

Разработка веб-тренажёра публичных выступлений

Даниил Плющенко
(руководитель: Марк Маркович Заславский)

Введение

Большое количество людей, например, студентов, аспирантов и исследователей, регулярно выступают с докладами о проделанной работе.

Чтобы дать полезную обратную связь, нужны квалификация в предметной области и время для прослушивания, анализа и оценки доклада.

Однако всегда можно найти или сформулировать набор критериев и свойств, которые могут быть оценены автоматически, т.е. провести первичную оценку качества доклада.

Обзор литературы и аналогов

Были рассмотрены приложения, помогающие улучшить речь и/или навыки публичных выступлений, выдающие некоторую оценку, а также статьи ^{1, 2, 3} на эту тему.

В основном исследуется поведение докладчика ^{1, 3} и критерии хорошего выступления ².

Общие выводы по приложениям:

- в основном поддерживается только английский язык
- нет возможности прикрепить презентацию, настроить критерии
- нет интеграции с внешними системами (LTI -- протокол взаимодействия учебных систем, REST API)

¹ S. Kopf et. al. A Real-time Feedback System for Presentation Skills, 2015

² J. Schneider et. al. Presentation Trainer, your Public Speaking Multimodal Coach, 2015

³ A. Hanani et.al. Automatic Estimation of Presentation Skills Using Speech, Slides and Gestures, 2017

Название	API/LTI?	Возможность прикрепить презентацию	Языки	Возможность настроить критерии?	Платное?
Speechace	Да	Нет	Английский	Нет	Демо
Voice Notebook / Speechpad	Нет	Нет	8	Нет	Нет
Speakit	Нет	Нет	Американский английский	Нет	Нет
Aksent	Нет	Нет	20+	Нет	Нет
Speeko	Нет	Текстовые заметки для текущей записи	Английский	Нет	Бесплатный Trial
Likeso	Нет	Нет	Английский	Подмножество слов-паразитов	Да
Orai	Нет	Текстовые заметки	Английский	Подмножество слов-паразитов	Да
ElsaSpeak	Нет	Нет	Американский английский	Нет	Бесплатное демо
Говорилло	Нет	Нет	Русский	Нет	Нет

Компоненты публичного выступления

- Речь докладчика
- Презентация, состоящая из слайдов
- Временные метки переключения слайдов

Примеры критериев:

- Для речи: длительность речи, её темп, количество слов-паразитов
- Для презентации: поиск ключевых слов на слайде, перегруженность слайда (работа другого студента)
- Общие: проверка, связан ли слайд с тем, что говорит докладчик

Цели и задачи

Цель -- разработать модульное и масштабируемое веб-приложение с открытым исходным кодом для первичной оценки качества публичного выступления.

Задачи:

- Провести обзор библиотек распознавания речи
- Разработать расширяемую архитектуру, позволяющую добавлять и заменять компоненты, связанные с анализом публичного выступления
- Разработать веб-приложение, используя предложенную архитектуру

Сравнение библиотек распознавания речи

Размер шингла	1		2		3		4	
	Среднее, %	Отклонение	Среднее, %	Отклонение	Среднее, %	Отклонение	Среднее, %	Отклонение
Vosk	74.84	10.03	60.27	12.28	50.08	12.94	42.24	12.80
Speech-to-Text (Russian)	49.14	14.41	28.27	12.33	17.78	10.34	11.82	8.30
Wit.ai	55.29	11.70	35.70	10.99	25.47	9.76	18.85	8.57

Алгоритм шинглов -- алгоритм сравнения схожести текстов. По сути сравнивает между собой набор слов в окне определённого размера.

Vosk показал лучшие результаты на собранном наборе данных из публичных выступлений, проведённых на русском языке, поэтому выбрали его.

Обработка тренировки

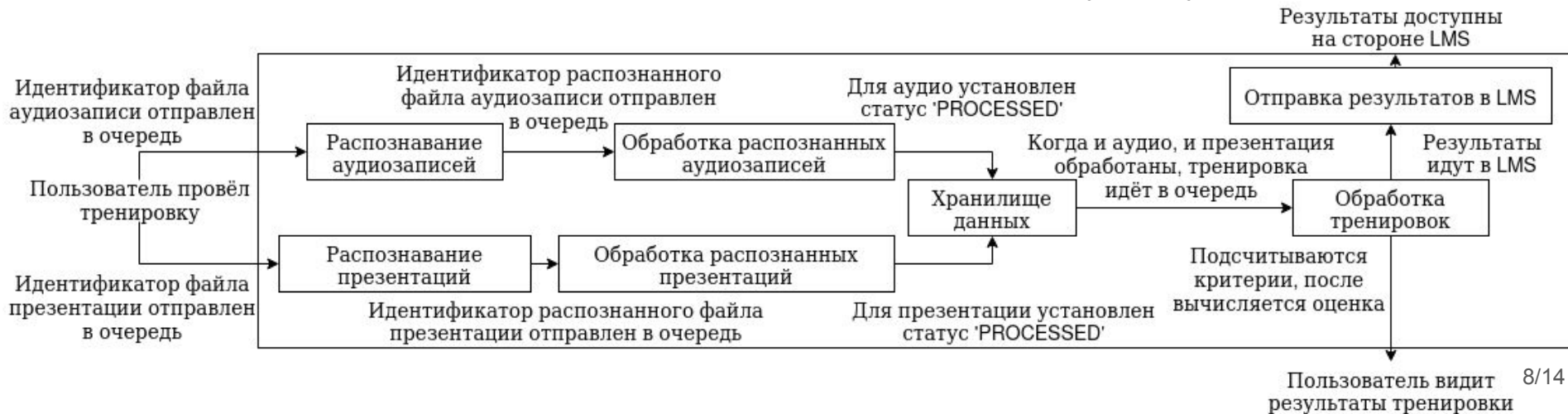
Сущности: аудиозапись, презентация, переключения слайдов.

Состояния: “сырые” данные, “распознанные” и “обработанные”.

Пример критерия: длительность доклада не превышает X. Результат -- дробное число.

Пример параметризованного критерия: длительность доклада не превышает 7 минут.

Объединяем параметризованные критерии в наборы, используем функцию оценки.

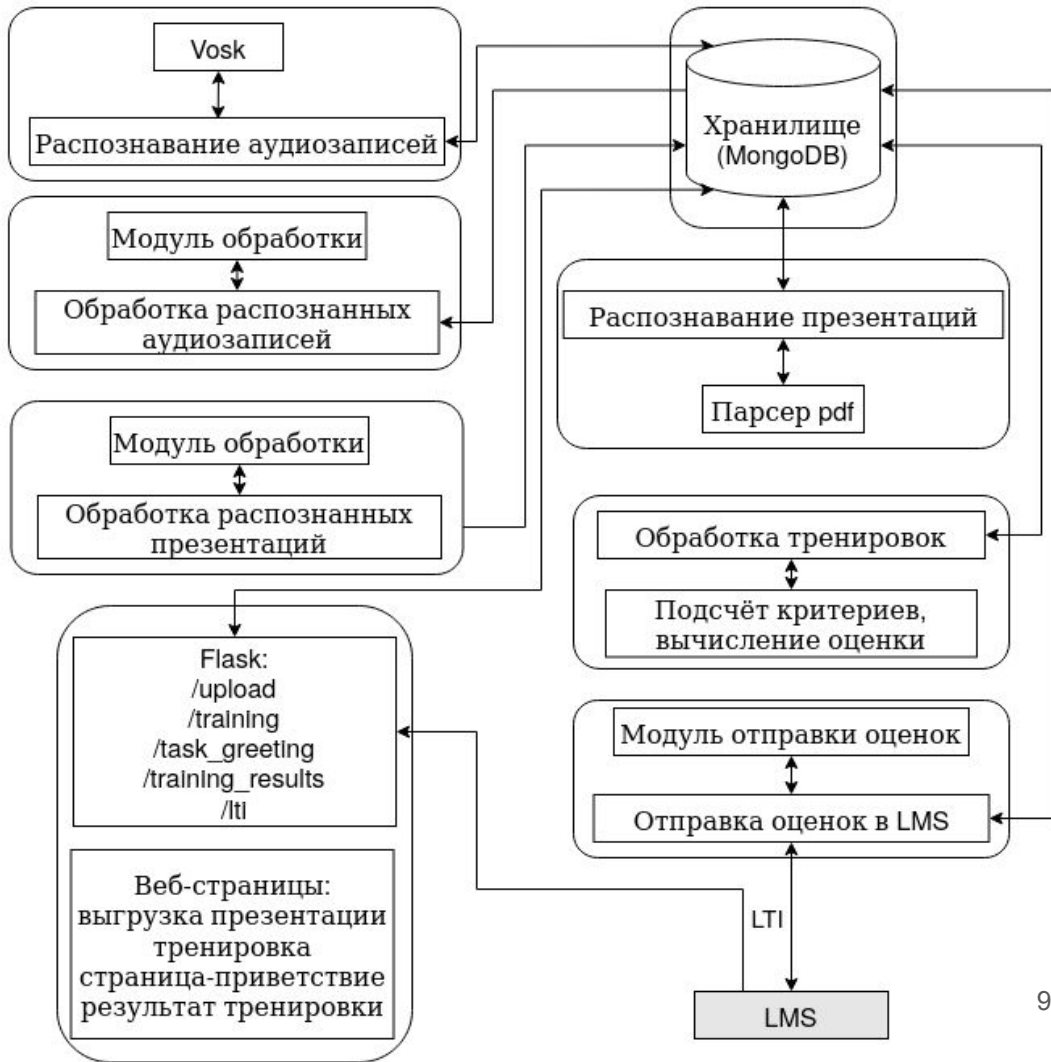


Архитектура

Сервисы: веб, хранилище данных и сервисы-очереди, к которым прикреплены обработчики.

Пользователь видит только веб-сервис.

Хранилище данных также перекладывает из очереди в очередь.



Интеграция с внешними приложениями

LTI (Learning Tools Interoperability) -- протокол, поддерживаемый многими LMS и MOOK платформами.

Суть: POST запрос с информацией о задаче и пользователе, а также данными для авторизации и ссылкой для передачи оценки обратно.

REST API -- набор методов для взаимодействия с приложением.

Были реализованы группы методов для:

- тренировок
- попыток
- файлов

Интерфейс

- Выгрузить презентацию
- Включить запись (Record)
- Проговорить доклад, переключая слайды (Next)
- Закончить тренировку (Done), дождаться оценки



Статистика тренировки с ID: 60a4ec04de901c8b5a7dede6

Оценка за тренировку = 0.61 = 0.600 * 0.69 + 0.200 * 1.00 + 0.200 * 0.00

StrictSpeechDurationCriterion = 0.69

SpeechPaceCriterion = 1.00, Оценки по слайдам:

Слайд 1: оценка = 0.98, слов в минуту = 73.66, слов сказано 16 за 00:13.

Слайд 2: оценка = 1.00, слов в минуту = 90.19, слов сказано 59 за 00:39.

Слайд 3: оценка = 1.00, слов в минуту = 77.99, слов сказано 72 за 00:55.

...

FillersNumberCriterion = 0.00, Использование слов-паразитов по слайдам:

Слайд 5: [значит, это, то есть, это, это, это, ну, то есть].

Слайд 6: [то есть, так].

Слайд 7: [значит, значит, это, значит, то есть, просто, ну]

...

id тренировки: 60a4ec04de901c8b5a7dede6. Статус: Обработана

id презентации: 60a4ec04de901c8b5a7dedde. Статус: Обработана

Название презентации: предзащита.pdf

id аудиозаписи: 60a4f108de901c8b5a7dee04. Статус: Обработана



[Сообщить о багах / проблемах сайта](#)

Тестирование нагрузки

Было взято 12 копий одной и той же тренировки длительностью 7:39 и простой (длительность, темп, слова-паразиты) набор критериев.

В такой конфигурации самая долгая часть -- распознавание аудиозаписи, время обработки тренировки почти равно времени распознавания аудиозаписи.

Количество обработчиков распознавания речи	Суммарное время обработки всех тренировок, мм:сс	Длительность тренировки, обработанная за 1с работы приложения, с
1	25:31	3.60
2	17:47	5.16
3	14:31	6.32

Апробация

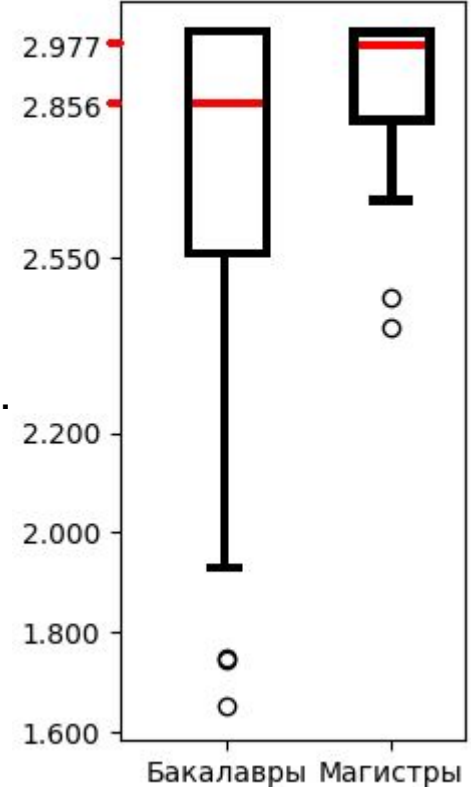
Приложение использовалось студентами (~80 человек) МОЭВМ ЛЭТИ как один из шагов аттестации за преддипломную практику, доступ через курс на Moodle ¹.

Было взято 5 случайных тренировок (из ~350 проведённых).

Средняя разница количества слов -- 4.23%.

Средняя разница темпа речи -- 4.50%.

Магистры в целом получают баллы больше.

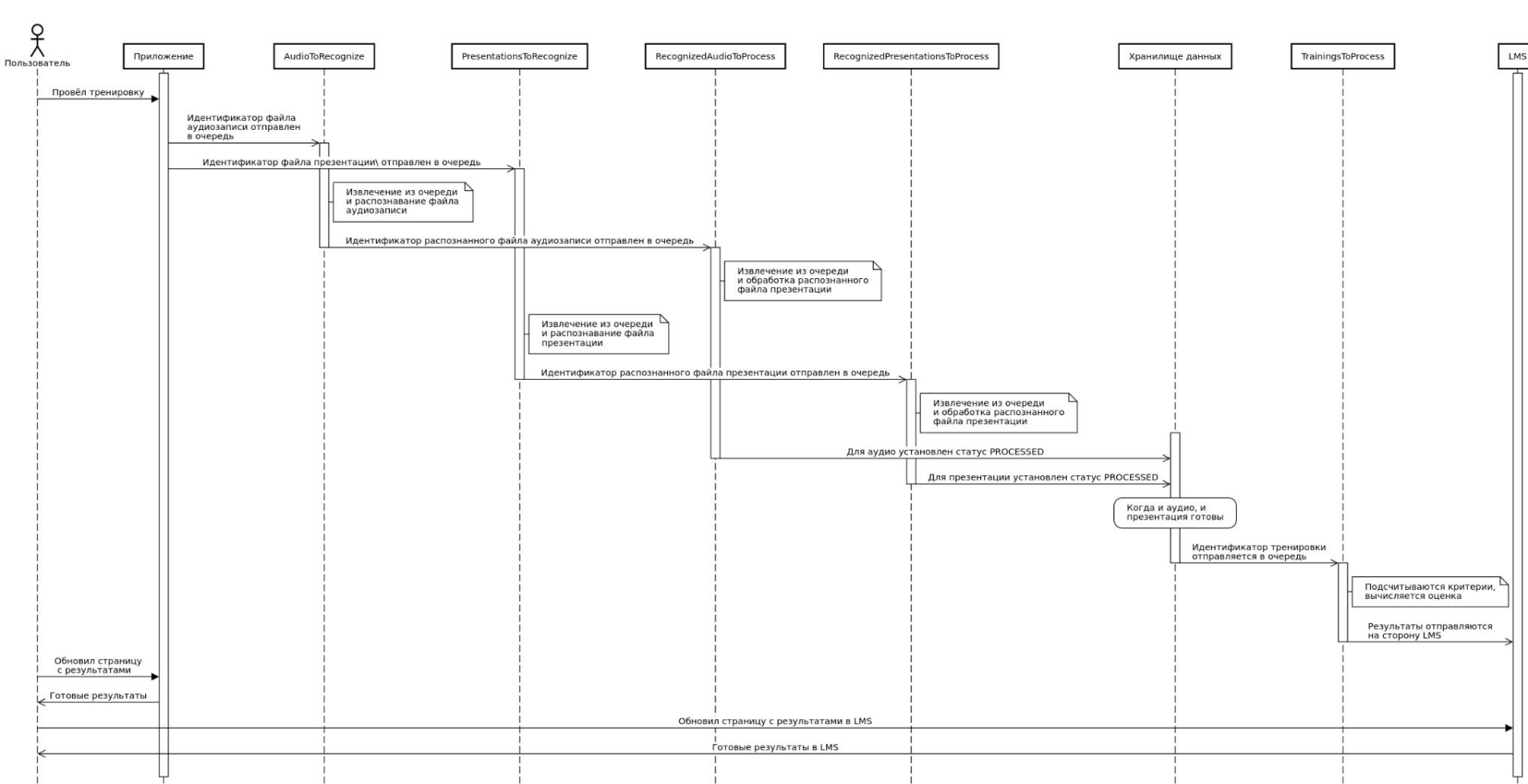


¹ <http://e.moevm.info/enrol/index.php?id=39>

Результаты

- Проведён обзор литературы и приложений на тему автоматической оценки публичного выступления и сравнение библиотек распознавания речи на собранном наборе данных
- Предложена архитектура приложения, учитывающая возможность изменения критериев и способов обработки данных
- Было разработано веб-приложение для автоматической оценки публичного выступления ¹ с реализованным REST API и поддержкой LTI
- Статья представлена на конференции FRUCT29 (май 2021)

¹ Репозиторий: https://github.com/OSLL/web_speech_trainer



Улучшения

Использовать специализированное ПО для очередей

- Нагрузка не такая большая, хватает запросов “самый старый объект” к коллекциям в БД

Сделать специализированные критерии

- Тогда, к примеру, критерии, зависящие только от аудиозаписи вычислять сразу, не дожидаясь готовности всей тренировки

Сравнение библиотек распознавания речи

Требования: бесплатная, офлайн, есть временные метки.

Данные для сравнения: собрал 10 выступлений, из них 5 -- защиты дипломов, ещё 5 -- уроки на Stepik.

Для получения результатов использовал <https://utext.rikuz.com/>