РЕФЕРАТ

СИСТЕМА ДЛЯ ПОЛУАВТОМАТИЗИРОВАННОГО СОЗДАНИЯ ДАТАСЕТОВ ИЗ БОЛЬШИХ АУДИОФАЙЛОВ ДЛЯ ОБУЧЕНИЯ МОДЕЛЕЙ ГЕНЕРАЦИИ РЕЧИ: дипломная работа / Прокопенко Д.Д. – Гомель: ГГТУ им. П.О. Сухого, 2026. – Дипломная работа: 117 страниц, 9 рисунков, 11 таблиц, 17 источников, 6 приложений.

Ключевые слова: датасет, синтез речи, выравнивание аудио и текста, *Vosk*, обработка речи, машинное обучение, транскрибация, *PyQt*5, *Python*, *JSON*.

Объектом разработки является программный продукт для персональных компьютеров под управлением операционной системы *Windows*, реализующий полуавтоматизированное создание размеченных речевых датасетов из больших аудиофайлов для обучения моделей генерации речи.

Целью работы является разработка системы, позволяющей автоматизировать процесс транскрибации и временного выравнивания текста с аудиодорожкой, а также предоставляющей инструменты для визуального редактирования временных меток и экспорта готового датасета в формате *JSON*.

Этапы выполнения работы состоят из: анализа существующих методов автоматического распознавания речи и выравнивания аудио и текста; обзора существующих открытых речевых датасетов для русского и английского языков; проектирования архитектуры системы; реализации модулей автоматического распознавания речи на базе *Vosk*, визуального редактора временных меток с поддержкой *waveform*-визуализации, модулей постобработки текста (расстановка знаков препинания, разметка предложений по паузам); проведения экспериментального тестирования точности выравнивания и оценки экономии времени; произведения экономического расчёта и обоснования рентабельности разработки; анализа внедрения разработки с точки зрения вопросов энерго- и ресурсосбережения.

Студент-дипломник подтверждает, что дипломная работа выполнена самостоятельно, приведенный в дипломной работе материал объективно отражает состояние разрабатываемого объекта, пояснительная записка проверена в системе «Антиплагиат» (*https://antiplagiat.ru*). Процент оригинальности составляет 88.04%. Все заимствованные из литературных и других источников теоретические и методологические положения и концепции сопровождаются ссылками на источники, указанные в «Списке использованных источников».

Резюме

Тема дипломной работы: «Система для полуавтоматизированного создания датасетов из больших аудиофайлов для обучения моделей генерации речи».

Объект исследования: библиотека автоматического распознавания речи *Vosk*, фреймворк для создания графических интерфейсов *PyQt*5.

Цель работы: разработка десктопного приложения, обеспечивающего автоматическую транскрибацию аудиофайлов, визуальное редактирование временных меток и экспорт размеченного датасета в формате *JSON*.

Основной результат работы: программный продукт для персонального компьютера с операционной системой *Windows*, реализующий полуавтоматизированное создание речевых датасетов с поддержкой русского и английского языков.

Рэзюмэ

Тэма дыпломнай працы: «Сістэма для паўаўтаматызаванага стварэння датасетаў з вялікіх аўдыёфайлаў для навучання мадэлей генерацыі маўлення».

Аб'ект даследавання: бібліятэка аўтаматычнага распазнавання маўлення *Vosk*, фрэймворк для стварэння графічных інтэрфейсаў *PyQt*5.

Мэта працы: распрацоўка дэсктопнага прыкладання, якое забяспечвае аўтаматычную транскрыбацыю аўдыёфайлаў, візуальнае рэдагаванне часавых метак і экспарт размечанага датасета ў фармаце *JSON*.

Асноўны вынік працы: праграмны прадукт для персанальнага камп'ютара з аперацыйнай сістэмай *Windows*, які рэалізуе паўаўтаматызаванае стварэнне моўных датасетаў з падтрымкай рускай і англійскай моў.

Abstract

Subject of the work: «Semi-automated dataset creation system from large audio files for speech generation model training».

Object of study: Vosk automatic speech recognition library, PyQt5 graphical interface framework.

Purpose of work: development of a desktop application that provides automatic audio transcription, visual timestamp editing, and labeled dataset export in JSON format.

The main result of the work: a software product for a personal computer with the Windows operating system, implementing semi-automated speech dataset creation with support for Russian and English languages.

Перечень условных обозначений и сокращений

В настоящей пояснительной записке применяются следующие термины, обозначения и сокращения.

API – (англ. *Application Programming Interface*) – программный интерфейс приложения.

ASR – (англ. *Automatic Speech Recognition*) – автоматическое распознавание речи.

CTC – (англ. *Connectionist Temporal Classification*) – классификация временных соединений.

FLAC – (англ. *Free Lossless Audio Codec*) – свободный кодек сжатия аудио без потерь.

HMM – (англ. *Hidden Markov Model*) – скрытая марковская модель.

JSON – (англ. *JavaScript Object Notation*) – текстовый формат обмена данными.

MFCC – (англ. *Mel-Frequency Cepstral Coefficients*) – мел-частотные кепстральные коэффициенты.

MFA – (англ. *Montreal Forced Aligner*) – инструмент для принудительного выравнивания.

*MP*3 – (англ. *MPEG*-1 *Audio Layer* 3) – формат сжатия аудио с потерями.

NLTK – (англ. *Natural Language Toolkit*) – инструментарий для обработки естественного языка.

SDK – (англ. *Software Development Kit*) – набор средств разработки программного обеспечения.

TTS – (англ. *Text-to-Speech*) синтез речи из текста.

UML – (англ. *Unified Modeling Language*) – унифицированный язык моделирования.

UTF – (англ. *Unicode Transformation Format*) – формат преобразования *Unicode*.

WAV – (англ. *Waveform Audio File Format*) – формат аудиофайлов.

БД – база данных.

ИС – информационная система.

ПК – персональный компьютер.

ПО – программное обеспечение.

ПП – программный продукт.

ПС – программное средство.

РБ – Республика Беларусь.

ТКП – технический кодекс установившейся практики.

ЭВМ – электронно-вычислительная машина.