

Київський національний університет імені Тараса Шевченка

Факультет комп'ютерних наук та кібернетики

Кафедра дослідження операцій

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА
за спеціальністю 113 «Прикладна математика»

на тему:

Машинний переклад



Студентки 4-го курсу
Тутушкіної Анна Іванівної

Науковий керівник:
доцент, доктор фізико-математичних наук
Самойленко Ігор Валерійович

Робота заслухана на засіданні кафедри обчислювальної математики та
рекомендована до захисту в ЕК, протокол № ...9.... від ..23.05.... 2023 р.

Завідувач кафедри ДО  Іксанов Олександр Маратович

Київ – 2023

РЕФЕРАТ

Обсяг роботи 30 сторінок, 6 рисунків, 10 використаних джерел.

МАШИННИЙ ПЕРЕКЛАД, АВТОМАТИЗОВАНИЙ ПЕРЕКЛАД, ПРОГРАМА-ПЕРЕКЛАДАЧ З УКРАЇНСЬКОЇ НА КРИМСЬКОТАТАРСЬКУ, СТАТИСТИКА, ПРАВИЛА УКРАЇНСЬКОЇ МОВИ, ПРАВИЛА КРИМСЬКОТАТАРСЬКОЇ МОВИ.

Об'єкт розроблення програмного додатку: автоматизований переклад .

Мета роботи:

1. Дослідити основні алгоритми машинного перекладу.
2. Проаналізувати необхідні ресурси для створення кожного.
3. Зробити порівняльний аналіз української та кримськотатарської мов.
4. Ознайомитися з необхідними інструментами для програмної реалізації задачі.
5. Створити програму-перекладач на основі можливостей, отриманих в результаті аналізу необхідних та доступних ресурсів.

Методи та інструменти розроблення:

- Мова програмування Python.
- Програмне середовище PyCharm 2021.
- Фреймворк для розробки веб систем Django.
- Мова розмітки HTML.

Результати роботи: В ході роботи були дослідженні різні алгоритми машинного перекладу, їх сильні слабкі сторони та необхідні ресурси для їхнього створення. Створено програмне забезпечення для реалізації автоматизованого перекладу. Результатом роботи є застосунок, що здійснює автоматизований переклад, побудований на основі доступних ресурсів.

Отримані результати коректні і повністю відповідають темі кваліфікаційної роботи.

Інформація щодо впровадження: Даний застосунок можна використовувати як основу для створення більш точної та якісної програми-перекладача.

Сфера застосування: Здійснення автоматизованого перекладу, теоретичне застосування.

Зміст

РЕФЕРАТ.....	1
СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ.....	5
ВСТУП.....	6
РОЗДІЛ 1. ОСНОВНІ ПОНЯТТЯ ТА ТЕХНОЛОГІЇ МАШИННОГО ПЕРЕКЛАДУ.....	8
1.1 Визначення основних термінів і понять.....	8
1.2 Опис технологій, що використовуються в машинному перекладі.....	8
1.3 Проблеми машинного перекладу.....	9
РОЗДІЛ 2. АЛГОРИТМИ ТА МОДЕЛІ МАШИННОГО ПЕРЕКЛАДУ.....	11
2.1 Переклад на основі правил.....	11
2.2 Статистичний переклад.....	12
2.3 «Нейронний» машинний переклад.....	13
2.4 Гібридний машинний переклад.....	14
РОЗДІЛ 3. ПЕРЕКЛАДАЧ З УКРАЇНСЬКОЇ НА КРИМСЬКОТАТАРСЬКУ.....	16
3.1 Правила української мови.....	16
3.2 Правила кримськотатарської мови.....	17
3.3 Порівняльний аналіз.....	19
РОЗДІЛ 4. ОГЛЯД WEB-ФРЕЙМОРКУ DJANGO.....	21
4.1 Опис технологій Django.....	21
4.2 Загальна структура проекту на Django.....	22
РОЗДІЛ 5. ПРОГРАМНА РЕАЛІЗАЦІЯ.....	24
5.1 Структура web-застосунку.....	24
5.2 Вигляд готової програми.....	25
ВИСНОВОК.....	29
ДЖЕРЕЛА.....	30
ДОДАТОК.....	31

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ

МП – машинний переклад.

RNN (Recurrent neural network) – рекурентні нейронні мережі

SVO (subject, verb, object) – один з методів типологічної класифікації
мов.

SOV (subject, object, verb) – один з методів типологічної класифікації
мов.

Кримськотатарська – кримська.

ВСТУП

Оцінка сучасного стану об'єкта дослідження або розробки.

Машинний переклад, як галузь комп'ютерної лінгвістики продовжує активно розвиватися та вдосконалюватися. Існуючі програми-перекладачі навіть зараз пропонують доволі неточний та неякісний переклад, не дивлячись на всі сучасні технології та роботу спеціалістів. Також деякі мови, які менш популярні за китайську, англійську, іспанську тощо, потребують більшої уваги та роботи над перекладачами. Наприклад, кримськотатарська мова потребує особливого ставлення та більш кропіткої роботи при вивченні та дослідженні ніж, наприклад, українська, або турецька. Кримська мова не представлена в такому відомому перекладачі, як Google Translate, а також Bing Translator від Майкрософт, Translatedict тощо.

Актуальність роботи та підстави для її виконання. Основна мета роботи це дослідження можливостей і створення підґрунтя для повноцінної програми-перекладача з української на кримськотатарську, адже це мова ще одного корінного народу України. Кримськотатарська мова має бути представлена світові на одному рівні з українською, також мати свої доступні онлайн-ресурси для вивчення, мають існувати перекладачі, книжки, розмовники, школи, садочки, тощо.

Мета й завдання роботи. Дослідити різні типи та технології машинного перекладу для створення майбутньої програми-перекладача. Для цього необхідно:

1. Проаналізувати всі існуючі на даний момент типи.
2. Визначити їхні сильні та слабкі сторони.
3. Проаналізувати основні проблеми, що можуть виникнути під час розробки.
4. Створити ту модель машинного перекладу, на яку вистачає ресурсів.
5. Створити програму-перекладач.
6. Сформулювати необхідні бази даних.

Можливі сфери застосування. Машинний переклад активно використовується звичайними людьми як допомога при вивченні нової мови, при подорожах в інші країни, як підтримка людей-перекладачів.

РОЗДІЛ 1. ОСНОВНІ ПОНЯТТЯ ТА ТЕХНОЛОГІЇ МАШИННОГО ПЕРЕКЛАДУ

1.1 Визначення основних термінів і понять

Машинний переклад – це технологія автоматичного перекладу з однієї природної мови на іншу за допомогою різних алгоритмів і програм. Задачами машинного перекладу, його дослідженням та вдосконаленням займається така галузь мовознавства, як комп'ютерна лінгвістика.

На базовому рівні, суть даної технології полягає у створенні програми-перекладача, яка замінює слова та словосполучення вхідної мови на слова та словосполучення цільової.[1]

1.2 Опис технологій, що використовуються в машинному перекладі

У машинному перекладі використовуються різні технології, які допомагають здійснювати автоматичний переклад тексту з вхідної мови на цільову. Основні технології, що застосовуються у машинному перекладі, включають:

1. *Статистичні моделі.* Статистичні моделі використовуються для аналізу статистичних відповідностей між словами, фразами та іншими лінгвістичними одиницями в паралельних текстах. Вони забезпечують основу для вибору найбільш ймовірного перекладу на основі попередньо зібраних статистичних даних.

2. *Нейронні мережі.* Нейронні мережі, такі як рекурентні нейронні мережі (RNN) або трансформери, є ключовою технологією для нейронного перекладу. Вони використовуються для моделювання складних мовних відношень та здійснення перекладу на основі великої кількості паралельних текстів.

3. *Моделі засновані на увазі (Attention-based models).* Моделі з увагою (attention) використовуються для врахування контексту і уваги до різних

частин вихідного тексту під час перекладу. Вони надають важливість різним словам та фразам в процесі перекладу, що допомагає вирішити проблему неоднозначності та покращити якість перекладу.

4. *Застосування мовних ресурсів.* Машинний переклад використовує мовні ресурси, такі як словники, лексичні бази даних, граматичні моделі тощо для поліпшення якості перекладу. Ці ресурси забезпечують інформацію про лексичні відповідності, граматичні структури та інші мовні аспекти, що використовуються при перекладі.

5. *Автоматичне навчання.* Технології автоматичного навчання використовуються для тренування моделей машинного перекладу на великій кількості паралельних текстів. Це дозволяє моделям "навчитися" перекладати, адаптуватися до специфіки мови та поліпшувати свої навички з кожним новим тренуванням.

Ці технології можуть поєднуватися для забезпечення кращого результату перекладу.

1.3 Проблеми машинного перекладу

Головна проблема машинного перекладу полягає в тому, що мови мають складну семантику, граматику та культурні особливості, що часто ускладнює точний та вірний переклад. Основні проблеми, з якими стикаються системи машинного перекладу, це:

1. *Семантична неоднозначність.* Багатозначність слів та фраз може призводити до неправильного тлумачення та перекладу. Одне й те саме слово може мати різні значення в залежності від контексту, тому необхідно знайти вірне тлумачення залежно від контексту перекладу.

2. *Культурні відмінності.* Кожна мова має свою унікальну культурну специфіку та виразність. Деякі вислови, фрази або ідіоми можуть бути складно або навіть неможливо перекласти без врахування контексту і культурних нюансів.

3. *Граматичні відмінності*. Мови можуть мати різні граматичні структури та порядок слів. Трансформація речень з однієї мови на іншу може вимагати значних зусиль для забезпечення граматично правильного перекладу.

4. *Відсутність контексту*. Під час перекладу важливо враховувати контекст вихідного тексту, але інформація про контекст не завжди доступна. Це може призводити до неточностей або неправильного розуміння того, що необхідно перекласти.

5. *Різні стилі та регіональні варіанти (діалекти)*. Мови можуть мати різні стилі мовлення та регіональні варіанти, що ускладнює вибір правильного перекладу для конкретного контексту.

Хоча системи машинного перекладу продовжують розвиватися та поліпшуватися, вирішення цих проблем є складним завданням, оскільки вони вимагають глибокого розуміння мови та контексту, а також врахування культурних та лінгвістичних особливостей.

РОЗДІЛ 2. АЛГОРИТМИ ТА МОДЕЛІ МАШИННОГО ПЕРЕКЛАДУ.

2.1 Переклад на основі правил

МП на основі правил, або Rule-based MT є одним з класичних методів підходу до автоматичного перекладу тексту. Цей метод будує переклад, здійснюючи морфологічний, граматичний та семантичний аналіз вхідної та цільової мов.

Виділяють три основні типи МП на основі правил:

1. *МП на основі трансформації.* Один з найпопулярніших методів МП. При цьому методі перекладу виконуються три кроки: граматичний, морфологічний та семантичний аналіз вхідного тексту, трансформування отриманої структури в відповідну структуру цільової мови і генерація перекладеного тексту.

2. *МП на основі словників.* Цей метод перекладає слово за словом без особливого аналізу зв'язку між ними. МП на основі словників, не дивлячись на свою велику неточність, ідеально підходить для перекладу довгих списків фраз на рівні підречення, тобто каталоги продуктів, прості описи, тощо.

3. *МП з використання проміжної мови (інтерлінгви).* Цей метод передбачає перетворення вихідного тексту у форму інтерлінгви – абстрактне представлення, що не залежить від конкретної мови. Потім цільова мова генерується на основі цієї інтерлінгви.

Найголовнішою перевагою цього типу машинного перекладу є точність та контрольованість. Є можливість вручну коригувати та покращувати якість перекладу. Це може бути корисним в різних галузях, де важлива термінологічна точність.

Найголовніший недолік це те, що створення та підтримка великої кількості правил вимагає значних зусиль та великої роботи спеціалістів. Багатозначність та нюанси мови можуть бути складними для врахування за

допомогою правил. Крім того, цей підхід може бути менш гнучким і неадаптивним до нових типів текстів або змін у мовному вжитку.

2.2 Статистичний переклад

Статистичний переклад, або Statistical Translation базується на використанні статистичних моделей, які побудовані на паралельних текстах. Цей метод являється другим класичним підходом до МП текстів і аналізують статистичні відповідності між словами, словосполученнями та іншими лінгвістичними одиницями текстів вхідної та цільової мов.

Першим кроком при створенні моделі статистичного МП збирається великий набір паралельних текстів (корпус паралельних текстів, або навчальний), який складається з текстів вихідною мовою та їх відповідних перекладів цільовою мовою. Цей корпус служить основою для навчання статистичних моделей. Матеріалами для створення корпусу послугує документація міжнародних організацій таких як ООН, ЄС, Всесвітній банк, Організація охорони здоров'я тощо. Ці організації ведуть звітності паралельно декількома мовами.

Наступним кроком будується статистична модель з урахуванням частоти фраз, які зустрічаються в навчальному корпусі. Ця модель зберігає фрази та їх кількість повторень у всьому навчальному корпусі у вигляді таблиці. Чим частіше фраза зустрічається у навчальному корпусі, тим вища ймовірність правильного перекладу цільовою мовою. Довжина фраз, збережених у таблиці, може варіюватися від одного до п'яти слів.

Далі, вже маючи готову навчену модель, можна вводити текст вхідною мовою. Він розбивається на невеликі фрази або слова, які підлягають перекладу. Це може включати розділення на рівні речень, фраз або навіть слів. Наступним кроком для кожної фрази або слова знаходяться найбільш ймовірні переклади залежно від статистичних моделей. Ці переклади можуть бути вибрані на основі максимізації ймовірності перекладу або інших критеріїв, таких як мінімізація модельних втрат. Далі найбільш ймовірні

переклади фраз або слів збираються разом, щоб сформувати переклад вихідного тексту в цільову мову. Цей процес може включати додаткові евристики та модифікації для забезпечення граматично коректного та звучного перекладу.

Отже, головною перевагою даного класичного методу є його еволюційність, здатність адаптуватися до різних мов, джерел та типів текстів. Після кожного використання навчальний корпус уточнюється, що дає змогу отримати точніший і якісніший переклад у майбутньому.

Однак, головною проблемою цього методу МП є залежність від якості та розміру навчального корпусу. Якщо корпус навчання не є достатньо репрезентативним або містить помилки, це може негативно позначитися на якості перекладу. Також недоліками є важкість у вирішенні амбігвітності та обмежена обробка контексту та культури.

Ці недоліки статистичного перекладу виправляються новітніми підходами, такими як нейромережеві моделі, які демонструють покращену якість перекладу та більш гнучку обробку мовних нюансів.[2]

2.3 «Нейронний» машинний переклад

«Нейронний» МП, або Neural MT заснований на методах глибокого засвоєння інформації, або deep learning і сучасним методом підходу до машинного перекладу. Замість статистичних моделей та правил, які використовуються в інших методах, нейромережевий переклад використовує глибокі нейронні мережі, зокрема рекурентні RNN та трансформери, для автоматичного вивчення залежностей між текстами вхідною та цільовою мовою.

Як і в статистичному перекладі, спочатку створюється корпус паралельних текстів, а потім вже будується та навчається нейронна мережа. Створюється нейромережева модель, яка має вхідний шар для прийому

тексту вхідною мовою, приховані шари для аналізу та засвоєння залежностей, а також вихідний шар для генерації цільового перекладу.

На початку процесу перекладу вхідний текст проходить через кодувальний шар нейромережі. Це дозволяє моделі засвоїти контекст та залежності вихідного тексту.

Після кодування вхідного тексту модель починає генерувати цільовий переклад. Вона послідовно створює слова або словосполучення цільовою мовою, враховуючи контекст та залежності, які вона засвоїла в процесі навчання.

Після генерації модель може застосовувати додаткові методи декодування та пост-обробки, щоб покращити якість перекладу, враховуючи лексичні, синтаксичні та семантичні властивості мови.

«Нейронний» МП має кілька переваг порівняно з іншими перекладами. Він може краще враховувати контекст та залежності між словами, що сприяє точнішому та більш зрозумілому перекладу. Він також здатний краще працювати зі складними реченнями і мовними конструкціями та здатний швидко адаптуватися.

Однак «нейронний» МП вимагає значної обчислювальної потужності та ресурсів для навчання та використання моделі і є вразливим до недостатньо представленої інформації.[3]

2.4 Гібридний машинний переклад

Гібридний МП, або Hybrid MT комбінує в собі різні методи машинного перекладу, компенсуючи слабкі сторони одних сильними сторонами інших, для досягнення оптимального результату.

Комбінацій існує безліч. Наприклад, статистичний переклад може використовуватись як основний метод, а переклад на основі правил може використовуватись для корекції та покращення перекладу на основі

статистики. Іноді переклад на основі правил може використовуватись для обробки специфічних випадків або важких фраз, які статистичний переклад може не зрозуміти належним чином.

З появою «нейронного» МП стали з'являтися гібридні машинні переклади, де поєднуються всі три технології перекладачів.

Головна перевага гібридного машинного перекладу полягає в тому, що він поєднує найкращі аспекти різних підходів, дозволяючи досягти кращої якості перекладу та кращої адаптації до різних типів текстів та мовних задач. Проте, розробка та підтримка гібридних систем можуть бути складними, оскільки вони вимагають інтеграції різних компонентів та налаштування їх взаємодії. [4]

РОЗДІЛ 3. ПЕРЕКЛАДАЧ З УКРАЇНСЬКОЇ НА КРИМСЬКОТАТАРСЬКУ

3.1 Правила української мови

Українська мова – це мова східнослов'янської групи слов'янських мов, що входять до індоєвропейської мовної родини. Для запису слів використовується кирилиця.

Українська мова відноситься до флективних мов. Тобто включає в себе зміну форми слова (інфлексію) для вираження різних граматичних категорій, таких як відмінювання, числові та родові форми, відмінки, часові форми дієслова, ступені порівняння прикметників та прислівників тощо.

Наприклад, українські іменники відмінюються за числами (однина, множина) та відмінками (називний, родовий, давальний, знахідний, орудний, місцевий, кличний), а прикметники змінюються за родами, числами та відмінками, враховуючи також вищий та найвищий ступені порівняння.

Така флективна система дозволяє точно передавати граматичні відношення та структуру речення в українській мові, але вимагає уваги до правильного використання граматичних форм при формулюванні речень.

При побудові перекладу необхідно звертати увагу на правила. Наприклад:

1. *Граматичні відмінювання.* Українська мова має велику кількість відмінюваних частин мови, таких як іменники, займенники, прикметники та дієслова. Перекладач повинен враховувати правильні форми залежно від числа, роду, відмінка та інших граматичних категорій.

2. *Складні речення.* Українська мова дозволяє будувати складні речення з допомогою різних зв'язків, таких як сполучники та пунктуаційні

знаки. Перекладач повинен розуміти структуру речення і правильно передавати зв'язки та смислові зв'язки між різними частинами речення.

3. *Часові форми дієслів.* Українська мова має різні часові форми дієслів, такі як теперішній, минулий, майбутній часи, а також дійсний, умовний, наказовий способи. Важливо правильно визначати часову форму дієслова в контексті перекладу, щоб передати відповідний час та спосіб дії.

4. *Синтаксичні правила.* Українська мова має свої синтаксичні правила, які стосуються порядку слів у реченні, розташування залежних і незалежних частин речення та інших синтаксичних конструкцій. При перекладі слід дотримуватися цих правил, щоб зберегти структуру та зв'язок між словами у перекладі.

5. *Вживання правильних лексичних одиниць.* Українська мова має свою унікальну лексику та фразеологію. При перекладі на основі правил важливо використовувати відповідні лексичні одиниці, щоб передати точний сенс та виразність оригінального тексту.

Важливо також брати до уваги контекст, специфіку тексту та інші мовні особливості для досягнення якісного перекладу.[5][6]

3.2 Правила кримськотатарської мови

Кримськотатарська мова – це мова кипчацько-половецької групи тюркських мов, гіпотетичної алтайської мовної сім'ї. Для запису слів використовується латиниця, іноді кирилиця.

Кримськотатарська мова це аглютинативна мова з певними елементами аналітизму. Аглютинація – це мовний процес, в якому граматичні та семантичні значення виражаються шляхом додавання афіксів до основи слів. У кримськотатарській мові афікси додаються до кореня слова, щоб утворити

різні граматичні форми, такі як відмінки, числа, часові форми дієслів та інші граматичні категорії.

Наприклад, у кримськотатарській мові іменники можуть мати різні відмінкові форми, які утворюються шляхом додавання відповідних афіксів до основи слова. Також, прикметники та дієслова можуть змінюватися за числами, родами та іншими граматичними категоріями шляхом додавання афіксів. Наприклад, bağça – сад, bağçaları – сади, bağçada – в саду, bağçalarda – в садах.

Аналітичність же – це мовний характер, коли граматичні відношення та синтаксичні структури виражаються переважно за допомогою слів, а не шляхом змін форм частин мови чи використанням аглютинативних афіксів.

У кримськотатарській мові речення зазвичай будуються за допомогою словесних конструкцій та функціональних слів, які визначають граматичні відношення між словами. Вона має розвинутий словосполучувальний потенціал, де окремі слова можуть виконувати різні функції в реченні залежно від контексту.

Наприклад, в кримськотатарській мові для вираження часових форм дієслів часто використовуються допоміжні дієслова або словосполучення, які показують часову категорію. Також, для вираження роду та числа іменників, використовуються відповідні слівні ознаки та словосполучення.

Аглютинативно-аналітична природа кримськотатарської мови дозволяє виражати багато граматичної інформації за допомогою додавання афіксів до основи слів, робить її більш прозорою та легкою для сприйняття, оскільки граматичні відношення виражаються через використання окремих слів та словосполучень. Це також додає гнучкості та точності при вираженні граматичних відношень у мові.

При побудові перекладача також необхідно звернути увагу на такі правила:

1. *Словотворення.* У кримськотатарській мові для утворення нових слів використовуються префікси, суфікси та інші афікси. Важливо враховувати правила словотворення та використовувати відповідні афікси для утворення слів у перекладі.

2. *Граматики.* Кримськотатарська мова має свої граматичні правила, які включають в себе відмінювання і сполучення слів у речення. Дотримання правил граматики допомагає побудувати правильну структуру речення в перекладі.

3. *Складність слів.* У кримськотатарській мові часто зустрічаються складні слова, які складаються зі злиття коренів та афіксів. Розуміння структури складних слів допоможе визначити їх значення та відтворити їх у перекладі.

4. *Відмінювання і залежності.* У кримськотатарській мові існує система відмінювання і залежностей, яка використовується для вираження категорій, таких як рід, число, відмінок тощо. Важливо дотримуватися цих правил при перекладі, щоб передати правильні граматичні відношення.

5. *Вживання слів та виразів.* Кримськотатарська мова має свої унікальні слова та вирази. Використовуйте відповідні слова та вирази, щоб передати ідеї та значення у перекладі.

При побудові перекладу на основі правил важливо ознайомитися з граматиною та правилами кримськотатарської мови, а також враховувати контекст та специфіку перекладу для точного та зрозумілого передачі інформації.[7]

3.3 Порівняльний аналіз

Українська та кримська мова доволі різні несхожі між собою. Вони відносяться до різних мовних груп, класифікуються за різною морфологією, частіше використовуються різні абеткові системи письма. Нижче наведена порівняльна таблиця деяких основних мовних аспектів:

	Українська	Кримська
Мовні родини	Слов'янська	Тюркська
Морфологічна класифікація	Флективна	Аглютинативна, з елементами аналітичної
Алфавіт	Кирилиця	Латиниця, іноді кирилиця
Синтаксис	Стандартний порядок слів - SVO	Гнучкий порядок слів, часто SOV
Вплив інших мов	Російська, польська та інші	Тюркські, особливо турецька

Загалом, українська та кримськотатарська мови мають свої унікальні особливості, але їх зв'язок полягає у спільній мовній сфері України, де обидві мови відіграють важливу роль у мовному розмаїтті країни.

Для побудови МП на основі правил необхідно буде враховувати всі правила двох мов, всі їхні особливості та різницю між ними і створити правила та формули для програми-перекладача. Для побудови статистичного або «нейронного» МП необхідно мати достатньо велику базу слів, словосполучень та речень аби мати змогу навчити статистичну модель, або нейронну мережу. Для гібридного МП необхідно мати і правила, і базу.

В цьому і полягає основна проблема створення програми-перекладача. Кримськотатарська мова внесена всесвітньою організацією ЮНЕСКО в список мов, що вимирають. На даний момент існує відносно небагато спеціалістів, відсутня підходяща база речень і не надається фінансування для дослідження. Все це та ряд ще деяких причин не дозволяє створити повноцінний автоматичний перекладач з української на кримську.

РОЗДІЛ 4. ОГЛЯД WEB-ФРЕЙМОРКУ DJANGO

4.1 Опис технологій Django

В процесі розробки програми був використаний web-фреймворк для Python під назвою Django.

Django - це високорівневий веб-фреймворк для розробки веб-додатків на мові програмування Python. Він надає потужні засоби для швидкої розробки безпечних і масштабованих веб-додатків.

Основні риси та можливості Django:

1. *Модель-представлення-контролер (MVC)*. Django використовує архітектурний патерн MVC, що дозволяє логічно розділити компоненти додатку для кращої організації та обслуговування коду.

2. *ORM (Object-Relational Mapping)*. Django надає ORM шар, який дозволяє взаємодіяти з базою даних з використанням об'єктно-орієнтованого підходу. Це спрощує роботу з базою даних, дозволяє використовувати різні СУБД і робить код більш портативним.

3. *Автоматична адміністративна панель*. Django надає готовий модуль адміністрування, що дозволяє швидко створювати потужну адміністративну панель для керування даними вашого додатку.

4. *URL маршрутизація*. Django дозволяє легко визначати маршрутизацію URL, спрощуючи навігацію між сторінками вашого веб-додатку.

5. *Шаблонний двигун*. Django має вбудований шаблонний двигун, що дозволяє розділити логіку відображення веб-сторінок від логіки додатку. Шаблони Django підтримують наслідування, вбудовані фільтри та функції шаблонізації.

6. *Безпека*. Django включає в себе різні заходи безпеки, такі як захист від CSRF-атак, захист від SQL-ін'єкцій, хешування паролів, автоматичну очистку та екранування вхідних даних.

7. *Розширюваність.* Django заснований на принципі модульності, що дозволяє додавати сторонні пакети і розширювати функціональність вашого додатку.[8]

Django є популярним вибором для розробки веб-додатків завдяки своїм потужним функціям, гнучкості і широкому спектру підтримуваних інструментів та бібліотек. Він використовується в багатьох великих проектах і має активну спільноту розробників, що забезпечує стабільність та підтримку фреймворку.

4.2 Загальна структура проекту на Django

Проекти на Django будуються згідно з певною структурою, яка рекомендується фреймворком. Основні компоненти структури проекту Django включають наступні елементи:

1. *Каталог проекту.* Це основний каталог проекту, в якому містяться всі файлові компоненти проекту. Його назва може бути будь-якою, але зазвичай використовується назва проекту.

2. *Конфігураційний файл.* Файл settings.py містить налаштування проекту, такі як налаштування бази даних, шляхи до статичних файлів, налаштування мови тощо.

3. *URL-маршрутизація.* Файл urls.py визначає маршрутизацію URL-адрес до відповідних переглядів (views) у проекті.

4. *Перегляди (Views).* Файл views.py містить функції або класи, які обробляють HTTP-запити і повертають відповіді. Вони виконують логіку обробки запитів і взаємодіють з моделями та шаблонами.

5. *Моделі (Models).* Файл models.py визначає моделі даних проекту, які використовуються для взаємодії з базою даних. Моделі визначають таблиці бази даних, поля і взаємозв'язки між ними.

6. *Шаблони (Templates).* Каталог templates містить HTML-файли, які використовуються для відображення динамічного вмісту проекту. Вони

дозволяють вбудовувати дані з моделей та передавати їх до відповідного вигляду (view).

7. *Статичні файли (Static files)*. Каталог static містить статичні файли, такі як CSS-стилі, JavaScript-скрипти, зображення тощо. Ці файли використовуються для візуального оформлення та функціональності проекту.

8. *Міграції (Migrations)*. Каталог migrations містить файли міграцій, які використовуються для автоматичного оновлення структури бази даних при зміні моделей.

9. *Ресурси (Assets)*. Залежно від потреб проекту, можуть бути використані інші каталоги для зберігання ресурсів, таких як медіафайли, додаткові модулі, конфігураційні файли тощо.

Це лише загальна структура проекту на Django. Також він надає можливість додати власні каталоги та компоненти відповідно до потреб власного проекту. Django забезпечує гнучкість та організованість у будівництві веб-додатків.[9][10]

РОЗДІЛ 5. ПРОГРАМНА РЕАЛІЗАЦІЯ

5.1 Структура web-застосунку

Через певні труднощі, що унеможлиблюють створення автоматичної програми-перекладача, був розроблений автоматизований перекладач, чия суть полягає в тому, аби допомагати людині-перекладачу здійснювати переклад.

Структура проекту зображена на рисунку 1.

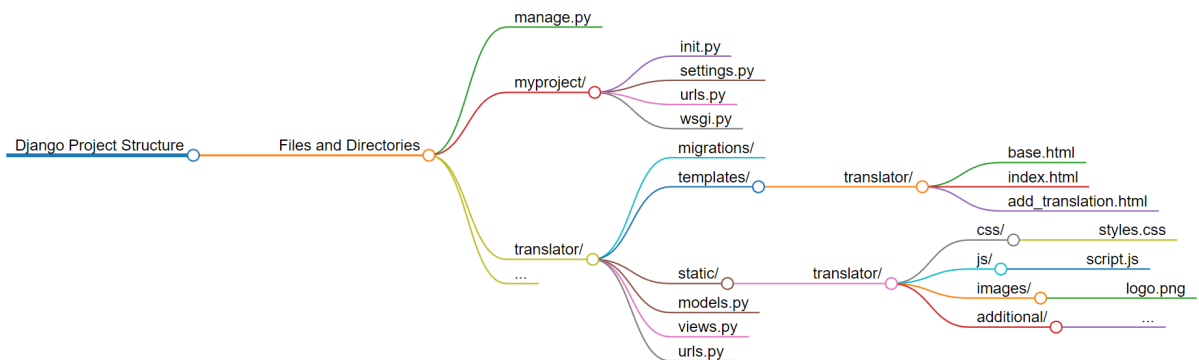


Рис 1. Структура проекту.

Файл `manage.py` запускає весь проект.

В папці `myproject` знаходяться стандартні файли фреймворку Django, які необхідні для створення проекту і запуску web-сторінки.

В папці `translator` знаходяться основні файли проекту, які і утворюють основну програму.

В папці `migration` знаходяться файли міграції, які оновлюють структури при введенні нової пари “слово-переклад”, або редагуванні, видаленні вже існуючої.

В папці `templates` знаходяться html-файли, за допомогою яких утворюється структура самого сайту.

В папці static знаходяться css-, js-файли, картинки тощо, що утворюють зовнішній вигляд web-сторінок.

Файл models.py містить в собі клас і функцію у ньому, що утворює і повертає пару “слово-переклад”.

В файлі views.py містяться функції, що обробляють запит, який користувач вводить в форму, і в залежності від цього повертають відповідні дані.

Файл urls.py містить в собі URL-маршрутизацію для кожної web-сторінки застосунку.

5.2 Вигляд готової програми

ПРИКЛАД РОБОТИ ПРОГРАМИ

На рисунку 2 зображена головна сторінка web-застосунку, де користувач може здійснювати пошук необхідного перекладу слова з української на кримськотатарську. Також він містить історію пошуку і повний список усіх слів, які додані в базу даних.

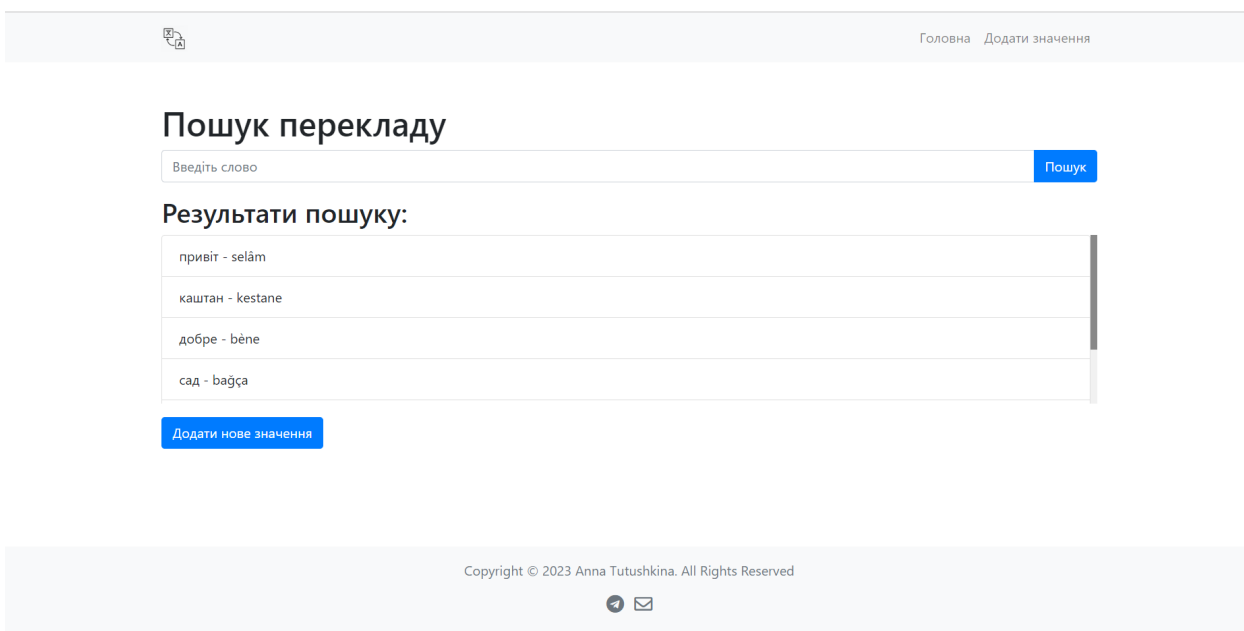


Рис 2. Головна сторінка застосунку.

Якщо слова в базі ще немає, то застосунок видає “Немає результатів”, що зображено на рисунку 3.

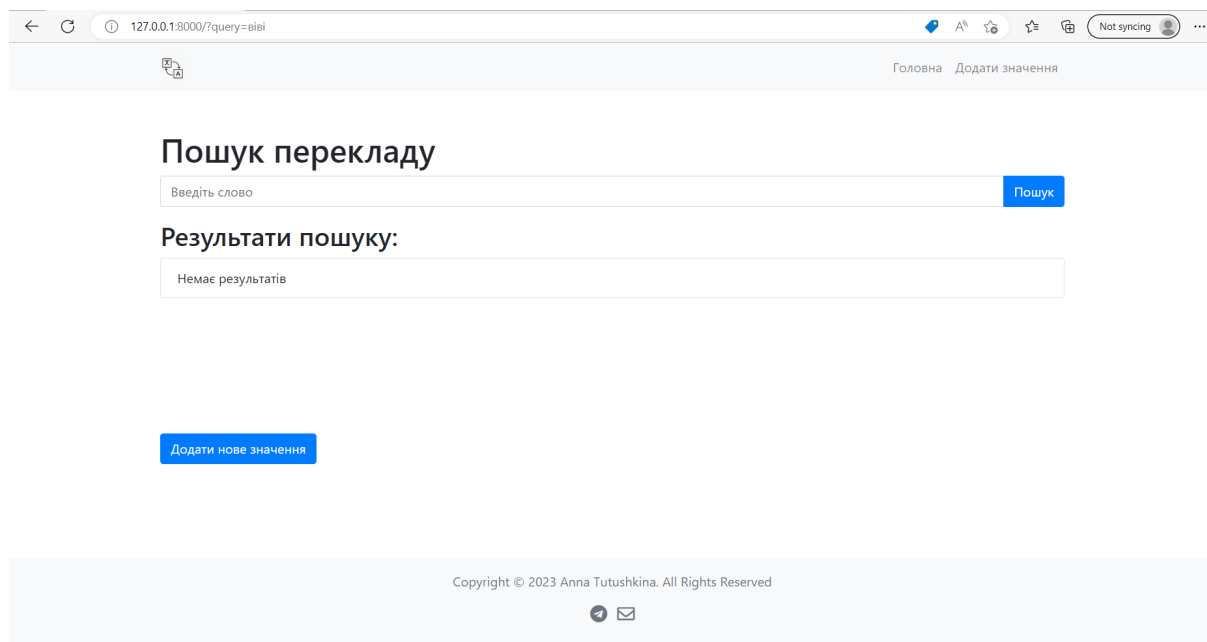


Рис 3. Функція повертає “Немає результатів” при введенні слова, що немає в базі.

Якщо слова немає, то при натисканні на кнопку “Додати нове значення” здійснюється перехід на сторінку “Додати переклад”, зображену на рисунку 4.

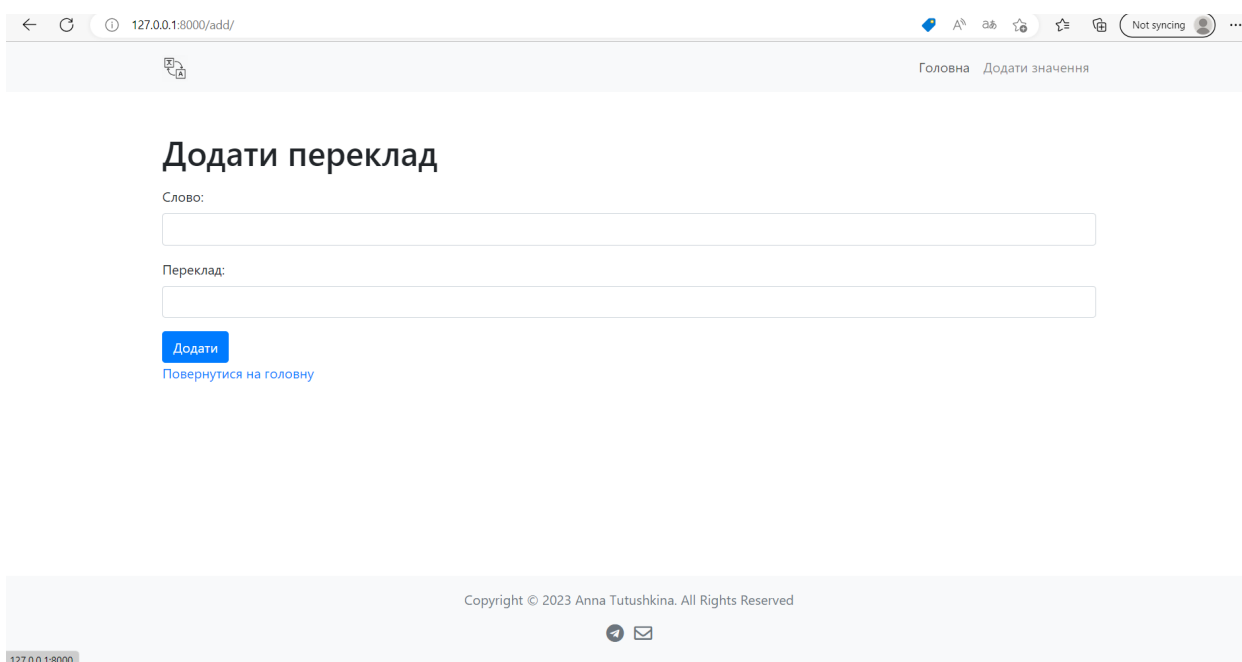


Рис 4. Вікно “Додати переклад”

На рисунку 5 зображена головна сторінка admin-панелі, яка містить в собі всі список усіх доступних проектів. Для того аби перейти в admin-панель необхідно до адреси головної сторінки додати /admin. Таким чином отримуємо адресу admin-сторінки: <http://127.0.0.1:8000/admin/> і доступ до неї.

Величезних плюсом фреймворку Django є те, що admin-панель є вбудованою в сам фреймворк. Admin-панель автоматично створює інтерфейс на основі моделей застосунку. Також ця панель використовує вбудовану систему автентифікації Django, що дозволяє контролювати доступ до адміністративних функцій за допомогою різних рівнів доступу та прав користувачів.

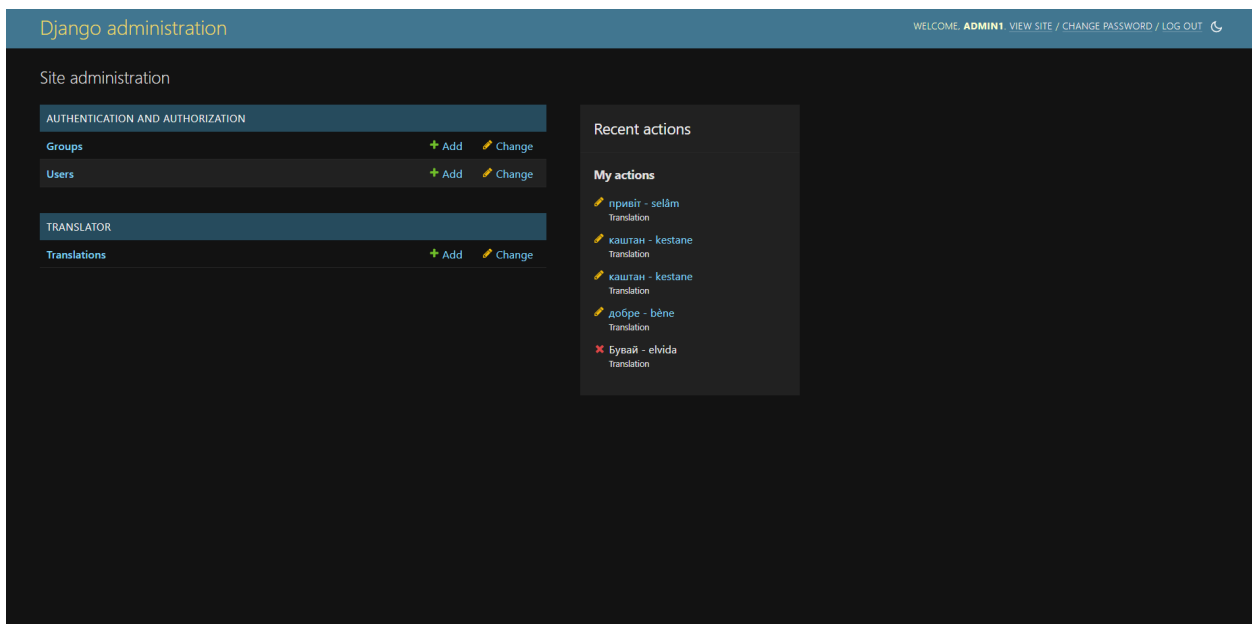


Рис 5. Головна сторінка admin-панелі.

Якщо обрати в проекті TRANSLATOR обрати пункт Translations до отримуємо доступ до бази всіх пар “слово-переклад”, зображеної на рисунку 6. На цій сторінці можна додавати, редагувати та видаляти пари “слово-переклад”.

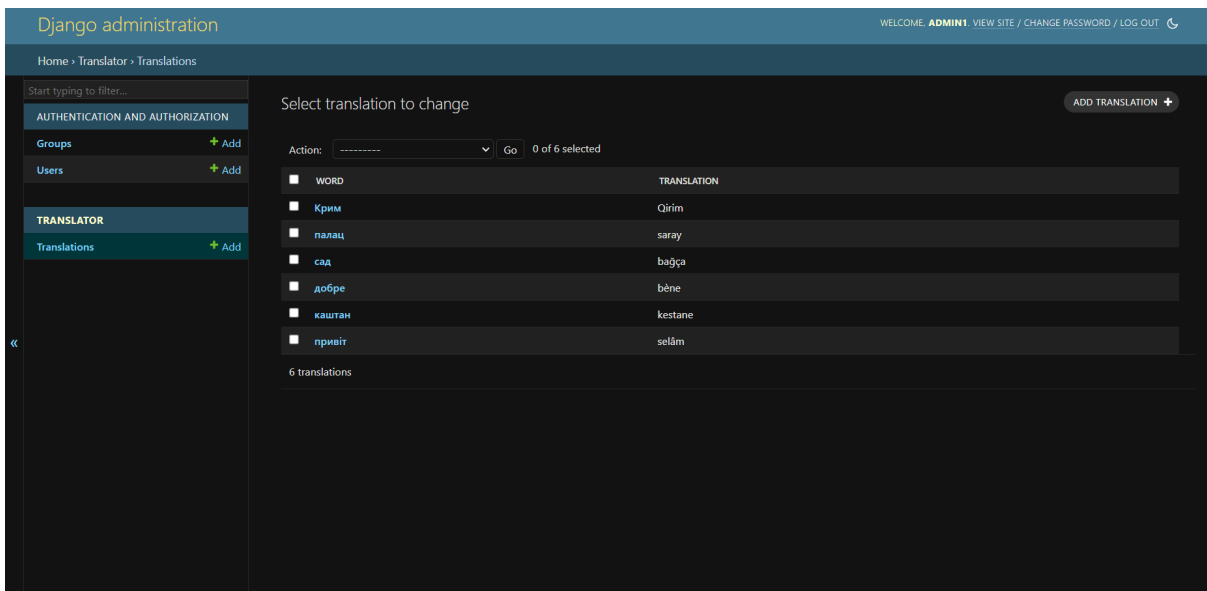


Рис 6. Вікно проекту TRANSLATOR в admin-панелі з можливістю редагувати існуючі пари “слово-переклад”

ВИСНОВОК

В ході кваліфікаційної роботи були розглянуті та дослідженні різні методи та підходи до машинного перекладу, особливості української та кримськотатарських мов, які можуть знадобитися для побудови програми-перекладача, а також основні проблеми.

Для програмної реалізації задачі було створено web-застосунок, що являє собою бази слів українською та відповідників їм кримськотатарською з можливістю кастомізації. Маючи акаунт адміністратора і маючи доступ до admin-панелі, є можливість додавати нові пари “слово-переклад”, редагувати вже існуючі та видаляти їх. Звичайні користувачі можуть використовувати пошук по базі для отримання необхідного перекладу.

Така програма може бути використана як допомога для людини-перекладача, а також стати основою під більш складні і точні автоматичні перекладачі.

Головною перевагою такої програми є зручність використання і великі можливості для розвитку.

ДЖЕРЕЛА

- [1]. Budiansky S. Lost in translation. Atlantic magazine. 1998. P. 81–84.
- [2]. Lehrberger J. Machine translation: Linguistic characteristics of MT systems and general methodology of evaluation. Amsterdam : J. Benjamins Pub. Co., 1988. 240 p.
- [3]. Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation / A. Toral et al. Proceedings of the Third Conference on Machine Translation: Research Papers. Brussels, 2018. P. 113–123.
- [4]. Boretz A. AppTek Launches Hybrid Machine Translation Software. Speech Tag Online Magazine. 2009 [Електронний ресурс] – Режим доступу до ресурсу: <https://www.speechtechmag.com/Articles/ReadArticle.aspx?ArticleID=5287>
1
- [5]. Безпояско О. К., Городенська К. Г. Морфеміка української мови. К., 1987
- [6]. Українська мова : енциклопедія / ред. В. М. Русанівський. Київ : НАН України, Ін-т мовознавства ім. О. О. Потебні, Ін-т укр. мови, 2000.
- [7]. KavitaKaya D. Crimean tatar. Berlin : LINCOM EUROPA, 2010. 129 p.
- [8]. DjangoBook — Безкоштовна книга про Django [Електронний ресурс] – Режим доступу до ресурсу: <https://djangobook.com/>.
- [9]. Biblioteka Pythona — Документація та мануали [Електронний ресурс] – Режим доступу до ресурсу: <https://web.archive.org/web/20180913143347/http://www.python.rk.edu.pl/>.
- [10]. Українська спільнота matrix присвячена Django [Електронний ресурс] – Режим доступу до ресурсу: https://matrix.to/###django_ukraine:matrix.opulus.space

ДОДАТОК

```
class Translation(models.Model):
    word = models.CharField(max_length=255, unique=True)
    translation = models.CharField(max_length=255)

    def __str__(self):
        return f'{self.word} - {self.translation}'

app_name = 'translator'

urlpatterns = [
    path('', views.search, name='index'),
    path('add/', views.add_translation, name='add_translation'),
]

class TranslationForm(forms.Form):
    word = forms.CharField(label='Слово', max_length=100)
    translation = forms.CharField(label='Переклад', max_length=100)

def main():
    """Run administrative tasks."""
    os.environ.setdefault('DJANGO_SETTINGS_MODULE',
        'translation_project.settings')

    try:
        from django.core.management import execute_from_command_line
    except ImportError as exc:
        raise ImportError(
            "Couldn't import Django. Are you sure it's installed and "
```

```
        "available on your PYTHONPATH environment variable? Did you "  
        "forget to activate a virtual environment?"  
    ) from exc  
execute_from_command_line(sys.argv)  
  
if __name__ == '__main__':  
    main()
```