|  | **Софийски университет „Св. Кл. Охридски”**  Факултет по математика и информатика  *Катедра „Софтуерни технологии”* |  |
| --- | --- | --- |

**ПРЕДДИПЛОМЕН ПРОЕКТ**

на тема

„Създаване на многомодален и многоезичен набор от данни за автоматично отговаряне на въпроси, давани на ученици от различни нива на обучение”

Дипломант: **Симеон Емилов Христов**

Специалност: **Софтуерни технологии**

Факултетен номер: **6MI3400191**

Научни ръководители:

проф. д-р Преслав Наков, катедра „Обработка на естествен език“, Университет „Мохамед Бин Зайед по Изкуствен интелект“, Обединените Арабски Емирства

проф. д-р Иван Койчев, катедра „Софтуерни технологии“, ФМИ, СУ „Св. Климент Охридски“

Консултант: докт. Димитър Димитров, катедра „Софтуерни технологии“, ФМИ, СУ „Св. Климент Охридски“

София, 2024 г.

**Съдържание**

**Глава 1. Увод** (3-5стр.)

1.1. Актуалност на проблема и мотивация (0,5-1стр.)

1.2. Цел и задачи на дипломната работа (1-2стр.)

1.3. Очаквани ползи от реализацията (1-2стр.)

1.4. Структура на дипломната работа (0,5-1стр.)

Големите езикови модели набират популярност в различни аспекти на човешката дейност. Отличителна тяхна черта е даването на кратък, точен и релевантен отговор на широка гама въпроси. Многомодалните многоезични езикови модели са вид езикови модели, който може да се използва в образователната система за получаване на отговор на въпроси от затворен тип.

Настоящият преддипломен проект се фокусира върху създаването на набор от данни, включващ три типа въпроси, давани на ученици. Въпросите могат да включват изцяло текст, а могат да включват и картинки с текст. Възможно е картинката да бъде част от въпроса, но също е възможно и самите отговори на включват (понякога изцяло) картинки. Разглеждат се само въпроси със затворен отговор, в които има само един верен отговор. Резултатите от този преддипломен проект ще послужат като основа за дипломна работа, която има за цел да направи сравнителен анализ на свободно достъпни многомодални многоезични езикови модели.

Целта на преддипломния проект е да се създаде многомодален и многоезичен набор от данни за автоматично отговаряне на въпроси, давани на ученици от различни нива на обучение, и прилагане на основни езикови модели за получаване на отправна точка за бъдещи сравнения.

**Глава 2. Обзор на съществуващи подходи и формиране на множества от подобни типове множества от данни** (10-15стр.)

2.1. Основни дефиниции

2.2. Подходи, методи (евентуално модели и стандарти) за решаване на проблемите

2.3. Съществуващи решения (практически реализации)

2.4. Избор на критерии за сравнение и сравнителен анализ на решения/методи/стандарти/...

2.5. Изводи

**Глава 3. Използвани технологии, платформи и/или методологии (за практическото решаване на проблема)** (10-15стр.)

3.1. Изисквания към средствата (технологии, платформи и методологии)

3.2. Видове средства (технологии, платформи и методологии) и начин и място за използването им – сравненителен анализ

3.3. Избор на средствата (технологии, платформи и методологии)

3.4. Изводи

**Глава 4. Анализ** (10-15стр.)

4.1. Концептуален модел

4.2. Потребителски (функционални) изисквания (права, роли, статуси, диаграми, ...)

4.3. Качествени (нефункционални) изисквания (като напр. преносимост, използваемост, скалируемост, поддръжка, ...)

4.4. Работни (бизнес) процеси

4.5. Изводи

**Глава 5. Проектиране** (10-15стр.)

5.1. Обща архитектура – напр. слоеве, модули, блокове, компоненти...

5.2. Модел на данните (напр. база данни, файлова структура, ...)

5.3. Диаграми (на структура и поведение - по слоеве и модули, с извадки от кода)

5.4. Потребителски интерфейс (опционално)

5.5. Ресурсни и спомагателни модули (опционално)

**Глава 6. Реализация, тестване/експерименти и (евентуално) внедряване** (10-15стр.)

6.1. Реализация на модулите

6.2. Системна интеграция (опционално)

6.3. Планиране на тестването - тестови сценарии, процедури, ...

6.4. Модулно и системно тестване

6.5. Анализ на резултатите от тестването и начин на отразяването им

6.6. Експериментално внедряване (технологични изисквания, инсталиране, условия, използване, ...)

* Създаване на подходяща файлова структура, в която да са достъпни подбрани изпити (в pdf формат), давани на ученици в различни държави.
* Преобразуване на страниците на pdf файловете в картинки.
* Използване на инструмент за анотиране за очертаване на границите на въпросите и маркиране на всеки въпрос в един от два класа: само текст или текст и картинка.
* Изрязване на всеки въпрос и създаване на анотация (в json формат) за всеки pdf файл. Анотацията включва информация за предмета, по който се провежда изпита, както и данни за всеки въпрос (път до изрязания въпрос, верен отговор, номер на въпроса, и т.н.).
* Създаване на програмен код, който проверява и валидира формата (т.е. структурата) на всяка анотация.

Глава 7. Използване на Bard за получаване на отправна точка за бъдещи сравнения.

**Глава 7. Заключение** (1-2стр.)

7.1. Обобщение на изпълнението на началните цели

7.2. Насоки за бъдещо развитие и усъвършенстване

**Използвана литература** (min 10 литературни източника – статии, книги, с, форматирани съгласно MLA Style - <http://www.library.mun.ca/guides/howto/mla.php>)

**Приложения (опционално)**

**Приложение 1: Терминологичен речник (опционално)** (1-2стр.)

**Приложение 2, 3, ....:** напр. код или извадки от кода, наръчник на потребителя, екрани от потребителския интерфейс, спецификации, диаграми, и др.

**Изисквания за оформяне на дипломната работа:**

1. Това е препоръчителен шаблон, в зависимост от конкретното задание.
2. Йерархията на структуриране на съдържанието да не бъде повече от 3 нива, номерирани с арабски цифри – напр. 1.2.3.
3. Чуждестранните термини да бъдат преведени, а където това не е възможно – цитирани в *курсив* и нечленувани.
4. Страниците да бъдат номерирани с арабски цифри, в долния десен ъгъл.
5. Използваният шрифт за основния текст на описанието да бъде Times 12 или Arial 10, и Courier 9 за кода, с междуредие 16pt.
6. Да се избягват пренасянията на нова страница на заглавия на секции, фигури и таблици.
7. Да се избягват празни участъци на страници вследствие пренасянето на фигури на нова страница.
8. Всички фигури и таблици да бъдат номерирани и именовани (непосредствено след фигурата или таблицата).
9. Всички фигури и таблици да бъдат цитирани в текста.
10. Използваните фигури от други източници да бъдат цитирани.
11. Всички цитати да бъдат отразени в списъка на използваната литература.
12. Всички източници от списъка на използваната литература да бъдат цитирани в текста.
13. Използваната литература да се цитира съгласно MLA Style - <http://www.library.mun.ca/guides/howto/mla.php>