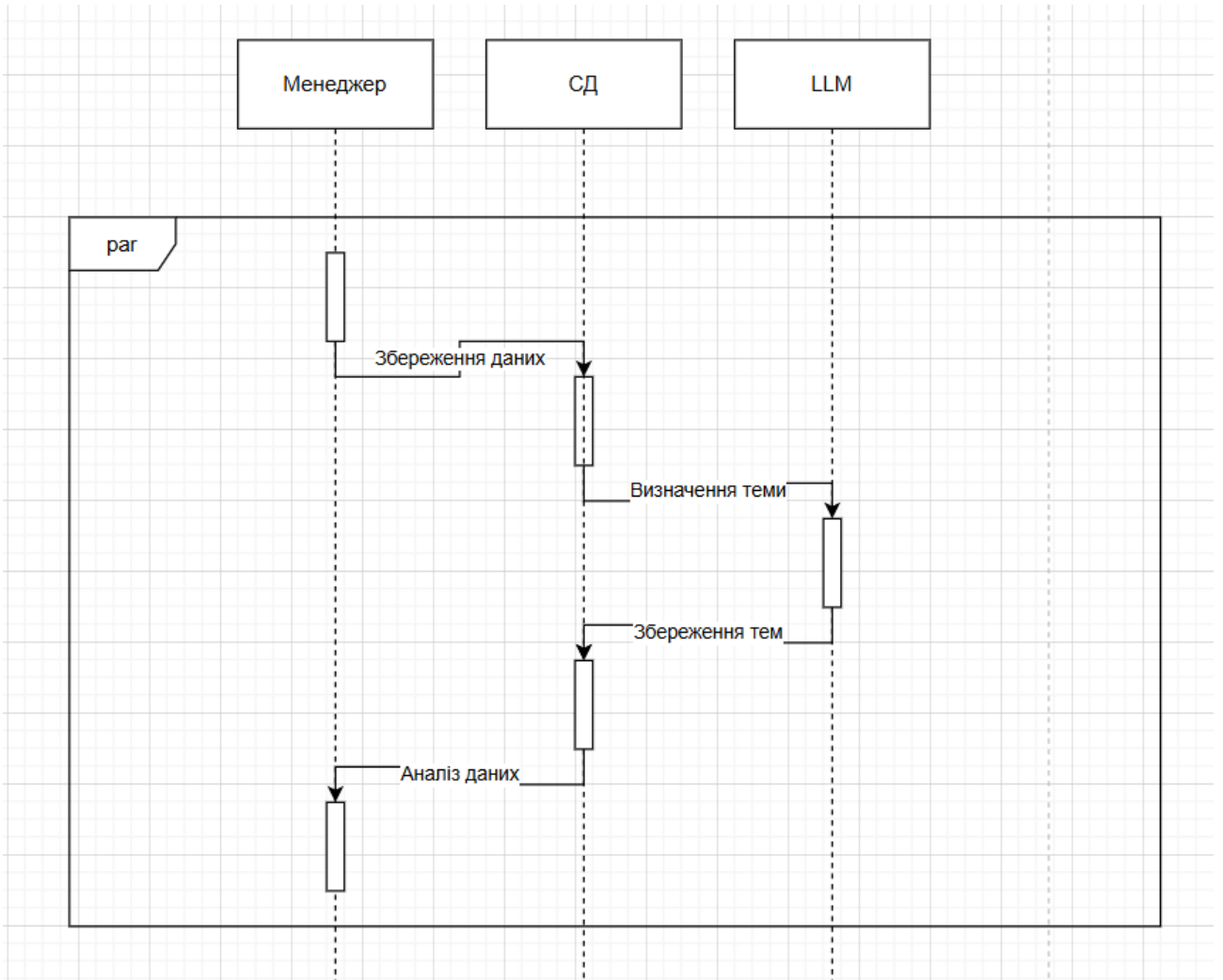


Рис. 2 Діаграма послідовностей

2.3 Діаграма активностей

На рис. 3 зображено діаграму активностей.

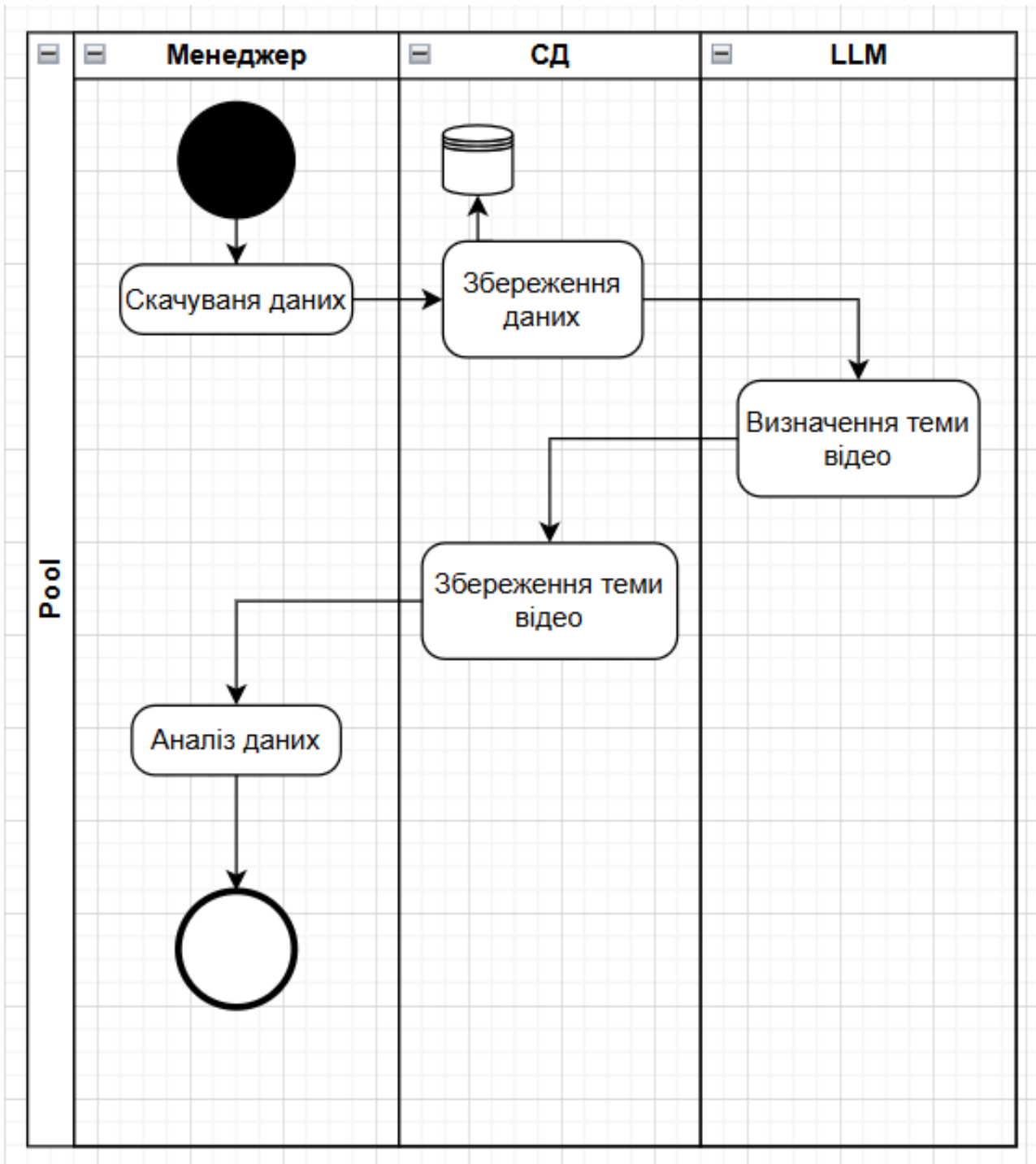


Рис. 3 Діаграма активностей

2.4 Діаграма компонентів

На рис. 4 зображено діаграму компонентів. На ній зображено Main.py, get_video_urls.py, get_video_topic.py, download_data_youtube.py.

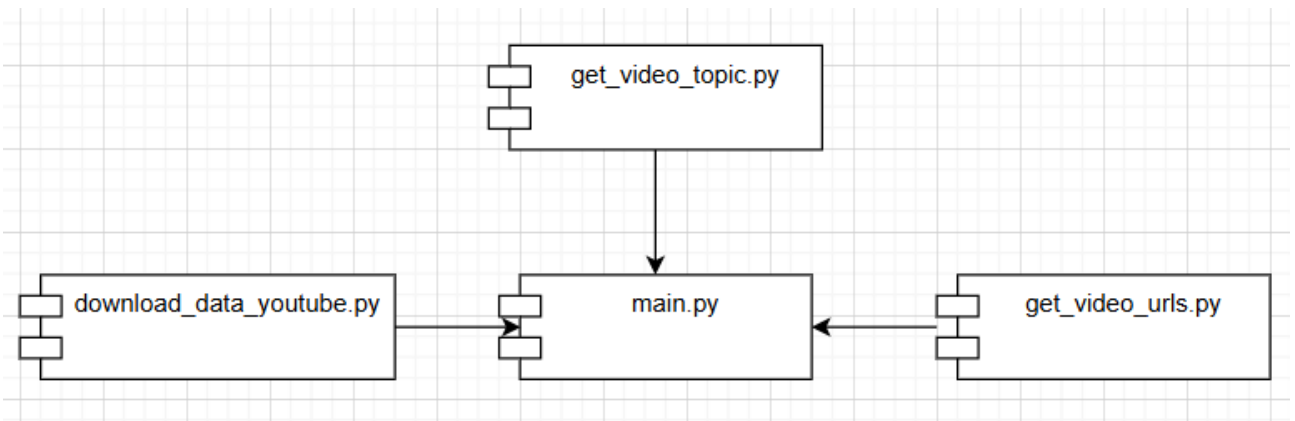


Рис. 4 Діаграма компонентів

2.5 Абстракції предметної області

На рис. 5 зображено абстракції предметної області: Відео та Канал.

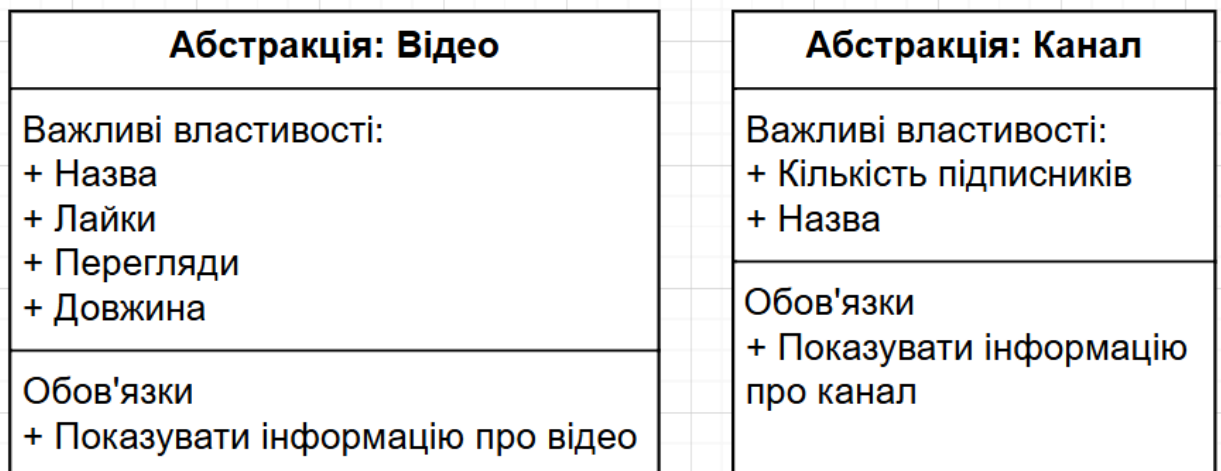


Рис. 5 Абстракції предметної області

3 РОЗРОБКА СИСТЕМИ

3.1 Структура сховища даних

Для аналізу була побудоване сховище даних в формі сніжинка яку можна переглянути на рис. 6 за допомогою Microsoft sql server.

“Microsoft SQL Server — система управління базами даних, яка розробляється корпорацією Microsoft. Як сервер даних виконує головну функцію по збереженню та наданню даних у відповідь на запити інших застосунків, які можуть виконуватися як на тому ж самому сервері, так і у мережі.

Мова, що використовується для запитів — Transact-SQL, створена спільно Microsoft та Sybase. Transact-SQL є реалізацією стандарту ANSI / ISO щодо структурованої мови запитів SQL із розширеннями. Використовується як для невеликих і середніх за розміром баз даних, так і для великих баз даних масштабу підприємства. Багато років вдало конкурує з іншими системами керування базами даних.”[2]

Логічна модель даних у вигляді діаграми зображена на рис. 1 та яка має наступні виміри:

- Topic_dim – вимір теми відео. Вимір має ключовий атрибут Topic_ID та не ключовий атрибут Topic_Name – назва теми.
- Channel_dim – вимір youtube каналу. Вимір має ключовий атрибут Channel_ID та не ключові атрибути Channel_Name – назва каналу, Subscriber_Count – кількість підписників.
- Date_dim – часовий вимір. Вимір має ключовий атрибут Date_ID та не ключові атрибути Year – рік, Mouth – місяць, Day – день.
- Video_dim – вимір відео. Вимір має ключовий атрибут Video_ID та не ключовий атрибут Video_Title – назва відео.
- Emoji_dim – вимір емодзі. Вимір має ключовий атрибут Emoji_ID та не ключові атрибути Video_ID, Emoji_Name – символ емодзі.

- Video_facts – факти. Факти має такі атрибути. Video_ID, Channel_ID, Date_ID, Topic_ID, View_Count – кількість переглядів в відео типом string, Video_Length – довжина відео типом string, View_Count_Int – кількість переглядів в відео, Video_Length_Int – довжина відео в секундах.

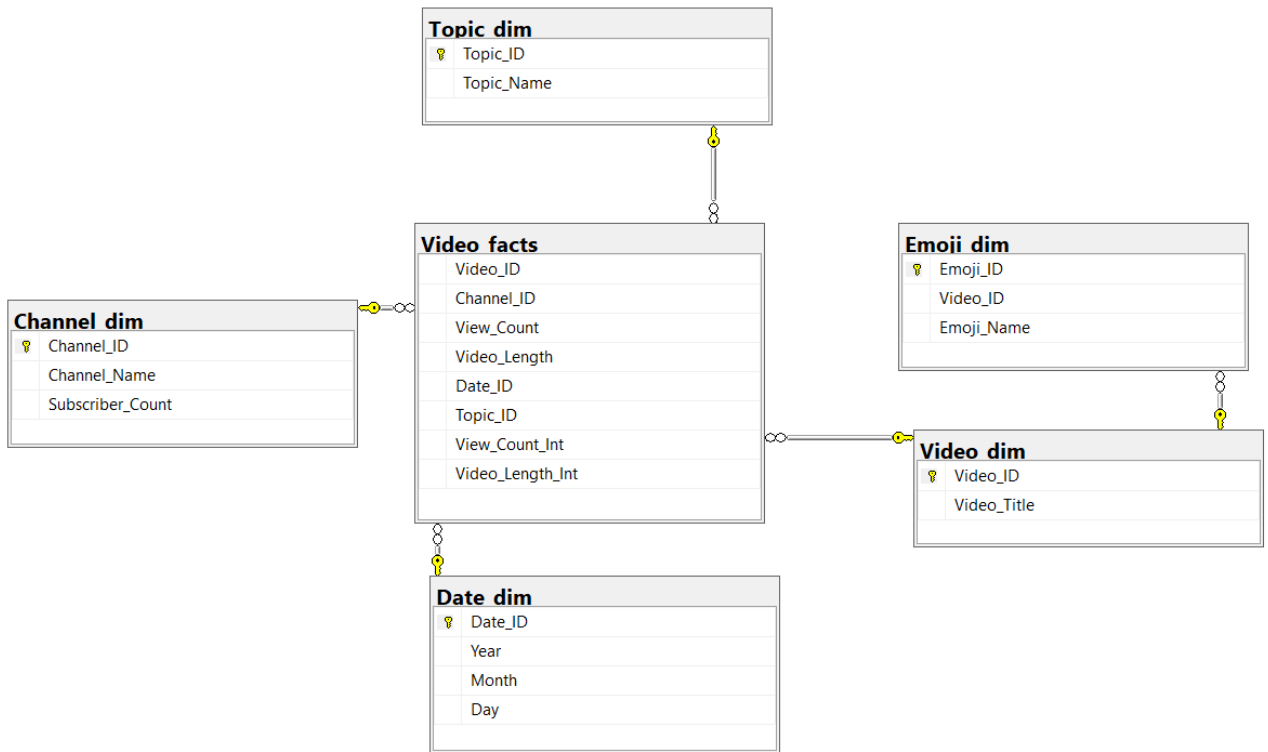


Рис. 6 Сховище даних

3.2 Структура LLM

Щоб визначати тему для того, щоб детермінувати для подальшого аналізу, був використаний сервер Ollama для запуску LLM llama 3.2:3b схему роботи якої можна переглянути на рис. 7.

“Легкі моделі 1B і 3B мають високі можливості для створення багатомовного тексту та можливостей виклику інструментів. Ці моделі дають змогу розробникам створювати персоналізовані агентські програми на пристрої з високою конфіденційністю, де дані ніколи не залишають пристрій. Наприклад, така програма може допомогти підсумувати 10 останніх отриманих повідомлень, витягнути елементи дії та використовувати інструмент виклику для прямого надсилання запрошень календаря на наступні зустрічі.

Локальний запуск цих моделей має дві основні переваги. По-перше, підказки та відповіді можуть здаватися миттєвими, оскільки обробка виконується локально. По-друге, запуск моделей локально забезпечує конфіденційність, не надсилаючи такі дані, як повідомлення та інформацію календаря, у хмару, що робить програму в цілому більш конфіденційною. Оскільки обробка обробляється локально, програма може чітко контролювати, які запити залишаються на пристрої, а які, можливо, потрібно буде обробити більшою моделлю в хмарі.

Як ми вже говорили про Llama 3.1, потужні моделі вчителів можна використовувати для створення менших моделей із покращеною продуктивністю. У моделях 1B і 3B ми використали два методи — обрізання та дистиляцію, що зробило їх першими високопродуктивними легкими моделями Llama, які можна ефективно встановити на пристрої.

Обрізка дозволила нам зменшити розмір наявних моделей у стадії лам, одночасно відновивши якомога більше знань і продуктивності. Для моделей 1B і 3B ми застосували підхід до використання структурованого обрізання за один раз від Llama 3.1 8B. Це передбачало систематичне видалення частин мережі та коригування величини ваг і градієнтів для створення меншої, ефективнішої моделі, яка зберігає продуктивність вихідної мережі.

Дистиляція знань використовує більшу мережу для передачі знань у меншу мережу з ідеєю, що менша модель може досягти кращої продуктивності за допомогою вчителя, ніж це можна було б зробити з нуля. Для 1B і 3B в Llama 3.2 ми включили логіти з моделей Llama 3.1 8B і 70B на етап попереднього навчання розробки моделі, де результати (логіти) з цих більших моделей використовувалися як цілі на рівні маркерів. Для відновлення продуктивності після обрізки було використано дистиляцію знань.”[1]

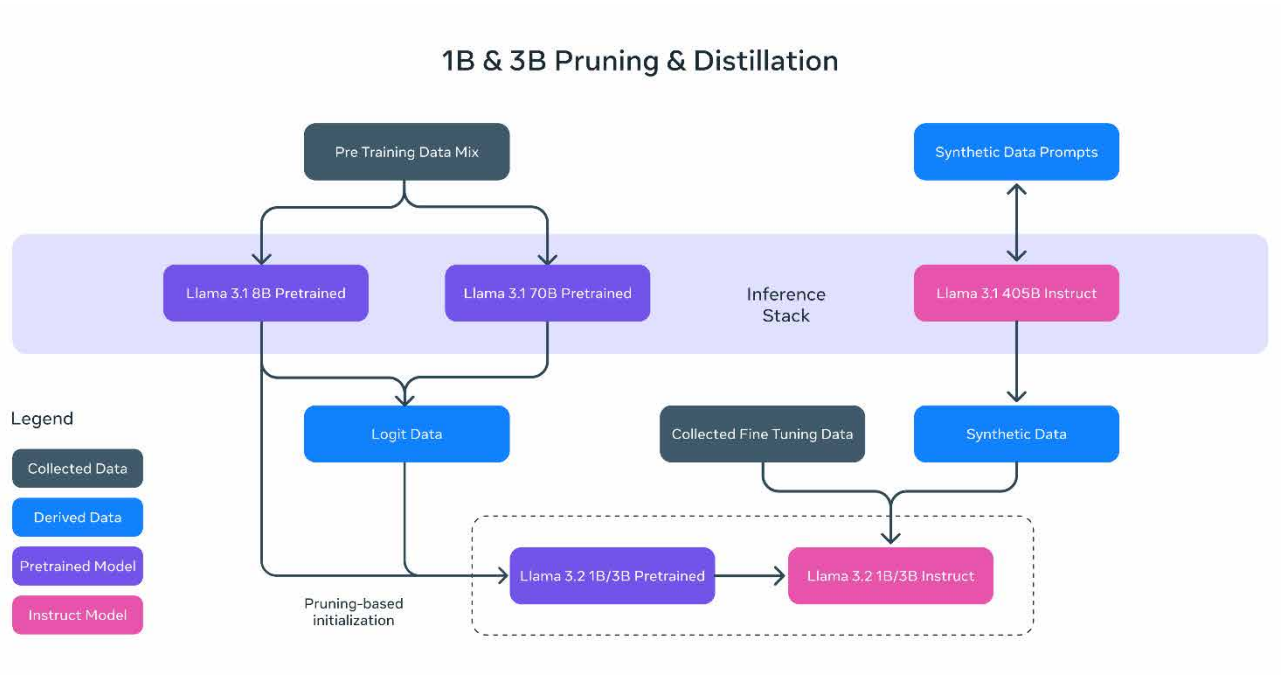


Рис 7 Схема роботи Llama 3.2

3.3 Діаграма пакетів

На рис. 8 зображено діаграму пакетів.

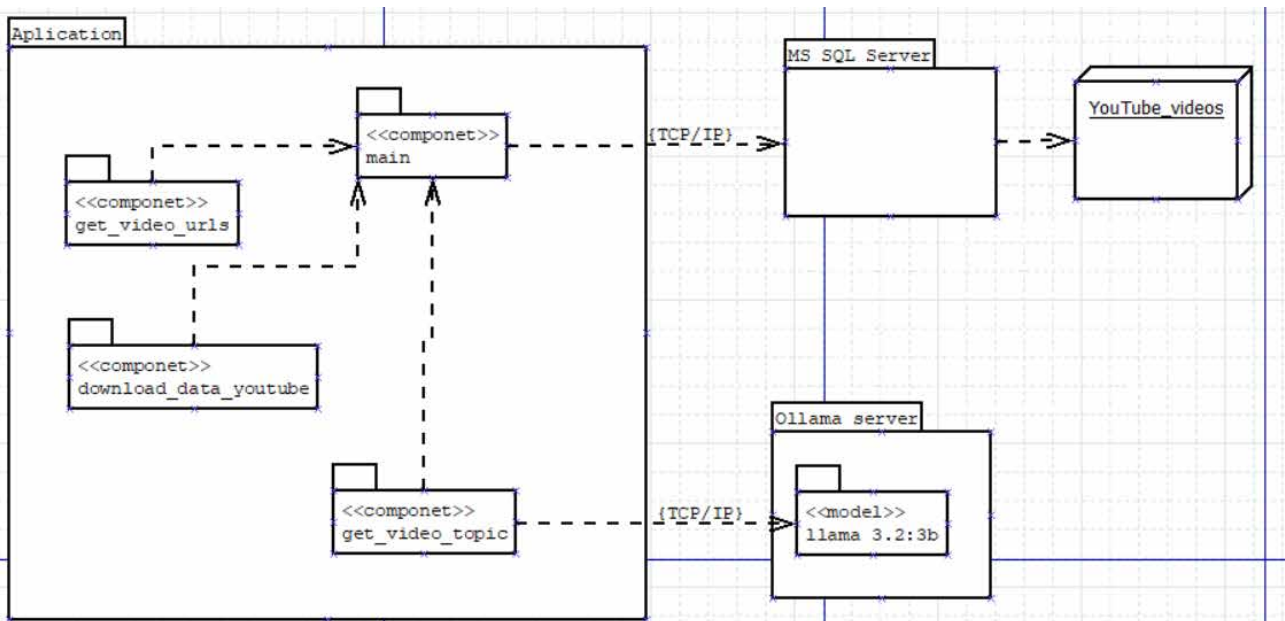


Рис. 8 Діаграма пакетів

3.4 Технології

Під час розробки було використано такі технології:

- Python - зручна і потужна мова з багатою екосистемою бібліотек для роботи з даними.
- yt-dlp - бібліотека для отримання даних про відео з різних платформ, що надає зручні інструменти для завантаження та обробки метаданих.
- Pandas “pandas має на меті бути основним будівельним блоком високого рівня для практичного аналізу реальних даних у Python. Крім того, він має ширшу мету — стати найпотужнішим і найгнучкішим інструментом для аналізу/маніпулювання даними з відкритим кодом, доступним будь-якою мовою.”[3]
- Matplotlib –“ Matplotlib — це комплексна бібліотека для створення статичних, анімованих та інтерактивних візуалізацій на Python. Matplotlib робить легкі речі легкими, а складні – можливими.”[4].
- Customtkinter – “CustomTkinter — це настільна бібліотека інтерфейсу користувача Python на основі Tkinter, яка надає сучасні та повністю налаштовані віджети. За допомогою CustomTkinter ви отримаєте узгоджений вигляд на всіх платформах настільних ПК (Windows, macOS, Linux).”[5]
- Selenium – “WebDriver нативно керує браузером, як і користувач, локально або на віддаленій машині за допомогою сервера Selenium. Це знаменує стрибок вперед у плані автоматизації браузера. Selenium WebDriver відноситься як до мовних прив’язок, так і до реалізацій окремого керуючого коду браузера. Це зазвичай називають просто WebDriver . Selenium WebDriver є рекомендацією W3C: WebDriver розроблений як простий і більш лаконічний інтерфейс програмування. WebDriver — це компактний об’єктно-орієнтований API. Він ефективно керує браузером.”[6]

4 ПОДАЛЬШИЙ РОЗВИТОК І ВДОСКОНАЛЕННЯ

Із за малих апаратних потужностей було неможливо або в не якісному вигляді реалізувати наступні технології, які б доповнили це дослідження.

4.1 Neo4j

Створення Knowledge(рис. 9) та Vector graph для глибшого аналізу відео за допомогою neo4j який можна переглянути на рис. 10.

“Ми на порозі усвідомлення того, що для того, щоб зробити щось суттєво корисне з GenAI, ви не можете залежати лише від авторегресійних LLM для прийняття рішень. Я знаю, що ви думаєте: «RAG — це відповідь». Або тонка настройка, або GPT-5.

Такі методи, як векторний RAG і точне налаштування, можуть допомогти. І вони досить хороші для деяких випадків використання. Але є ще цілий клас випадків використання, коли всі ці техніки натикаються на стелю. Векторна RAG – так само як і тонке налаштування – збільшує ймовірність правильної відповіді на багато питань. Однак жодна з цих методик не забезпечує впевненості у правильній відповіді. Часто їм також бракує контексту, кольору та зв'язку з тим, що ви знаєте як правду . Крім того, ці інструменти не залишають вам багато підказок про те, чому вони прийняли певне рішення.

Повертаючись до сьогоднішнього дня, інженери штучного інтелекту та передові науковці відкривають те саме, що й Google: секрет подолання цієї стелі — це графіки знань. Іншими словами, внесіть знання про речі в комбінацію статистичних текстових технік. Це працює так само, як будь-який інший тип RAG, за винятком виклику графа знань на додаток до векторного індексу. Або іншими словами, GraphRAG !

Ця публікація призначена для вичерпного та легкого для читання опису GraphRAG. Виявилось, що створення графа знань із ваших даних і використання його в RAG дає вам кілька потужних переваг. Є надійний масив досліджень, які

доводять, що він дає вам кращі відповіді на більшість, якщо не на ВСІ запитання, які ви можете поставити LLM, використовуючи звичайний векторний RAG.

Саме це стане величезним рушієм впровадження GraphRAG. Крім того, ви отримуєте легшу розробку завдяки видимості даних під час створення програми. Третя головна перевага полягає в тому, що графіки можуть бути легко зрозумілі та аргументовані як людьми, так і машинами. Тому будувати за допомогою GraphRAG легше, дає кращі результати, і – це вбивця в багатьох галузях – це можна пояснити та перевірити! Я вважаю, що GraphRAG включить лише векторну RAG і стане стандартною архітектурою RAG для більшості випадків використання. Ця публікація пояснює, чому.”[7]

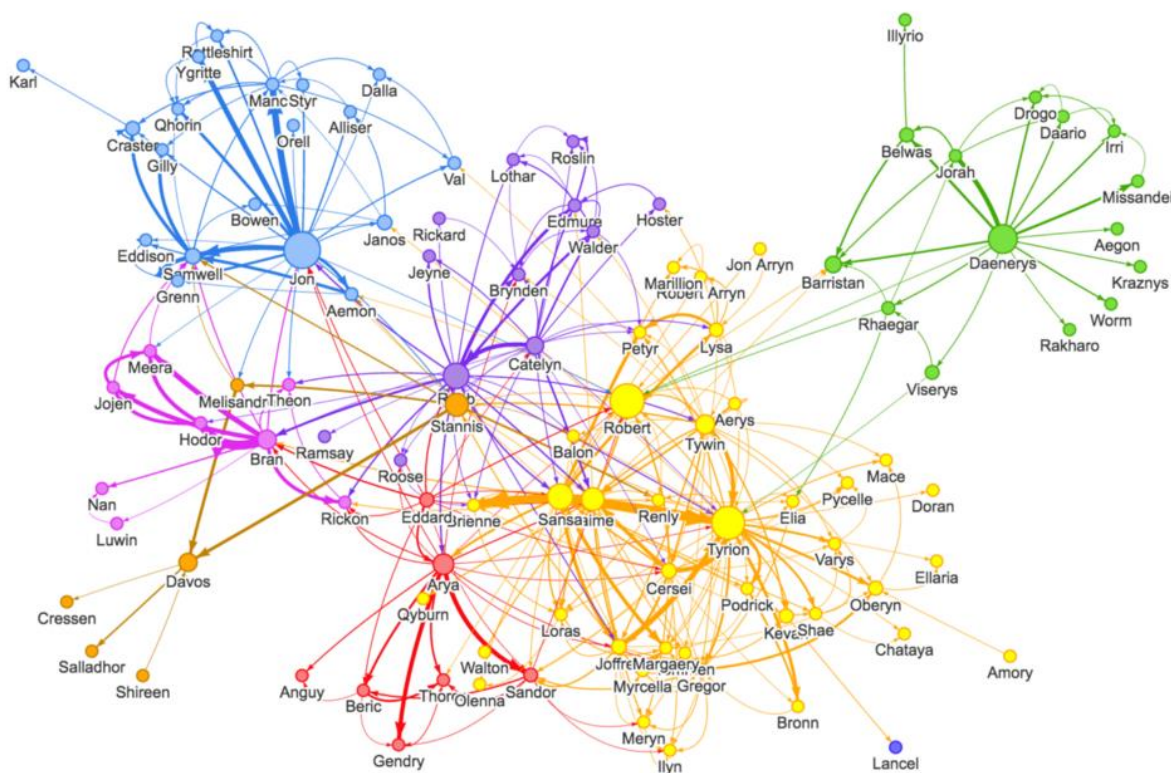


Рис. 9 Knowledge graph

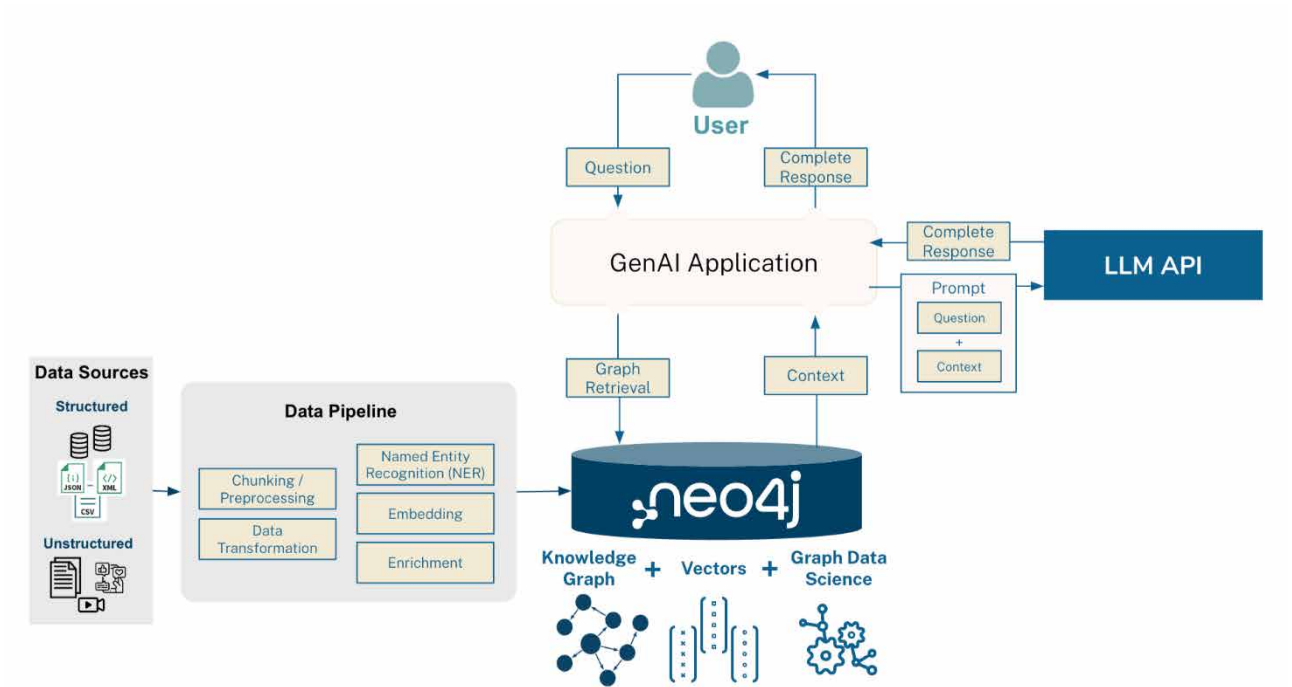


Рис. 10 Архітектура neo4j

4.2 Microsoft GraphRAG

Neo4j в поєднанні GraphRAG для створення Knowledge Graph, приклад на рис. 11.

“GraphRAG — це структурований ієрархічний підхід до Retrieval Augmented Generation (RAG), на відміну від простих підходів семантичного пошуку з використанням простих текстових фрагментів. Процес GraphRAG включає вилучення графа знань із необробленого тексту, створення ієрархії спільнот, генерування підсумків для цих спільнот, а потім використання цих структур під час виконання завдань на основі RAG.

Щоб дізнатися більше про GraphRAG і про те, як його можна використовувати для покращення здатності ваших магістрів LLM міркувати про ваші особисті дані, відвідайте публікацію в блозі Microsoft Research .”[8]

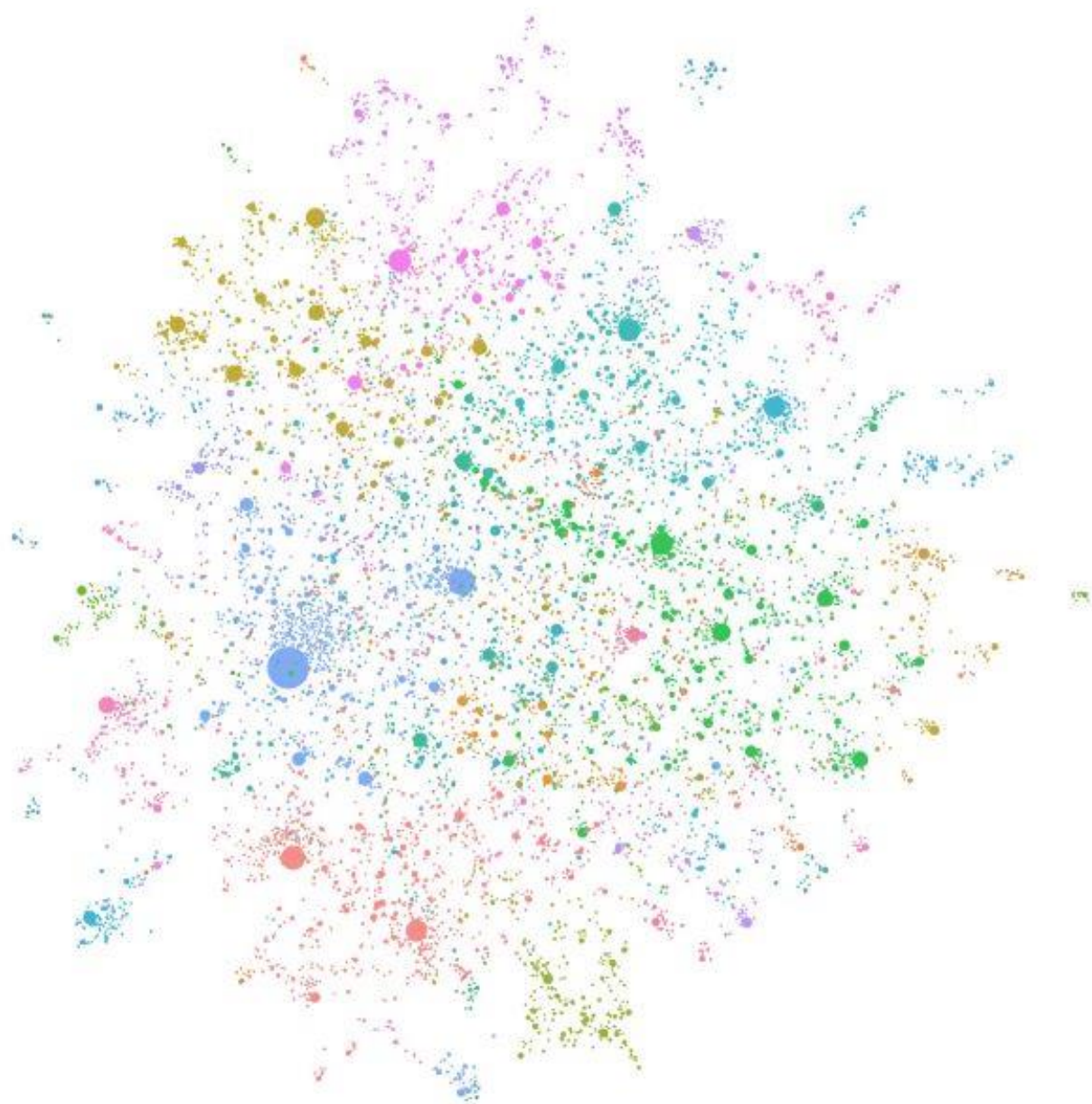


Рис. 11 Граф знань, створений LLM, побудований за допомогою GPT-4 Turbo

4.3 Аналіз субтитрів

Аналіз субтитрів дуже продуктивна затратна, але надає більший контекст про відео. Проблематика заключається в тім що в деяких відео не створюються субтитри і їх потрібно генерувати або їхня якість низька із за чого буде поганий контекст про відео.

4.4 Vision

Великою частиною популярністю це pre view картинка. Від аналізу можна отримати точніші знання про популярність відео.

“Як перші моделі Llama, які підтримували завдання зору, моделі 11B і 90B вимагали абсолютно нової архітектури моделі, яка підтримує логічний аналіз зображень.

Щоб додати підтримку введення зображення, ми навчили набір ваг адаптерів, які інтегрують попередньо навчений кодувальник зображень у попередньо навчену мовну модель. Адаптер складається з ряду рівнів перехресного звернення уваги, які передають представлення кодера зображення в мовну модель. Ми навчили адаптер на парах текст-зображення, щоб узгодити представлення зображення з представленнями мови. Під час навчання адаптера ми також оновили параметри кодувальника зображень, але навмисно не оновили параметри мовної моделі. Роблячи це, ми зберігаємо всі текстові можливості недоторканими, надаючи розробникам заміну для моделей Llama 3.1.

Наш навчальний конвеєр складається з кількох етапів, починаючи з попередньо підготовлених текстових моделей Llama 3.1. Спочатку ми додаємо адаптери зображень і кодери, а потім попередньо навчаємося на великомасштабних зашумлених (зображення, текст) парних даних. Далі ми навчаємося на парних даних високої якості в домені та з розширеними знаннями (зображення, текст).

У постнавчанні ми використовуємо рецепт, подібний до текстових моделей, виконуючи кілька раундів вирівнювання на контрольованому точному налаштуванні, вибірці відхилень і оптимізації прямих переваг. Ми використовуємо генерацію синтетичних даних за допомогою моделі Llama 3.1 для фільтрації та доповнення запитань і відповідей поверх зображень у домені, а також використовуємо модель винагороди для ранжування всіх відповідей кандидатів, щоб надати високоякісні дані точного налаштування. Ми також додаємо дані щодо зниження рівня безпеки, щоб створити модель із високим рівнем безпеки, зберігаючи корисність режиму

Кінцевим результатом є набір моделей, які сприймають як зображення, так і текстові підказки, а також глибоко розуміють і обґрунтовують поєднання. Це ще один крок до того, щоб моделі Llama мали ще багатші агентські можливості.”[1]

Результати можна переглянути на рис. 12.

Vision instruction-tuned benchmarks

Modality	Category Benchmark	Llama 3.2 11B	Llama 3.2 90B	Claude 3 - Haiku	GPT-4o-mini
Image	College-level Problems and Mathematical Reasoning MMMU (v1, 0-shot CoT, micro avg accuracy)	50.7	60.3	50.2	59.4
	MMMU-Pro, Standard (10 opts, test)	33.0	45.2	27.3	42.3
	MMMU-Pro, Vision (test)	23.7	33.8	20.1	36.5
	MathVista (testmini)	51.5	57.3	46.4	56.7
	Charts and Diagram Understanding ChartQA (test, 0-shot CoT relaxed accuracy)*	83.4	85.5	81.7	—
	A12 Diagram (test)*	91.1	92.3	86.7	—
	DocVQA (test, ANLS)*	88.4	90.1	88.8	—
	General Visual Question Answering VQAv2 (test)	75.2	78.1	—	—
Text	General MMLU (0-shot, CoT)	73.0	86.0	75.2 (5-shot)	82.0
	Math MATH (0-shot, CoT)	51.9	68.0	38.9	70.2
	Reasoning GPQA (0-shot, CoT)	32.8	46.7	33.3	40.2
	Multilingual MGSM (0-shot, CoT)	68.9	86.9	75.1	87.0

Рис. 12 Ефективність різних моделей

5 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

5.1 Результати

Аналіз був взятий за період з 01.03.2024 по 01.09.2024 (6 місяців).

Було обрано 10 youtube каналів різної величини на одну тематику – новинні канали.

Список каналів: 5 канал, BBC News Україна, Radio NV, Апостроф TV, Є ПИТАННЯ, Сейчас, Суспільне Новини, Телебачення Торонто, Українська правда, Цензор.НЕТ.

На перше що хотілося звернути увагу це на те що кількість підписників(рис. 13) не впливає на кількість переглядів на відео(рис. 14) але корелюється з загальним переглядами(рис. 15).

Пов'язано це з тим що великі канали мають змогу випускати велику кількість відео. Також це пов'язано з тим що малі канали можуть обирати іншу стратегію для просування відео збільшивши якість відео як, наприклад Телебачення Торонто яке має в середньому більше 250 тис. переглядів\відео. Так як найбільший канал в цьому списку 5 канал має лише 50 тис. переглядів\відео, но із за обсягів 5 канал найбільший за загальним кількістю переглядів.



Рис. 13 Графік підписників на каналі



Рис. 14 Середня кількість переглядів

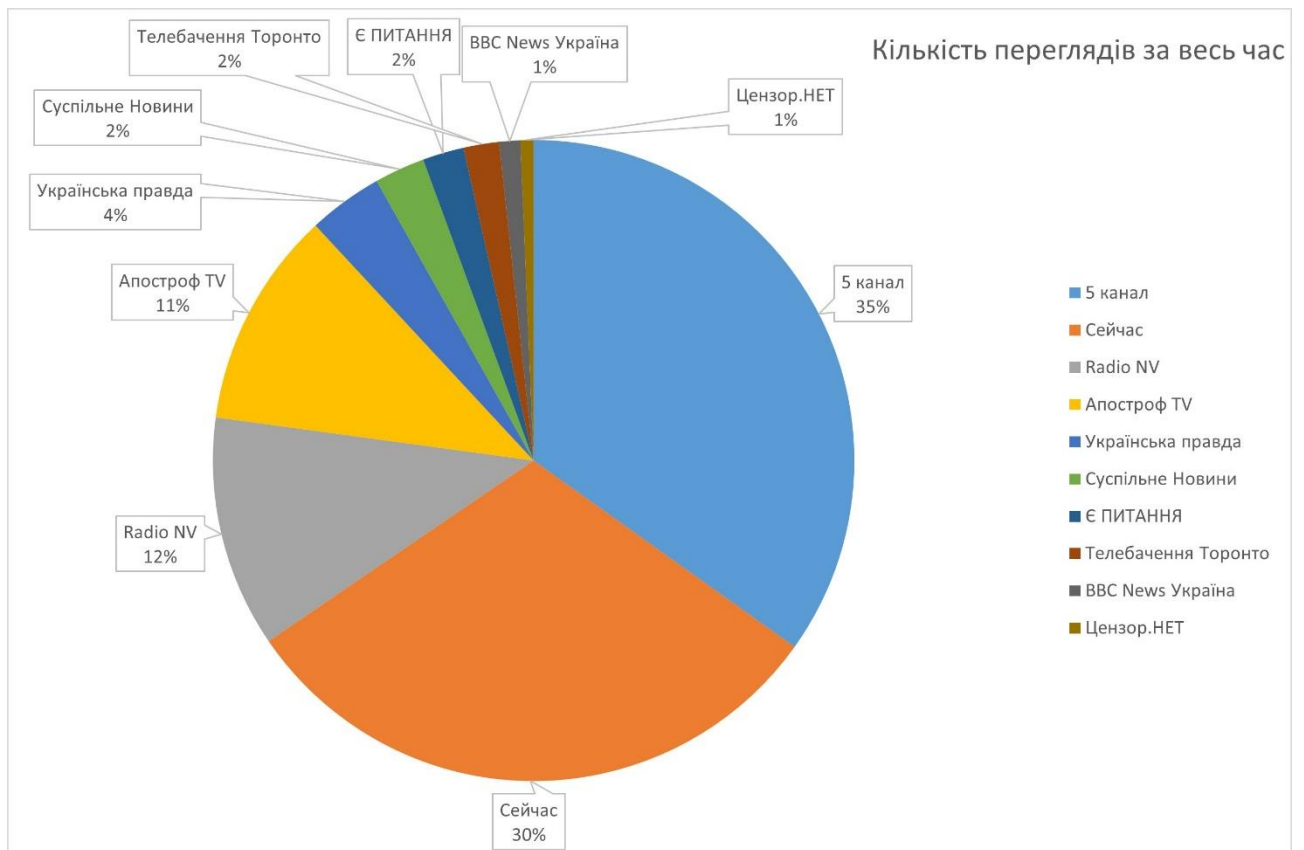


Рис. 15 Кількість переглядів за весь час

Наступне що хотілося проаналізувати це вплив емої в назві(рис. 16).

Було розбито на 3 категорії відео це не популярні відео(рис. 17), популярні відео(рис. 18) та дуже популярні відео(рис. 19). *І як видно в популярних і не популярних відео використання емої майже не дає ніякого результату, а ось в відео які набирають дуже багато переглядів частіше(72%) це ті які не використовують в назві емої, що показує про негативний вплив емої.*



Рис. 18 Популярні відео



Рис. 97 Дуже популярні відео

При аналізі впливу довжини відео до його популярності показало, **що чим довше відео тим воно популярніше і ця залежність майже лінійна**(рис. 20).

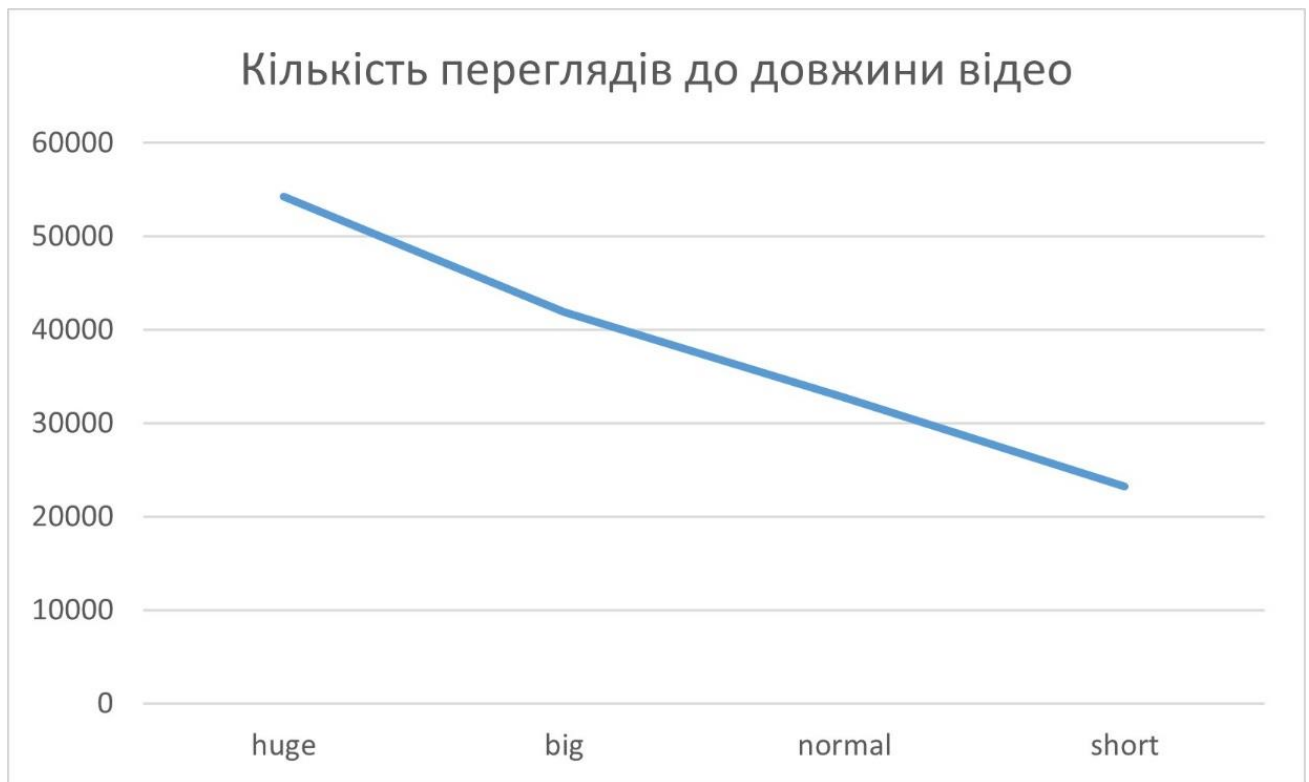


Рис. 20 Кількість переглядів до довжини відео

Після аналізу рис 21 та рис 22 можна надати такі рекомендації для різних каналів

Канал Сейчас видно з рис 22 що короткі відео мають непогану популярність, но з рис 20 видно що канал майже не знімають такого типу відео.

Канал Radio NV з рис 22 середні та короткі відео мають непогану ефективність. Також канал випуска переважно величезні відео, хоча відео класу великі набирають більше переглядів.

В інших каналах переважно рівномірна стратегія, яка не потребує змін.

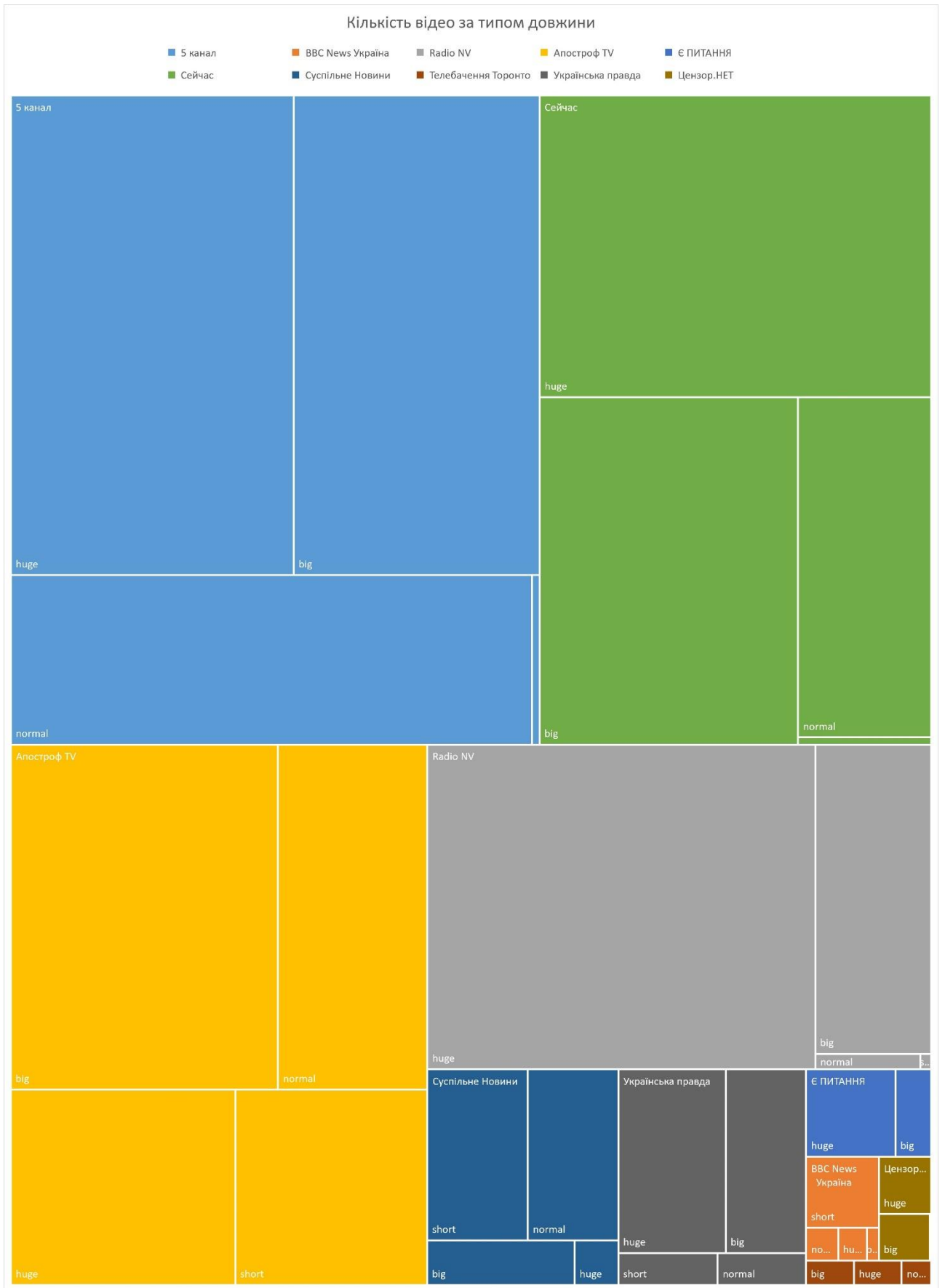


Рис. 21 Кількість відео за типом довжини

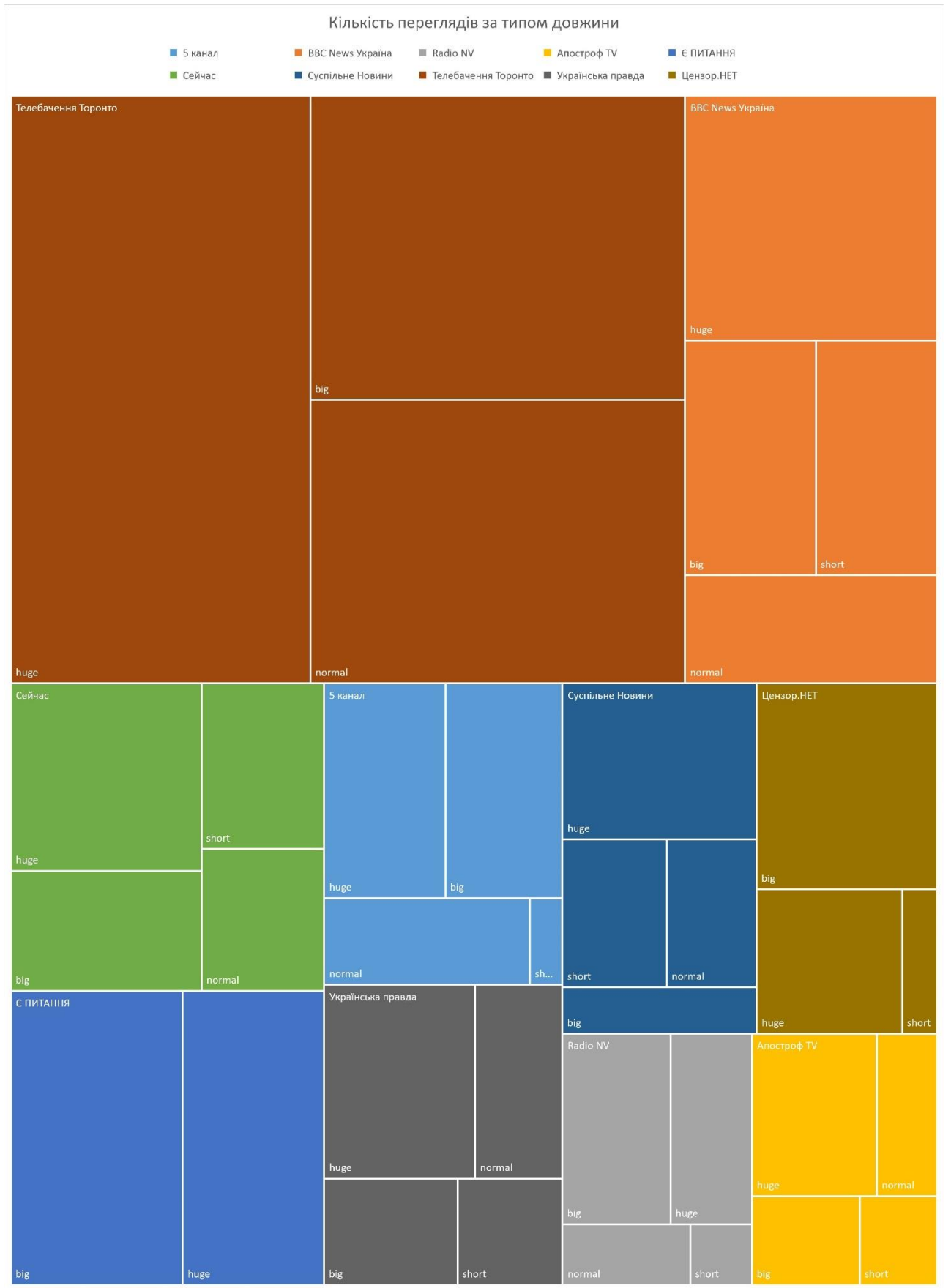


Рис. 22 Кількість переглядів за типом довжини

5.2 Апаратні та програмні вимоги

Апаратні та програмні вимоги до комп'ютера:

Мінімальні:

Операційна система: рекомендовано Windows 10 64 bit або вище.

Процесор: процесор або система на кристалі (SoC) з 4 ядрами та частотою 4 ГГц чи вищою.

Оперативна пам'ять: 8 ГБ

Відеокарта: з 4 Гб

Рекомендується:Ск

Операційна система: рекомендовано Windows 10 64 bit або вище.

Процесор: процесор або система на кристалі (SoC) з 8 ядрами та частотою 4 ГГц чи вищою.

Оперативна пам'ять: 16 ГБ

Відеокарта: з 8 Гб

ВИСНОВОК

У магістерській роботі було розроблено програмне забезпечення для аналізу популярності відео в інтернеті, яке:

- Збирає дані з обраних YouTube-каналів (кількість переглядів, лайків, коментарів, використання емодзі в назвах відео, тривалість відео тощо).
- Автоматизує обробку зібраних даних, забезпечуючи виявлення закономірностей та створення звітів.
- Візуалізує результати аналізу для кращого розуміння впливу різних факторів на популярність відео.

Результати дослідження:

Кількість підписників каналу корелює із загальною кількістю переглядів, але не впливає на середню популярність окремих відео. Великі канали мають перевагу завдяки великій кількості контенту, тоді як малі канали можуть досягати високих середніх переглядів завдяки якості відео.

Використання емодзі в назвах відео не сприяє популярності. Навпаки, для дуже популярних відео зафіксовано, що 72% з них не використовують емодзі в назвах, що свідчить про можливий негативний вплив.

Довжина відео прямо пропорційно впливає на популярність: чим довше відео, тим більше переглядів воно набирає, причому ця залежність є майже лінійною.

Для каналів з різними стратегіями випуску відео було виявлено специфічні рекомендації. Наприклад:

- Канал Сейчас може зосередитися на коротких відео, які вже демонструють непогану ефективність.

- Канал Radio NV повинен приділити більше уваги середнім та коротким відео, які мають вищу ефективність у порівнянні з величезними роликами.
- Інші канали демонструють стабільну стратегію без необхідності змін.

Рекомендації щодо можливості подальшого використання отриманих результатів:

- Розроблене програмне забезпечення може бути адаптоване для аналізу інших тематик або платформ, наприклад, TikTok чи Instagram.

Таким чином, результати дослідження досягли поставленої мети, забезпечуючи інструменти для глибокого аналізу популярності відео в інтернеті та надаючи практичні рекомендації для оптимізації контент-стратегії.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Llama 3.2: революційний передовий ШІ та бачення з відкритими настроюваними моделями. URL: <http://1.https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/>.
2. Microsoft SQL Server. URL: https://uk.wikipedia.org/wiki/Microsoft_SQL_Server
3. Pandas. URL: <https://pandas.pydata.org/about/>
4. Matplotlib: Візуалізація за допомогою Python. URL: <https://matplotlib.org/>
5. Сучасна та настроювана бібліотека інтерфейсу користувача Python на основі Tkinter. URL: <https://customtkinter.tomschimansky.com/>
6. Selenium. URL: <https://www.selenium.dev/documentation/webdriver/>
7. Маніфест GraphRAG: додавання знань до GenAI. URL: <https://neo4j.com/blog/graphrag-manifesto/>
8. Ласкаво просимо до GraphRAG. URL: <https://microsoft.github.io/graphrag/>