Специфікація програмного забезпечення

СПЕЦИФІКАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

«Дослідження застосування моделей графового представлення

знань для задач відповіді на запитання стосовно зображення»

Автор:

ст. гр. ІПЗм-23-2

Тєльний Максим

Харків, 2025

**ВСТУП**

* 1. Огляд програмного забезпечення

Сучасні системи штучного інтелекту для відповідей на запитання щодо зображень (Visual Question Answering – VQA) часто стикаються з труднощами у випадках, коли відповідь вимагає додаткової контекстної інформації, яка виходить за рамки даних, безпосередньо доступних на зображенні чи в тексті запитання. Це означає, що ефективність таких систем значною мірою залежить від можливості інтегрувати зовнішні джерела знань, які містять важливі семантичні зв'язки між об’єктами.

Більшість наявних реалізацій VQA орієнтовані на завдання класифікації, розпізнавання або простого опису зображень. Проте їхні можливості обмежені, якщо запитання передбачає складніші аналітичні дії, пов’язані з додатковою інформацією про об’єкти та їхні взаємозв’язки. Як наслідок, виникають ситуації, коли модель не може дати правильну відповідь через недостатність контекстних даних.

З огляду на це є актуальною потреба створення програмного рішення, здатного:

– адаптивно інтегрувати зовнішні графи знань залежно від конкретного змісту запитання та розпізнаних на зображенні об’єктів;

– швидко та ефективно виконувати фільтрацію отриманої інформації з графа знань, що дозволить запобігти надлишковому навантаженню моделі непотрібними фактами;

– використовувати переваги сучасних трансформерних моделей для мультимодальної інтеграції графових та візуально-текстових даних;

– забезпечити збереження високої продуктивності та можливість легкого масштабування.

Система, запропонована в рамках цієї роботи, орієнтована саме на вирішення цих завдань, що дозволить суттєво підвищити точність і якість формування відповідей у VQA-системах, забезпечивши стабільну та ефективну роботу у різних практичних сценаріях застосування.

* 1. Мета

Метою цієї кваліфікаційної роботи є дослідження ефективності інтеграції графів знань у задачі автоматизованих відповідей на запитання щодо зображень (Visual Question Answering) та розробка програмного рішення, здатного динамічно адаптуватися до особливостей конкретних запитань і контексту зображень. Головним завданням є створення системи, яка забезпечує точні й змістовні відповіді завдяки використанню додаткових семантичних знань.

Для досягнення зазначеної мети визначені такі задачі:

– здійснити аналіз сучасних методів і алгоритмів, що використовуються у сфері відповідей на запитання щодо зображень, із акцентом на застосуванні зовнішніх графів знань (Knowledge Graphs);

– сформулювати вимоги до програмного рішення, що здатне інтегрувати графи знань із нейромережевими моделями, визначивши основні критерії вибору релевантних фактів і тріплетів з графа;

– розробити модуль, який забезпечує швидке вилучення та фільтрацію інформації з графа знань у залежності від результатів попереднього аналізу зображення і тексту запитання;

– реалізувати мультимодальний механізм, який ефективно поєднує інформацію з графів знань та мультимодальні ознаки (візуальні та текстові) з використанням сучасних трансформерних архітектур;

– провести експериментальну оцінку розробленої системи за показниками точності відповідей, швидкості генерації відповідей, а також здатності системи справлятися зі складними контекстними запитаннями, що вимагають додаткових знань.

Виконання поставлених завдань дозволить створити універсальну та практично застосовну систему, що може бути легко інтегрована у різні програмні середовища і значно покращує загальну якість та пояснювальність відповідей у мультимодальних сценаріях.

# **2 ЗАГАЛЬНИЙ ОПИС**

2.1. Перспективи програмного забезпечення

Розвиток систем, здатних автоматично генерувати відповіді на запитання щодо змісту зображень (Visual Question Answering – VQA), є важливим напрямком у сфері штучного інтелекту. Водночас зростає потреба у більш точних та семантично обґрунтованих відповідях, що враховують додатковий контекст, не представлений прямо у візуальних чи текстових даних. У цьому аспекті інтеграція графів знань із нейромережевими моделями відкриває нові перспективи для глибшого розуміння змісту зображень та контекстуальних зв’язків.

Використання системи, що адаптивно залучає графи знань залежно від типу запитання та результатів обробки зображень, дозволяє суттєво підвищити якість відповідей, особливо в умовах, коли потрібна інформація відсутня безпосередньо на зображенні або є недостатньо зрозумілою для моделі. Такий підхід значно зменшує кількість помилкових відповідей, покращує стійкість системи до складних контекстних ситуацій та оптимізує процес взаємодії між нейромережами і зовнішніми базами знань.

Завдяки гнучкій модульній архітектурі запропоноване рішення легко адаптується до нових завдань і контекстів, таких як додавання альтернативних графових баз знань або інтеграція з іншими моделями генерації відповідей. Крім того, система підтримує ефективне масштабування без значних змін архітектури чи коду, що дозволяє її застосовувати у різноманітних галузях — від медицини до освітніх і комерційних платформ.

Подальше вдосконалення програмного забезпечення може передбачати інтеграцію додаткових інструментів моніторингу та аналізу роботи системи, що дасть змогу автоматично оцінювати точність і якість відповідей, а також забезпечити оперативне реагування на зміну зовнішніх умов і внутрішніх показників моделі.

2.2. Функції програмного забезпечення

На поточному етапі розробки програмного рішення передбачено реалізацію таких основних функцій системи:

– модуль розпізнавання об’єктів на зображеннях, що дозволяє оперативно виявляти й ідентифікувати ключові компоненти сцени з високою точністю;

– компонент отримання зовнішньої інформації з графа знань (Wikidata), який забезпечує швидке вилучення необхідних семантичних тріплетів, що мають безпосереднє відношення до запитання та виявлених об'єктів;

– механізм адаптивної фільтрації отриманих із графа знань фактів, що мінімізує ризик перевантаження системи зайвими або нерелевантними даними;

– інтеграція зовнішніх знань із мультимодальними даними (текстовими та візуальними) через спеціалізований модуль на основі трансформерних архітектур, що забезпечує формування точних та контекстуально обґрунтованих відповідей;

– реалізація системи динамічного прийняття рішення про вибір найважливіших знань із графа з використанням рейтингового механізму на основі релевантності та значущості тріплетів;

– налаштування параметрів роботи модулів (наприклад, поріг впевненості для розпізнавання об’єктів, обсяг фактів, що вилучаються з графа знань), дозволяючи адаптувати програмне забезпечення під конкретні завдання й умови експлуатації.

Запланований функціонал створить ефективну основу для автоматичної обробки мультимодальних даних із залученням зовнішніх знань, яка здатна стабільно й точно працювати без необхідності постійного втручання користувача, навіть у складних та змінних сценаріях використання.

2.2. Залежності програмного забезпечення

Для повноцінного функціонування розробленої системи необхідне встановлення низки бібліотек і пакетів, які забезпечують коректну роботу її модулів. Серед них основними є:

– PyTorch: платформа для побудови й навчання глибинних нейронних мереж, яка є необхідною для роботи трансформерних моделей.

– Transformers (від Hugging Face): бібліотека, яка забезпечує доступ до великої кількості попередньо навчених нейромережевих моделей, таких як T5, які використовуються для генерації відповідей.

– Ultralytics YOLOv8: сучасний інструмент для швидкого й ефективного розпізнавання об’єктів на зображеннях.

– Requests: пакет, який використовується для взаємодії з API зовнішньої графової бази знань Wikidata.

– SPARQLWrapper: інструмент, що дозволяє формувати та виконувати SPARQL-запити до бази знань.

– Pandas та NumPy: бібліотеки для обробки й аналізу даних, що використовуються для маніпуляцій та фільтрації отриманих фактів із графів знань.

Дані залежності мають бути встановлені у віртуальному середовищі Python, що дозволяє уникнути конфліктів версій і забезпечує стабільність роботи програми. Всі залежності системи перелічені у файлі requirements.txt, що спрощує процес встановлення та налаштування програмного забезпечення.

2.3. Використання програмного забезпечення

Розроблене програмне забезпечення призначене для автоматичного надання відповідей на запитання, що стосуються змісту зображень, із використанням інтегрованих зовнішніх знань з графа Wikidata.

Процес взаємодії користувача із системою передбачає такі основні етапи:

1. Підготовка вхідних даних: користувач надає системі зображення (у форматі JPEG або PNG) та формулює текстове запитання щодо змісту цього зображення.
2. Запуск системи: користувач запускає програму через командний рядок, передаючи шлях до зображення та текст запитання як аргументи.
3. Автоматичний аналіз:

* Виконується розпізнавання об’єктів на зображенні з використанням YOLOv8.
* Система автоматично запитує відповідні семантичні знання із Wikidata та фільтрує їх за релевантністю до заданого запитання.
* Генерується відповідь за допомогою трансформерної мовної моделі, що поєднує отримані знання з мультимодальними ознаками.

1. Отримання результату: після завершення роботи, програма виводить сформульовану відповідь безпосередньо в командний рядок, що дозволяє користувачу легко й оперативно отримати необхідну інформацію.

Розроблене рішення може бути використано у різних прикладних сферах, включаючи освітні платформи, автоматизовані системи аналізу медичних зображень, інтелектуальні чат-боти, інтерактивні музейні додатки та інші проєкти, де важлива якісна та контекстно обґрунтована обробка візуальної інформації.

2.5. Вимоги до середовища

Для коректної та стабільної роботи програмного забезпечення необхідно забезпечити відповідність наступним вимогам:

Апаратні вимоги:

* Процесор: Intel Core i5 або аналогічний за продуктивністю.
* Обсяг оперативної пам’яті: не менше 8 ГБ (рекомендовано 16 ГБ).
* Відеокарта: NVIDIA GPU з підтримкою технології CUDA (рекомендується для прискорення роботи нейромережевих моделей).
* Обсяг вільного місця на диску: не менше 10 ГБ для зберігання моделі, графових даних та бібліотек.

Програмні вимоги:

* Операційна система: Windows 10 і новіші, Ubuntu 20.04 LTS і новіші, або будь-яка сумісна з Python ОС.
* Інтерпретатор мови програмування Python версії 3.8 і вище.
* Система керування пакетами: pip.
* Бібліотеки й пакети Python відповідно до файлу залежностей (requirements.txt).

Мережеві вимоги:

* Постійний доступ до мережі Інтернет для запитів до зовнішнього графа знань Wikidata.

Дотримання зазначених вимог дозволить забезпечити високу продуктивність, стабільність роботи, а також оптимальну швидкість виконання завдань програмного забезпечення.