

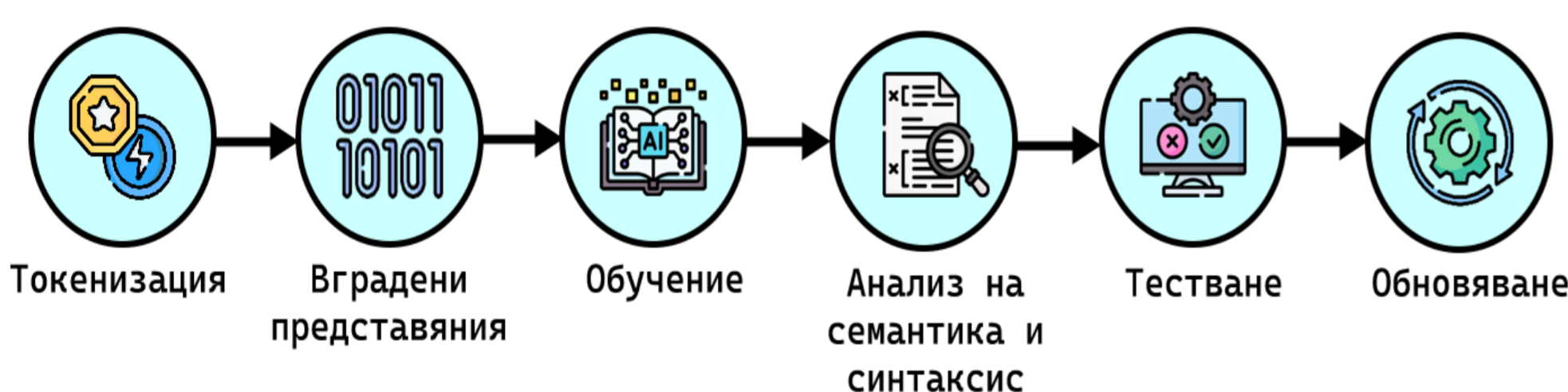
ФИТА

Проектиране и изграждане на система за текстово взаимодействие с потребителя

Ръководител на проекта: проф. д-р инж. Милена Карова, преподавател, КНТ
Пламена Атанасова Атанасова, студент, КНТ, председател на студентски клуб "Creative Code;"
Виктор Александров Иванов, студент, СИТ, заместник председател на студентски клуб "Creative Code;"
Божидар Иванов Иванов, студент, СИТ, член на студентски клуб "Creative Code;"
Калоян Панайотов Панайотов, студент, СИТ, член на студентски клуб "Creative Code;"

Въведение

Предложението за клубен проект се основава на разработването на интелигентна система за текстово взаимодействие с потребителя, която предоставя автоматизирани и точни отговори на въпроси. Изградената софтуерна инфраструктура включва база данни, сървърно приложение с API и мобилно приложение за Android. Потребителите изпращат заявки чрез мобилното приложение към API на сървъра, който генерира и връща отговори. Основната трудност е липсата на предварително обучени NLP модели за български език, което налага създаване на специален Data Set и нов модел за обработка на текст. Създаването на качествен набор от данни е трудоемко, включвайки събиране и структуриране на текстова информация на български. Обучението и оптимизацията на модела също са предизвикателство, тъй като системата трябва да разбира контекста и да генерира релевантни отговори. Осигуряването на добра интеграция между сървърното и мобилното приложение за бърз и сигурен обмен на данни е допълнително затруднение.



Фиг. 1 – Стъпки при изграждане на езиков модел

Методи за изграждане на системата

За подобряване на качеството на Data Set са използвани техники като текстово перифразиране, обратен превод, добавяне на случайни грешки. За изграждането на NLP модела са прилагани методи като токенизация, отбелязване на части от речта, разпознаване на именувани съществителни имена, анализ на настроението и класификация на текстове.

Заклучение

Текущата система предоставя удобен начин за достъп до информация относно Технически университет – Варна. Създаденият Data Set лесно може да се използва за обучение на бъдещи подобрени модели и да бъде надграден. Това поставя стабилна основа за бъдещо разширение и подобрения на системата, както и възможност за адаптация към различни потребителски нужди и технологични изисквания.

Резултати

Създаването на качествен Data Set с разнообразие и представителност на данните е осъществено чрез прилагането на различни техники. За постигане на по-голяма вариативност на текстовите примери са използвани следните два метода:

- текстово перифразиране - формулиране на нови изречения чрез замяна със синонимни думи, промяна на структурата, прилагане на граматически преобразувания;
- обратен превод - изреченията се превеждат на друг език и после обратно на оригиналния. Така се въвеждат вариации в структурата на изречението, избора на думи и формулировката;

Използвани са и техники като случайно вмъкване, изтриване и размяна на думи или фрази, както и добавяне на шум в текстовете чрез случайни грешки, за да се симулират сценарии от реалния свят и да се подобри устойчивостта на модела.

В изграждането на NLP модела са приложени следните техники:

- токенизация – текстът се разделя на по-малки единици като думи, изречения или части от думи. Улеснява се по-нататъшният анализ и обработка;
- отбелязване на части от речта (POS tagging) – присояват се граматични маркери на всяка дума в дадено изречение като съществително, глагол, прилагателно или наречие;
- разпознаване на именувани съществителни (NER) – идентифицира и класифицира именувани съществителни като имена на хора, организации, места, дати или други специфични лица в текста;
- анализ на настроението (Sentiment analysis) – определя изразеното мнение или настроение в даден текст дали е положително, отрицателно или неутрално;
- класификация на текст – категоризира даден текст в предварително определени класове и категории;

Диалоговата система е комбинирана, основана на правила и машинно обучение. В нея са използвани предварително дефинирани правила за покриване на специфични сценарии и алгоритми на машинното обучение. Това позволява на системата да се справя и с предвидими и с непредвидими въпроси.

Мобилното приложение предоставя удобен и интуитивен интерфейс, чрез който потребителите могат лесно да взаимодействат със системата. По този начин значително се подобрява достъпа до информация относно Технически университет – Варна.

Публикации по проекта

1. Карова М., Атанасова П., Иванов В., Проектиране и изграждане на контекстно ориентиран автоматизиран агент за диалог, СНС „Наука в бъдещето“, ТУ-Варна, 2024

Благодарности

Научните изследвания, резултати, от които са представени в настоящата публикация, са извършени по проекта в рамките на научноизследователската дейност на Технически университет – Варна, финансиран от държавния бюджет.