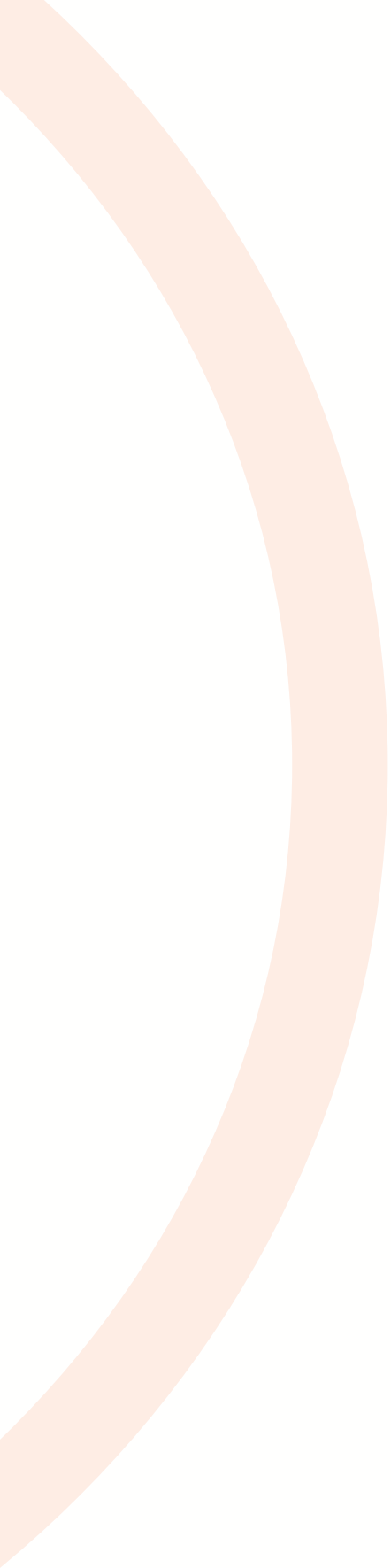
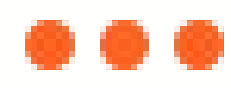




АВТОМАТИЗОВАНЕ НАПОВНЕННЯ ГРАФІВ ЗНАНЬ З НЕСТРУКТУРОВАНИХ ТЕКСТІВ

Виконав: Анісімов Є.О.

Керівник: Андрощук М.В.



Актуальність теми



- **Експоненційне зростання обсягів неструктурованих текстових даних.**
- **Висока вартість та складність ручного вилучення та структурування знань.**
- **Нагальна потреба в ефективних автоматизованих методах побудови графів знань.**

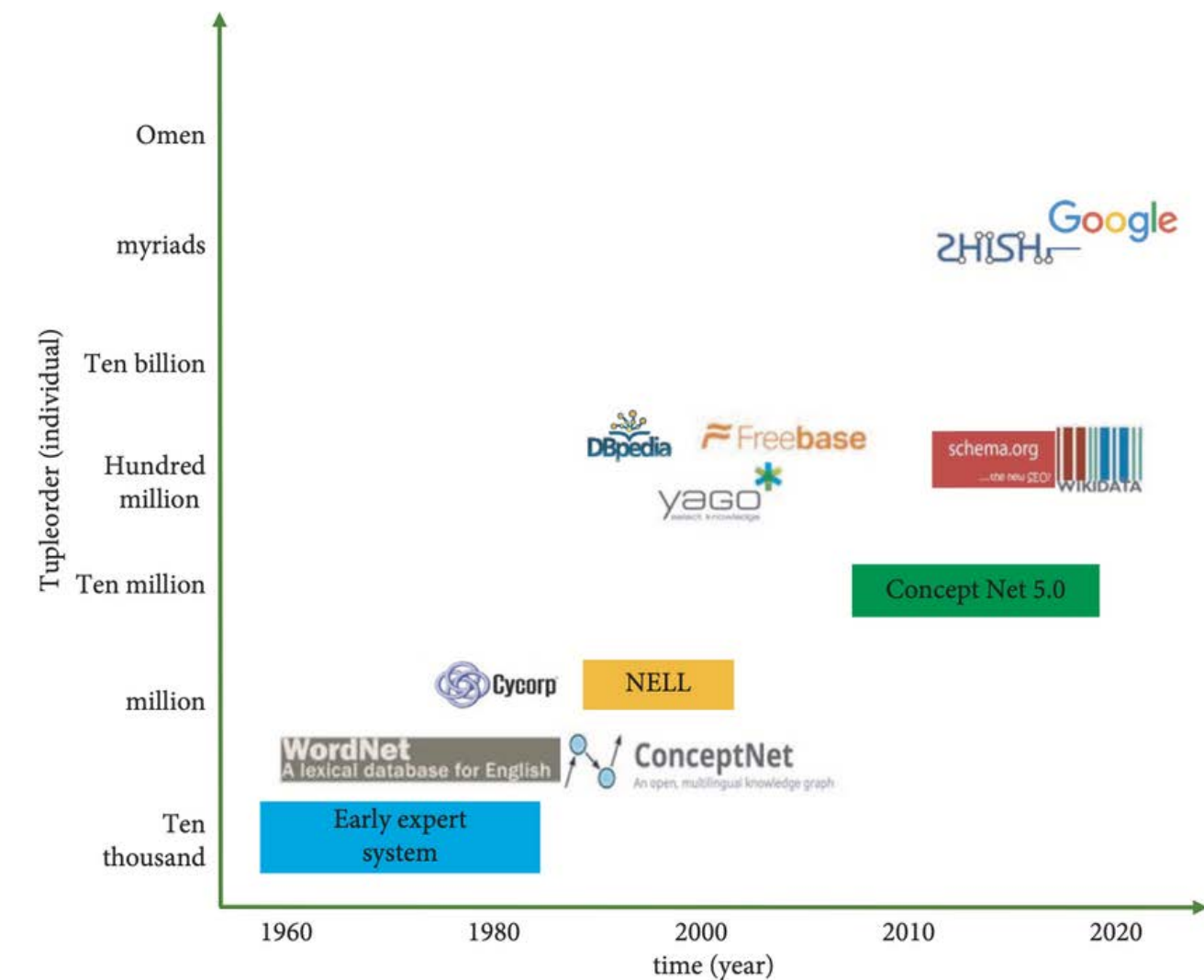
Мета



Розробка, реалізація та оцінювання програмної системи для автоматизованого наповнення графів знань з неструктурованих текстів з підтримкою функціональності відповідей на запитання.

Графи Знань: Еволюція та Потреба в Автоматизації

- Ранні експертні системи: ручні правила, обмежений масштаб.
- Семантичні мережі, онтології, перші графи знань.
- Сучасні графи знань: DBpedia, Google KG.

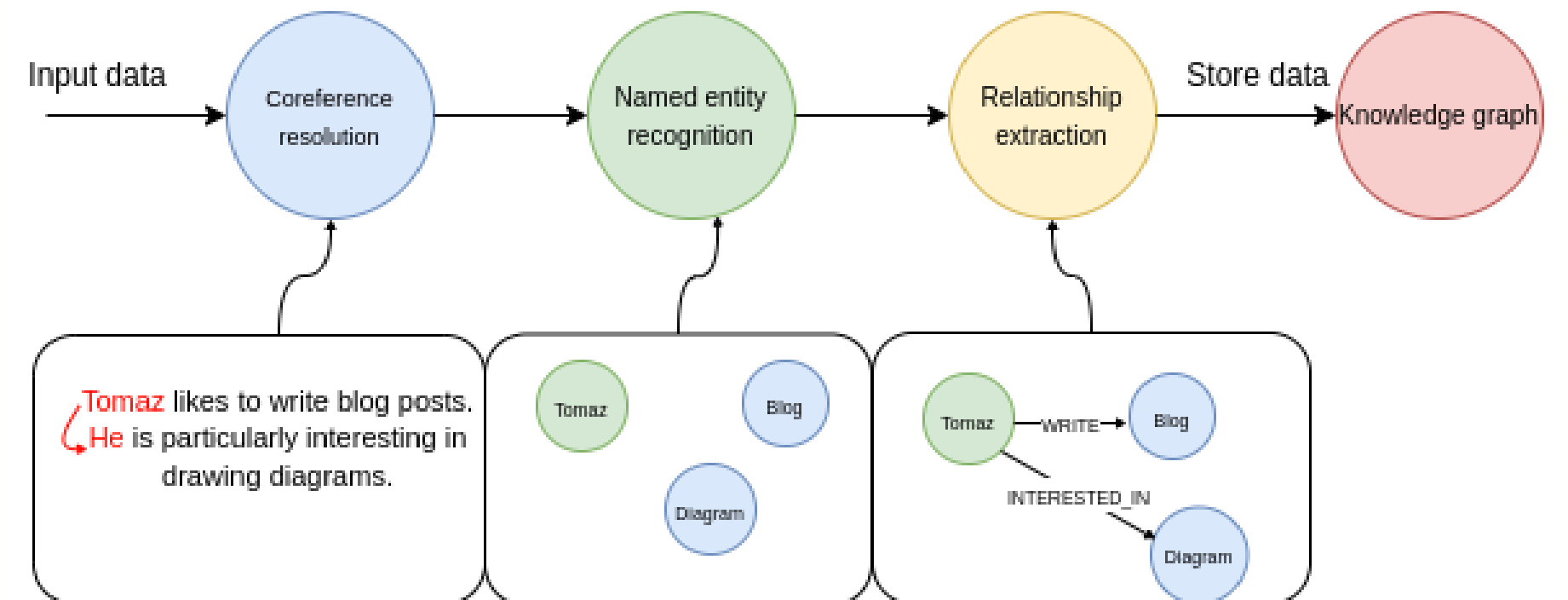


Етапи розвитку графів

знань

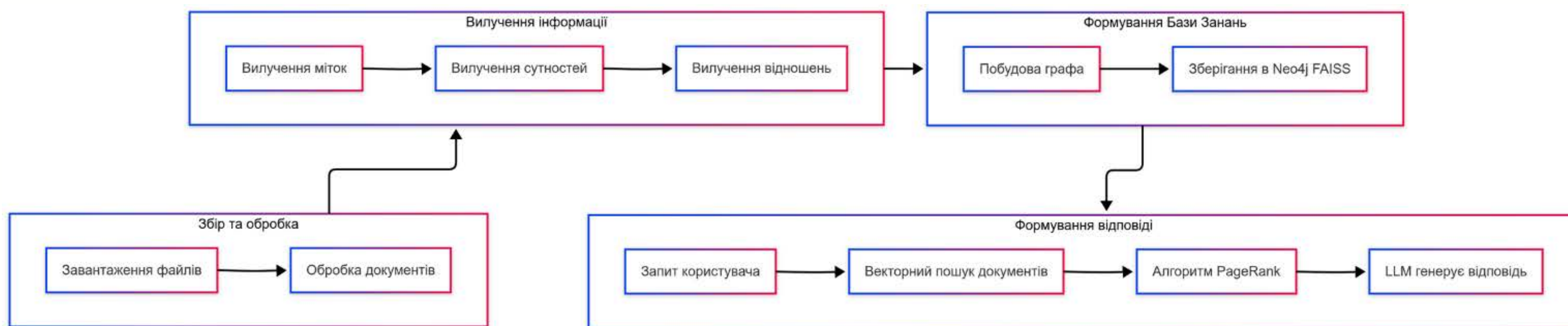
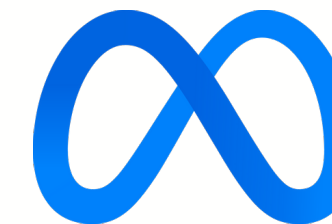
Технології для Автоматичного Створення Графів Знань

- Розпізнавання іменованих сутностей.
- Вилучення відношень.
- Роль LLM та спеціалізованих моделей на етапах ІЕ.

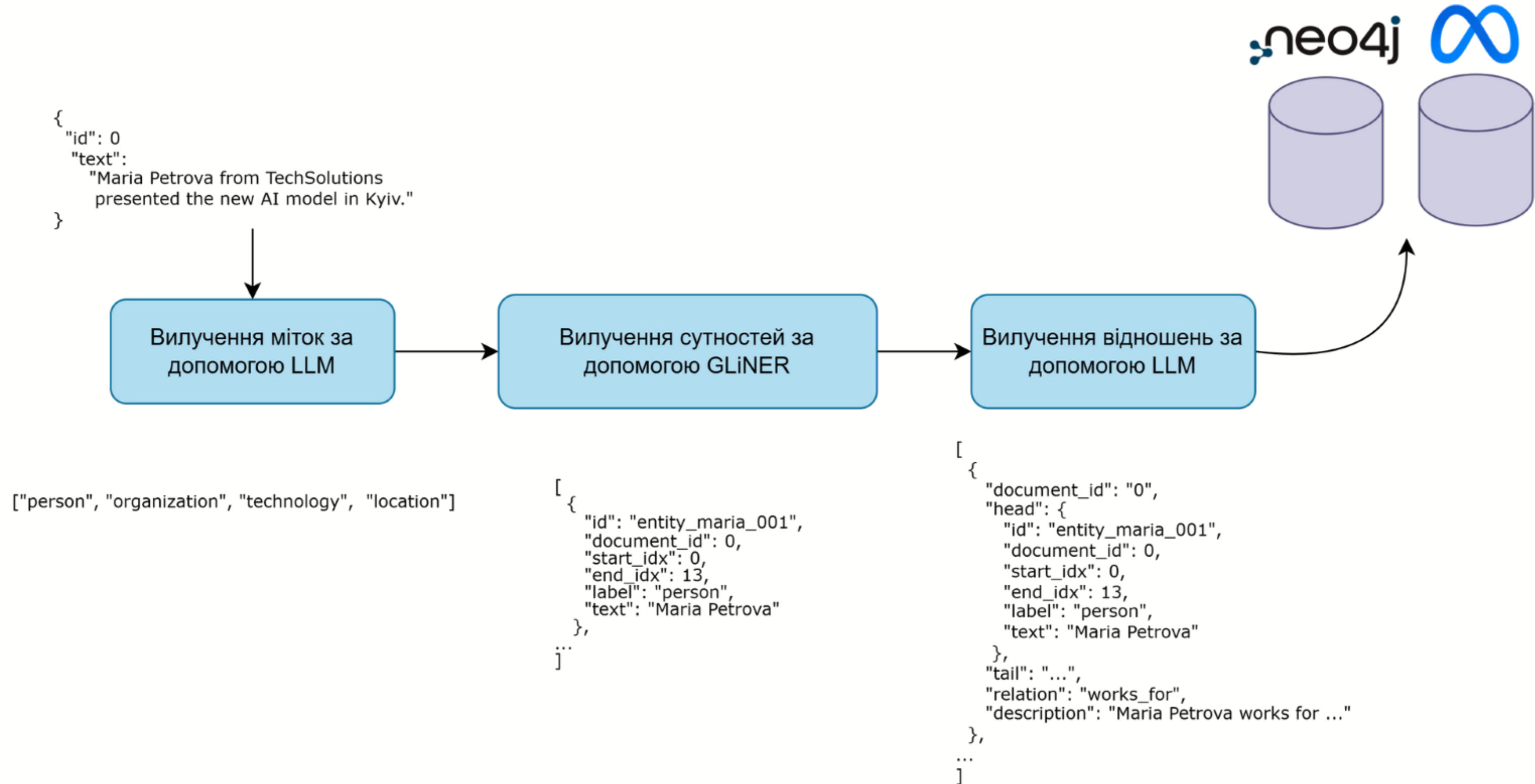


Етапи інформаційного вилучення

Архітектура системи



Етап Вилучення Інформації



Интерфейс системы

Configuration Settings

GLiNER Model

urchade/gliner_large-v2.1

Predefined GLiNER Labels

Sample Size NER Labels

50 - +

NER Threshold

0.50

0.00 1.00

Synonym Threshold

0.80

0.00 1.00

Knowledge Graph Generation and Question Answering

Document Processing Question Answering

Document Processing

Upload a file to begin processing.

Upload a text file

Drag and drop file here
Limit 200MB per file • TXT, JSONL, JSON

Browse files

Enter an optional query for processing:

Leave empty to just process the document

Process Document

Clear Neo4j Database

Insert Data into Neo4j

Интерфейс системы

Configuration Settings

GLiNER Model

urchade/gliner_large-v2.1

Predefined GLiNER Labels

Sample Size NER Labels

50 - +

NER Threshold

0.50

0.00 1.00

Synonym Threshold

0.80

0.00 1.00

Knowledge Graph Generation and Question Answering

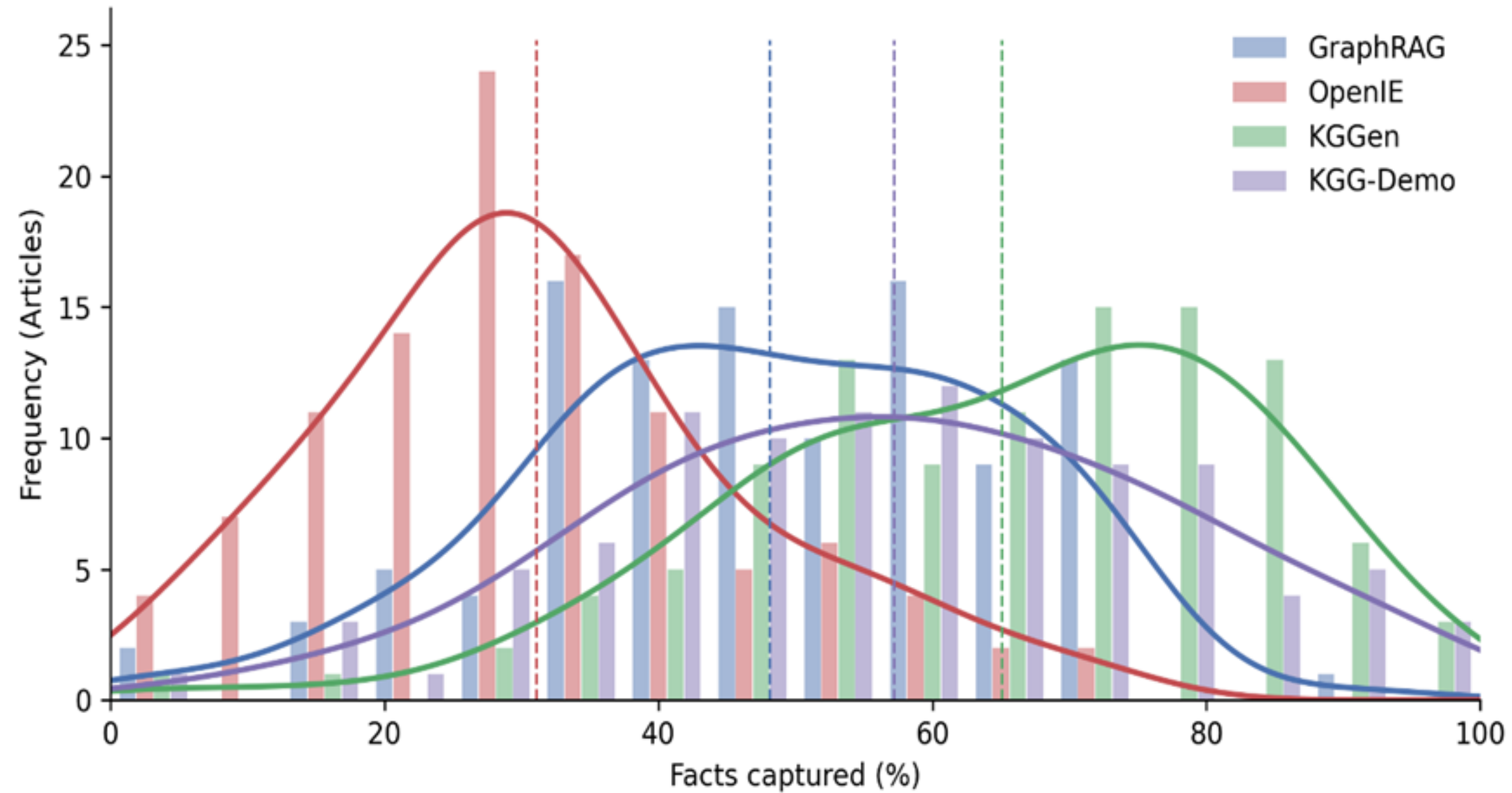
Document Processing Question Answering

Question Answering

Enter your question:

Get Answer

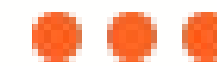
Тестування та оцінка ефективності



Висновки та напрямки розвитку



- Розроблено прототип системи для автоматизованого наповнення графів знань з текстів, використовуючи LLM.
- Система продемонструвала конкурентоспроможність на бенчмарку у вилученні фактів, підтвердивши практичну цінність.
- Подальші дослідження зосередяться на поглибленні лінгвістичної обробки, вдосконаленні вилучення, розширенні GQA та підвищенні продуктивності.



Дякую за увагу!