

I. Введение

Проект направлен на разработку мультязычного тезауруса нового поколения, ориентированного на узбекский и немецкий языки, с интеграцией ИИ, корпусной лингвистики и цифровых технологий (AR, TTS, JSON, NLP). Актуальность определяется необходимостью поддержки образования, миграции, терминологической точности и переводов. Повышенный интерес подтверждается и международным научным сообществом, где тема развивается в рамках проек...

II. Структура тезауруса и база данных

- **Lemma:** основная форма слова (например: ilm)
- **POS:** часть речи (например: s.m. — существительное мужского рода)
- **Definition:** краткое определение
- **Translation:** эквивалент на немецком языке
- **Domain:** предметная область (туризм, логистика, медицина)
- **Frequency:** частота в корпусе
- **Register:** стиль (официальный, разговорный, научный)

Пример JSON:

```
{
  "lemma": "kitob",
  "pos": "noun",
  "definition": "bilim man..."
```

Также создаются расширенные лексические гнезда (пример: музей), включающие определения, перевод, домены, подтипы, примеры, связанные термины и мультязычные конструкции (UZ, DE, EN).

III. Этапы разработки

Этап I. Подготовка источников

Словари: izohli lug'at, etimologik lug'at, tarixiy lug'at, DWDS.de, Multitran, Phrasebook

Этап II. Извлечение и аннотирование

Инструменты: pdfplumber, spaCy, stanza, regex, OCR. Форматы: JSON, TEI/XML, CSV

Этап III. Корпусная обработка и семантика...

IV. Тематические направления


А. Туризм: buchen, check-in, Trinkgeld → buyurtma qilmoq, ro'yxatdan o'tish, choypuli

В. Логистика: Fracht, Drohne, Telematik → yuk, dron, telematika

****C. Образование и миграция****: Studium, Sprachkurs, Zertifikat → ta'lim, til kursi, sertifikat

****D. Транспорт****: метро, трамвай, электротранспорт → metro, tramvay, elektr transporti → тематическая энциклопедия «Метро Ташкента»

V. AR/AI/EdTech Модули

-  ****AR-лексика****: визуализация предметов с переводом

...

VI. Научная и правовая база

- Диссертации (UZ/DE): структура, методология, цифровая реализация

- Журналы: фиксация этапов, JSON/корпус/гнезда

- Авторские декларации (июль 2025): охрана авторских прав и приоритета

- GreenUZ & Codex Terra: архивная защита культурного и научного наследия

VII. Вывод

Создание мультязычного тезауруса – ключевой вклад в цифровую лексикографию, межкультурную коммуникацию и языковое наследие. Платформа охватывает AI, AR, EdTech, лингвистику и прикладные дисциплины....