

Кафедра инженерной кибернетики

# ОТЧЕТ

ПО

## ЛАБОРАТОРНОЙ РАБОТЕ

*«Разработка демонстрационного прототипа приложения для  
решения специализированной задачи интеллектуальной  
обработки и анализа информации с использованием современных  
ИИ-сервисов»*

учебная дисциплина «Методы искусственного интеллекта»

Группа: БПМ-19-2

Учащийся: Комлев Д.А.

Преподаватель: Дранга Д., Кожаринов А.С.

Оценка:

Дата защиты: 12.09.2022

2022 г.

# **1. ВВЕДЕНИЕ**

## **1.1 Цель лабораторной работы**

Выработать у учащихся устойчивые умения (навыки) по организации взаимодействия на уровне программного кода между собственным программным обеспечением и функционалом существующих ИИ-сервисов нового поколения: Yandex Cloud, Microsoft Azure, Houndify. При решении некоторой конкретной типовой задачи интеллектуальной обработки и/или анализа информации.

Определить уровень качества, с которым основные современные ИИ-сервисы могут решать некоторую конкретную типовую задачу интеллектуальной обработки и/или анализа данных и выявить лучший из использованных ИИ-сервисов для конкретной задачи.

# **2. ОПИСАНИЕ ЗАДАЧИ**

## **2.1 Постановка задачи**

Разработать приложение, преобразовывающее аудиозаписи в текст с помощью ИИ-сервисов нового поколения. Сравнить ИИ-сервисов на нескольких одинаковых аудиозаписях.

## **2.2 Описание приложения**

Приложение должно использовать три ИИ-сервиса (Yandex Cloud, Microsoft Azure, Houndify), иметь возможность выбора файлов для распознавания. В результате работы приложения в текстовом окне должен выводиться скрипт аудиозаписи.

### **3. СРЕДСТВА ДЛЯ РАЗРАБОТКИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ**

#### **3.1. Языковые и инструментальные средства, использованные для создания приложения**

- Платформы для разработки программного обеспечения: .net6
- Язык программирования: C#
- Графический интерфейс: WPF (Windows Presentation Foundation)

#### **3.2. Используемый функционал:**

ИИ-сервисы, применяемы для преобразования речи в текст:

- Yandex Cloud
- Microsoft Azure
- Houndify

Используемое SDK:

SDK, предоставляемое Houdnfiy (оболочка над API)

Используемые программы:

FFMEG – конвертация аудиофайлов в различных форматах

## 4. ПРИМЕР РАБОТЫ

### 4.1 Исходное приложение

Приложение имеет кнопку для выбора аудиофайла, 3 кнопки для выбора ИИ-сервиса, который будет использоваться для распознавания, а также поле, в которой будет отображаться результат распознавания.

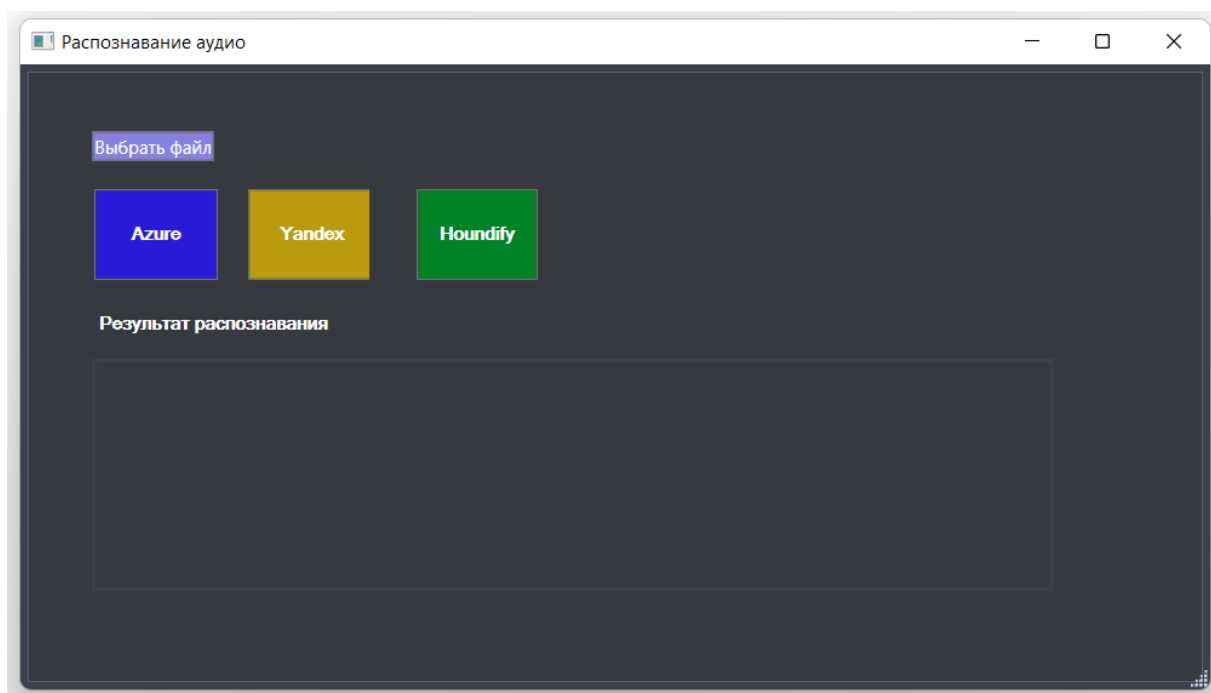


Рис. 1 Интерфейс приложения

### 4.2. Исходные данные

В качестве исходных данных могут выступать любые аудиофайлы форматов: mp3, wav, opus, ogg. За тем исключением, что API Yandex Cloud имеет ограничение на длину аудио: 30 сек. И ограничение на размер файла: 1мб.

Разные ИИ-сервисы требуют разные форматы аудиофайлов. А именно: Microsoft Azure поддерживает только формат wav, Yandex Cloud поддерживает форматы ogg, opus, Houndify поддерживает форматы wav и opus. В связи с этими ограничениями в программу встроены конвертаторы аудиофайлов, которые приводят выбранный пользователем файл в необходимому для ИИ-сервиса разрешению. Для этого используется утилита FFMPEG.

Для тестирования приложения были выбраны 4 аудиозаписи: 3 аудиозаписи

со стихами на английском и одна запись с разговором.

#### 4.3. Пример работы

При нажатии на кнопку “Выбрать файл” открывается диалоговое окно, в котором пользователь выбирает аудиофайл (Рис. 1).

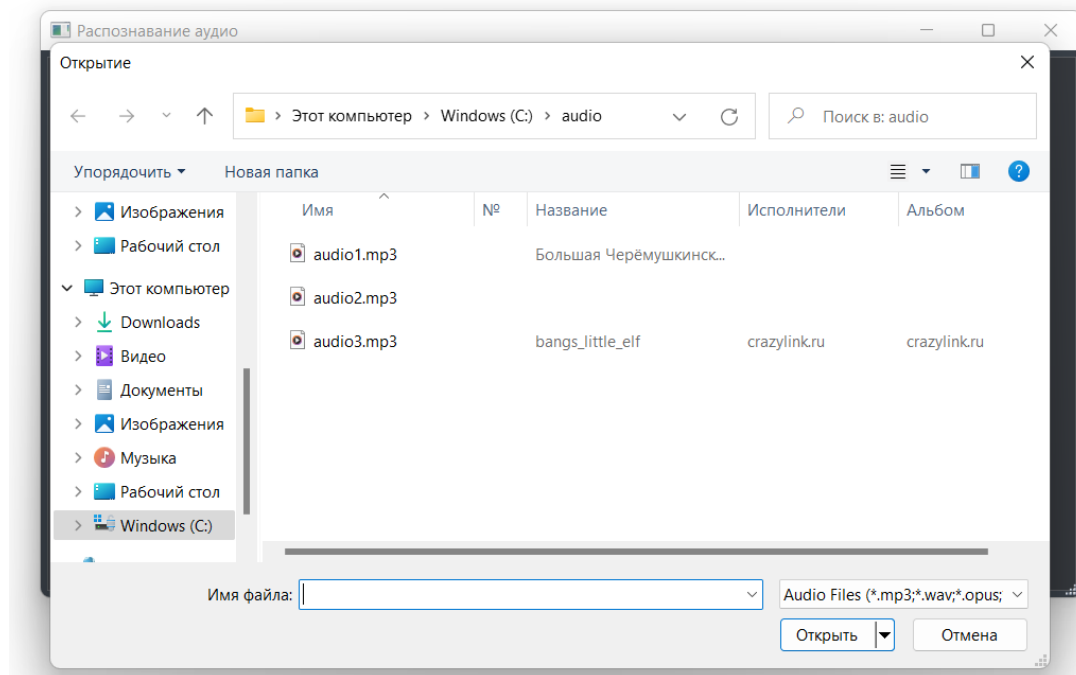


Рис. 2 Выбор аудиофайла

После того, как аудиофайл выбран, его название будет отображено справа от кнопки “Выбрать файл” (Рис. 2).

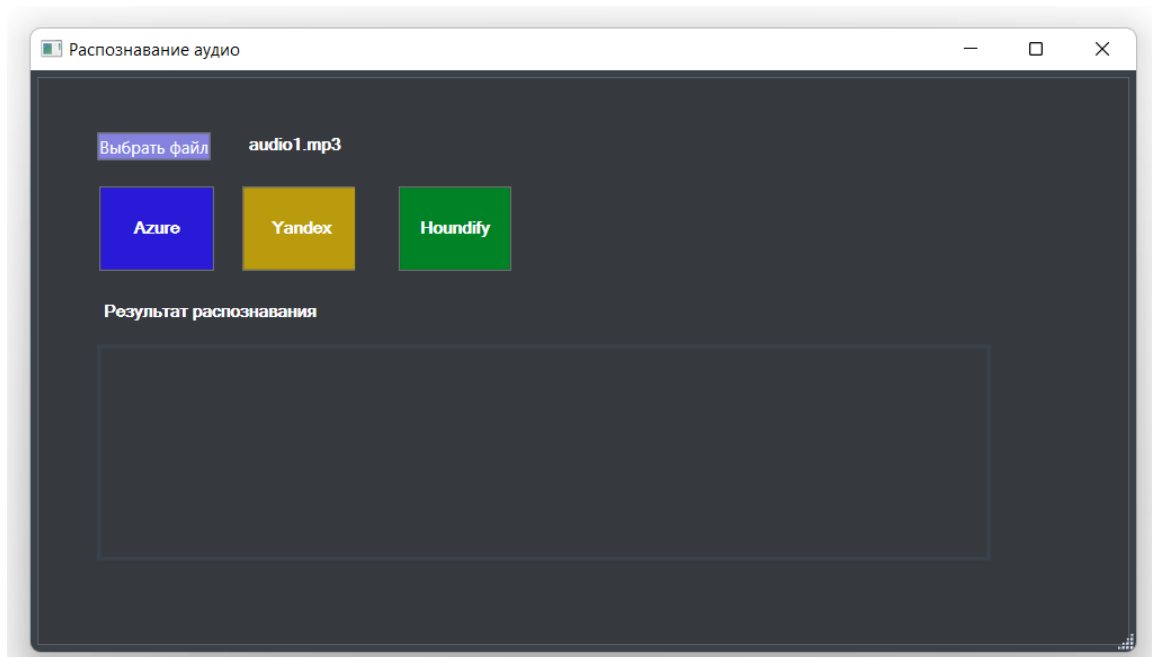


Рис. 3 Файл выбран.

После выбора файла можно приступить к распознаванию. Для этого необходимо выбрать один из ИИ-сервисов и дождаться завершения его работы.

После нажатия на кнопку с названием ИИ-сервиса, в области поле вывода будет отображён индикатор загрузки, говорящий о том что распознавание в процессе (Рис. 3).

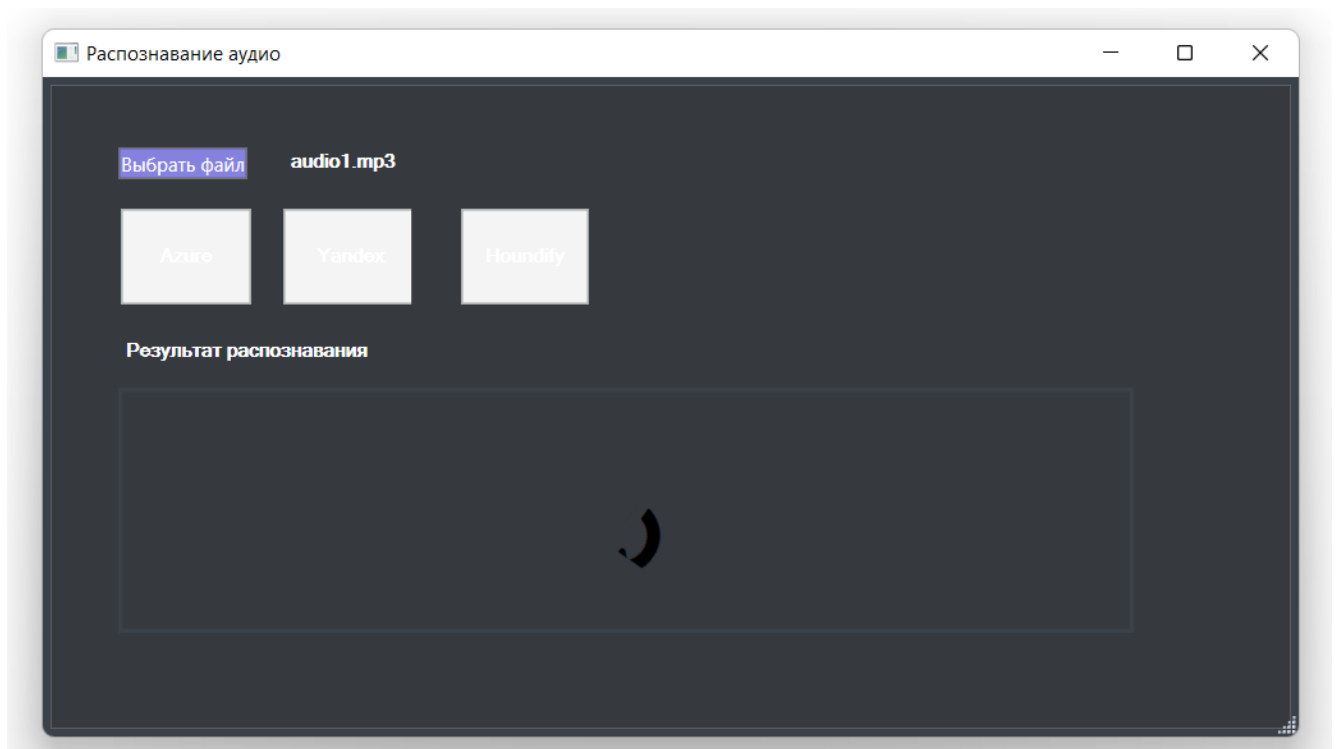


Рис. 3 Индикатор распознавания

Когда распознавание будет завершено, его результат будет выведен в поле вывода (Рис. 4).

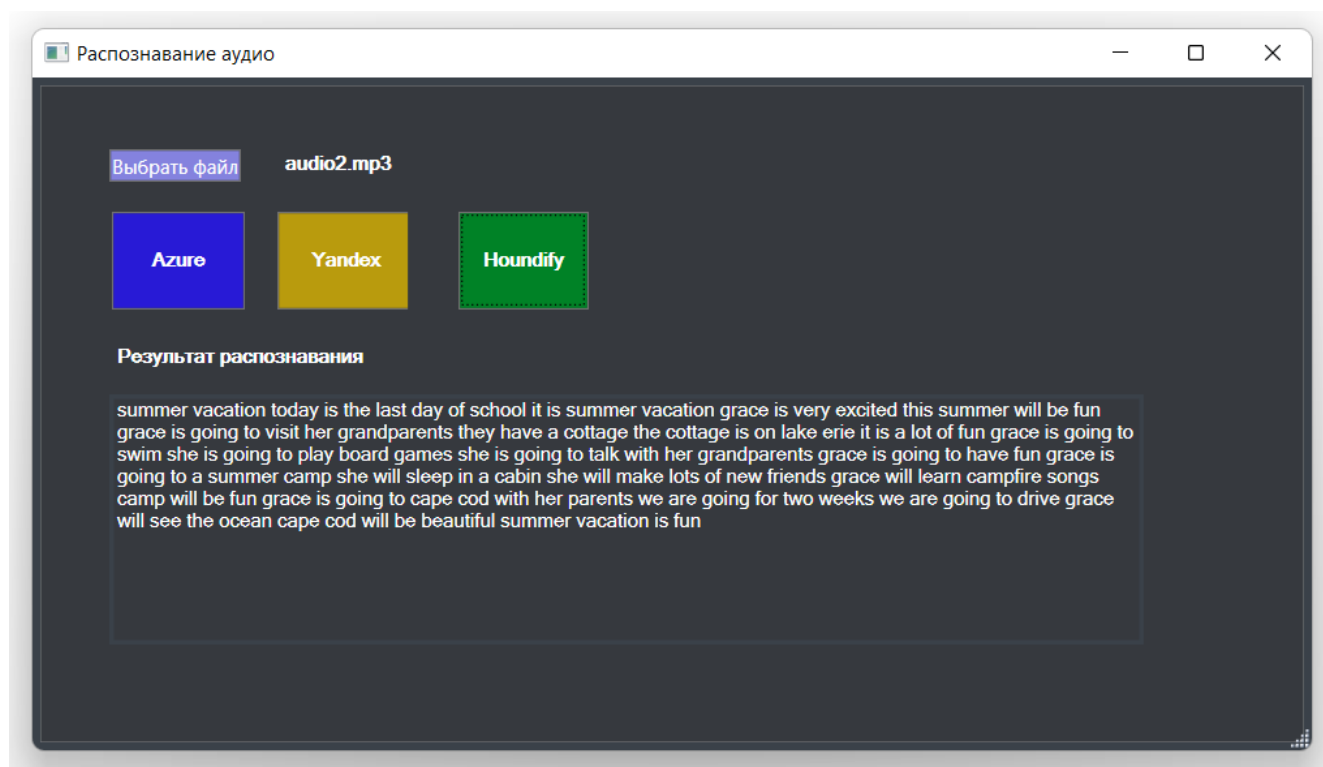


Рис. 4 Результат распознавания.

В случае возникновения каких-либо ошибок, сообщения о них будет отображено в модальном окне. Например, в случае нажатия кнопки распознавания без выбранного файла будет показано следующее сообщение:

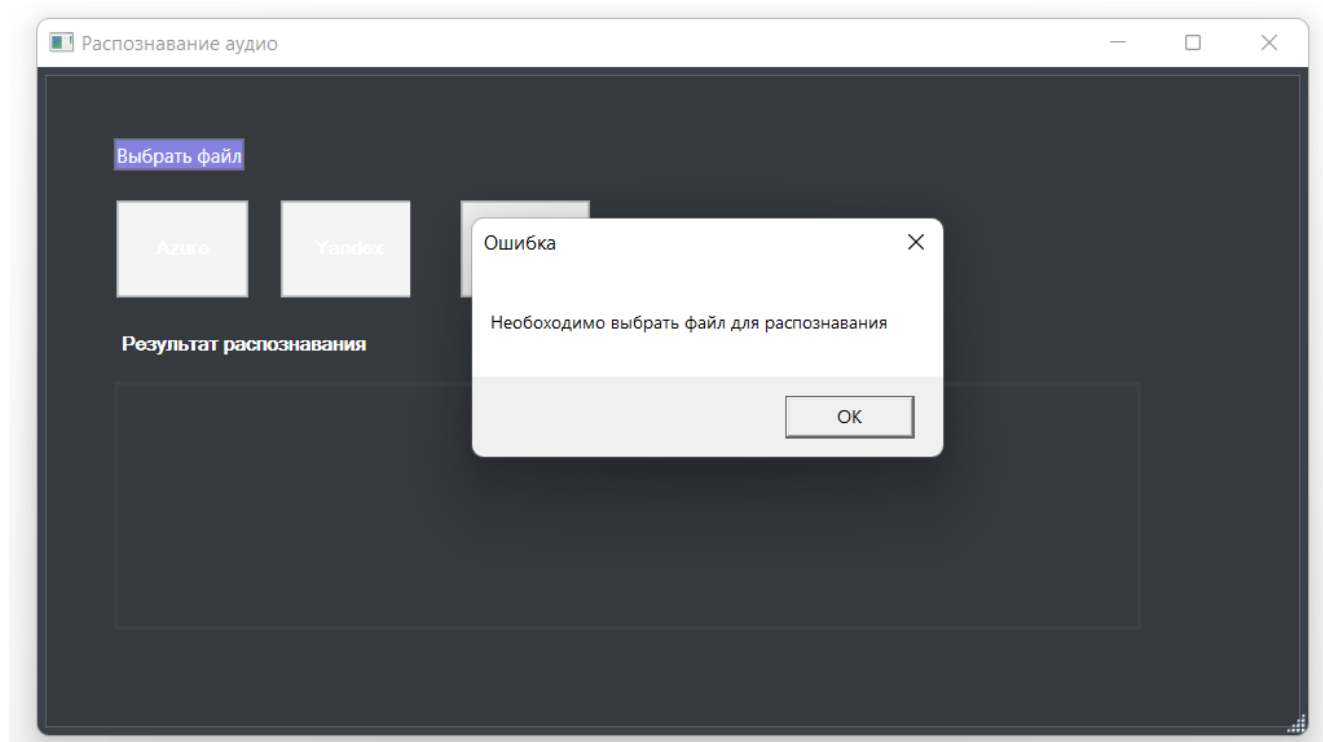


Рис. 5 Модальное окно с сообщением об ошибке

## **5. РЕЗУЛЬТАТЫ**

### **5.1. Microsoft Azure**

Audio 1 (Запись разговора человека, который последовательно отходит дальше от микрофона). Распознано около 30% записи. Распознавание завершилось когда человек отошел на 3 метра от микрофона

Audio3 (Стих с хорошим качеством звука) распознан на 100%, также частично расставлены знаки препинания (около 60%).

Audio2 (Стих с тихим звуком) распознан на 95%. Не распознаны последние 3 секунды. Расставлены все знаки препинания, 1 знак препинания расставлен неверно.

### **5.2. Yandex Cloud**

Audio3 (Стих с хорошим качеством звука) распознан на 90%, 3 слова из 56 распознаны неверно. Знаки препинания в распознанном тексте отсутствуют.

Audio 1 (Запись разговора человека, который последовательно отходит дальше от микрофона). Запись была разбита на части по 30 (в связи с ограничением API). Во всех частях аудио было распознано около 40-50% записи.

Audio2 (Стих с тихим звуком) распознан на 100% (запись была разбита на 2 части по 30 секунд). Знаки препинания отсутствуют.

### **5.3. Houndify**

Audio3 (Стих с хорошим качеством звука) распознан на 90%, 1 слово из 56 распознаны неверно. Знаки препинания в распознанном тексте отсутствуют.

Audio2 (Стих с тихим звуком) распознан на 100%. Знаки препинания отсутствуют.

Audio 1 (Запись разговора человека, который последовательно отходит дальше от микрофона). Распознано около 50% записи.



## **6. ВЫВОДЫ**

В этой лабораторной работе я получил опыт работы с API таких ИИ-сервисов, как Microsoft Azure, Yandex Cloud, Houndify. Было разработано десктопное приложения для распознавания речи в аудиофайлах.

Лучше всего с задачей справились Microsoft Azure и Houndify. Они показали высокий процент распознавания, а Microsoft Azure также расставлял знаки препинания. Yandex Cloud тоже хорошо справился с задачей, однако сильным минусом является то, что у него есть ограничение на размер аудиофайла – 1мб и на длину аудио – 30 сек.