Софийски университет “Св. Климент Охридски”

Факултет по математика и информатика

**Курсов проект**

По "Семантичен уеб”

Трансформация на текст в граф

Изготвили:

Стоян Николов, Технологии за големи данни

Николай Шиваров, Технологии за големи данни

Нели Узунова, Изкуствен интелект

Съдържание:

[Въведение](#_faebrgoeoo7m)

[Основни функционалности](#_bsk2qv25oxco)

[1. Въвеждане на текст](#_oz9ja6js5aty)

[2. Избор на схема](#_b5y7h66x6oqz)

[3. Валидация на Turtle формат](#_i0lpsgbctz2d)

[4. Генериране на граф от знания](#_cxnfmi1jjlg)

[Архитектура на приложението](#_vlvkuvm41wgh)

[Ключови елементи на функционалността:](#_b0zxzejbv2j9)

[Примери за употреба](#_wojz4t4sfxvd)

[Заключение](#_2cwkl9nyspgw)

## Въведение

Целта на проекта е да се разработи софтуерно приложение, което автоматично трансформира неструктуриран текст в структуриран граф от знания, съобразно зададени схеми като schema.org или FHIR. Идеята е да структурира и визуализира текстова информация във формат, който улеснява разбирането на връзките и понятията в текста.

## Основни функционалности

#### 1. Въвеждане на текст

Потребителите могат да въведат текстова информация в системата. Този текст се обработва автоматично, за да се извлекат ключови понятия и връзки.

#### 2. Избор на схема

При въвеждането на текст потребителите могат да изберат между различни схеми: schema.org за общи данни или FHIR за медицински данни. Тези схеми определят как данните ще бъдат структурирани и представени.

#### 3. Валидация на Turtle формат

След като текстът е преобразуван в граф от знания, системата предоставя възможност за валидация на генерирания Turtle формат. Това осигурява, че данните са коректно структурирани според избраната схема и съответстват на спецификациите на Turtle формата.

#### 4. Генериране на граф от знания

След като текстът е валидиран системата автоматично генерира граф от знания, който визуализира връзките между различните понятия и термини в текста.

## Архитектура на приложението

Приложението е разделено на три основни компонента, които сътрудничат за постигане на целта на проекта:

* Модул за извличане на информация: Този модул използва техники за обработка на естествен език (Natural Language Processing - NLP), за да анализира входния текст и да извлича ключови концепции и връзки между тях.
* Модул за трансформация в граф от знания: След като информацията е извлечена, този модул я преобразува в структуриран граф от знания, като създава връзки и атрибути според зададената схема (например schema.org).
* Интерфейс за взаимодействие: Потребителският интерфейс позволява на потребителите да въвеждат текст и да получават резултатите от трансформацията в удобен за тях формат.

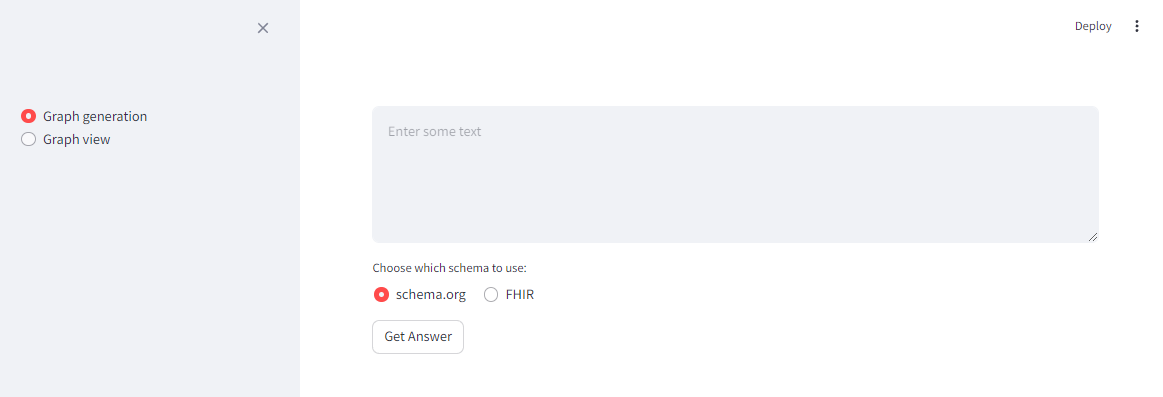
## Ключови елементи на функционалността:

1. Валидация на Turtle формат
2. Използване на LlamaAPI
   * LlamaAPI се използва за генериране на отговор в Turtle формат според зададена схема като schema.org или FHIR.
3. Генериране на граф от знания

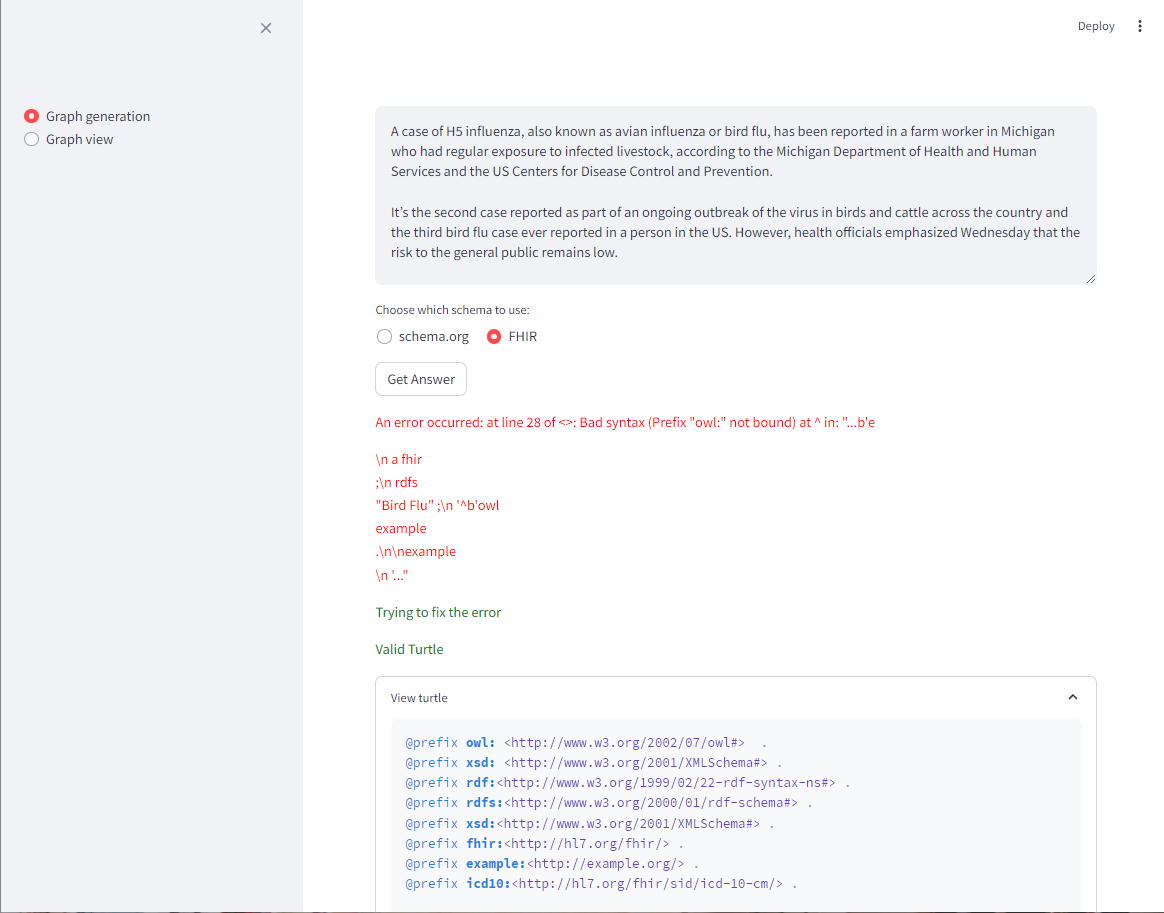
Тези ключови елементи предоставят възможност за валидация на входния текст в Turtle формат, използване на външно API за генериране на отговор в определена схема и визуализация на граф от знания чрез строене на възли и ребра в интерактивен графичен формат.

## Примери за употреба

Визията на потребителския интерфейс изглежда така:

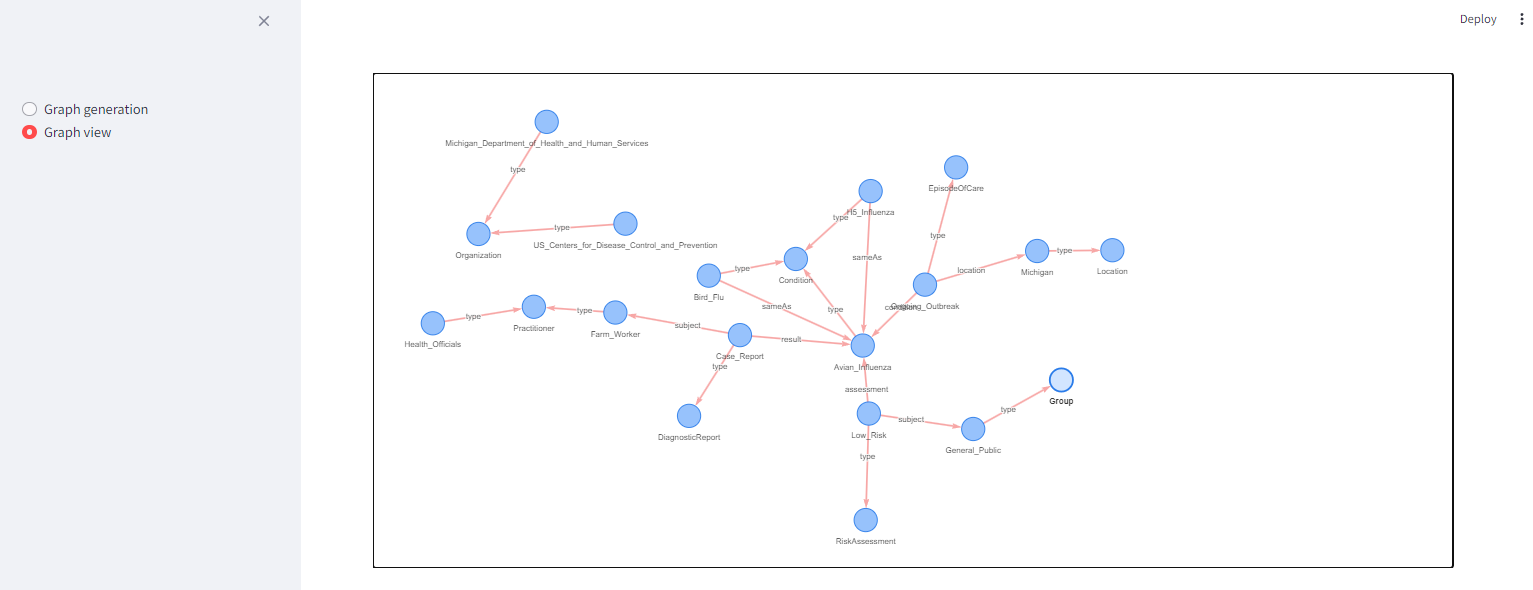


Потребителя може да въведе желания текст, и да избире схема между schema.org или FHIR и да види неговия текст превърнат във валиден turtle формат :





А във втория таб може да се види интерактивния граф, отговарящ на желания текст.



## Заключение

Проектът предоставя функционално приложение за автоматична трансформация на текст в граф от знания, използвайки зададена схема. Това улеснява структурираното представяне на текстови данни и подобрява анализирането и използването им в различни приложения.