

Разработка веб-тренажёра публичных выступлений

Даниил Плющенко

(руководитель: Марк Маркович Заславский, JetBrains Research)

Введение

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

К примеру, (пред-) защиты курсовых и дипломов, выступления на конференциях

Им также приходится тренироваться перед выступлением

Чтобы оценить выступление, нужны квалификация в предметной области и время для прослушивания, анализа и оценки доклада

Введение

Однако всегда можно есть набор критериев и свойств, которые могут быть оценены автоматически

Примеры:

- длительность доклада
- скорость речи докладчика
- порядок секций в презентации
- наличие нумерации слайдов
- наличие просторечий или “слов-паразитов”
- ...

LTI

Learning Tools Interoperability -- протокол, описывающий взаимодействие учебных/образовательных систем.

С его помощью можно “соединять” образовательные системы, например, проходить авторизацию в одной системе и работать во второй, а также пересылать результаты выполнения задания.

Stepik, edX, Moodle, Blackboard LMS и другие системы поддерживают этот протокол.

Цели и задачи

Цель - разработать модульное, а также масштабируемое с точки зрения нагрузки, новых шагов оценивания и конфигураций веб-приложение для подготовки к публичным выступлениям и первичной оценки его качества

Задачи:

- Провести обзор аналогов
- Провести обзор библиотек распознавания речи
- Разработать расширяемую архитектуру, позволяющую добавлять и заменять компоненты, связанные с хранением и анализом публичного выступления
- Разработать веб-приложение в соответствии с предложенной архитектурой

Аналоги

Были рассмотрены приложения и сервисы, помогающие улучшить речь и/или навыки публичных выступлений, выдающие некоторую оценку

Общие выводы:

- в основном поддерживается только английский язык
- нет возможности прикрепить презентацию
- нельзя настроить критерии
- мобильные приложения платформозависимы
- не поддерживается LTI

В 2018-2019 году велась разработка мобильного приложения с подобной целью. Приложение может показывать, сколько времени занял каждый слайд, сколько слов было сказано во время каждого слайда и с какой скоростью.¹

¹ https://github.com/OSLL/android_public_speech_trainer

Название	Платформы	API?	Возможность прикрепить презентацию	Ограничение на продолжительность речи	Запись речи	Языки	Возможность настроить критерии?	Платное?
Speechace	Web	Да	Нет	15 секунд в бесплатной версии	Да	Английский	Нет	Демо
Voice Notebook / Speechpad	Web	Нет	Нет	Нет	Да	8	Нет	Нет
Speakit	Android	Нет	Нет	Одна фраза	Фиксированные слова и фразы	Американский английский	Нет	Нет
Aksent	iOS	Нет	Нет	Одна фраза	Да	20+	Нет	Нет
Speeko	iOS	Нет	Текстовые заметки для текущей записи	Нет	Да	Английский	Нет	Trial
Likeso	iOS	Нет	Нет	30 минут	Да	Английский	Подмножество слов-паразитов	Да
Orai	iOS, Android	Нет	Текстовые заметки	Нет	Да	Английский	Подмножество слов-паразитов	Да
ElsaSpeak	iOS, Android	Нет	Нет	Одна фраза	Фиксированные слова и фразы	Американский английский	Нет	Демо
Говорилло	Android	Нет	Нет	Нет	Да	Русский	Нет	Нет

Библиотеки (сервисы) распознавания речи

Название	Офлайн?	Русский язык?	Временные метки?	Стоимость
Vosk	Да	Да	Да	Бесплатно
Speech-To-Text (Russian)	Да	Да	Нет	Бесплатно
Picovoice	Да	Нет	Нет	Бесплатно
at16k	Да	Нет	Да	Бесплатно
Google Web Speech API	Нет	Да	Да	Бесплатно час в месяц
Google Cloud Speech API	Нет	Да	Да	Бесплатно час в месяц
Microsoft Bing Speech API	Нет	Да	Да	Бесплатно 5000 запросов в месяц
IBM Speech to Text	Нет	Нет	Да	Бесплатно 50 часов нагрузки в месяц
Wit.ai	Нет	Да	Да	Бесплатно

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 похожести текстов. По сути сравнивает между собой набор слов в окне определённого размера.

Vosk показал лучшие результаты, поэтому выбрали его.

Обработка тренировки

Сущности: аудиозапись и презентация

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

Эти данные “перетекают” из очереди в очередь, сейчас это просто таблицы базы данных, из которых извлекаются данные для обработки

Данные -- структуры с дополнительной статистической информацией, которую можно использовать для подсчёта критериев

Обработка тренировки

Пример критерия: длительность доклада не превышает X.

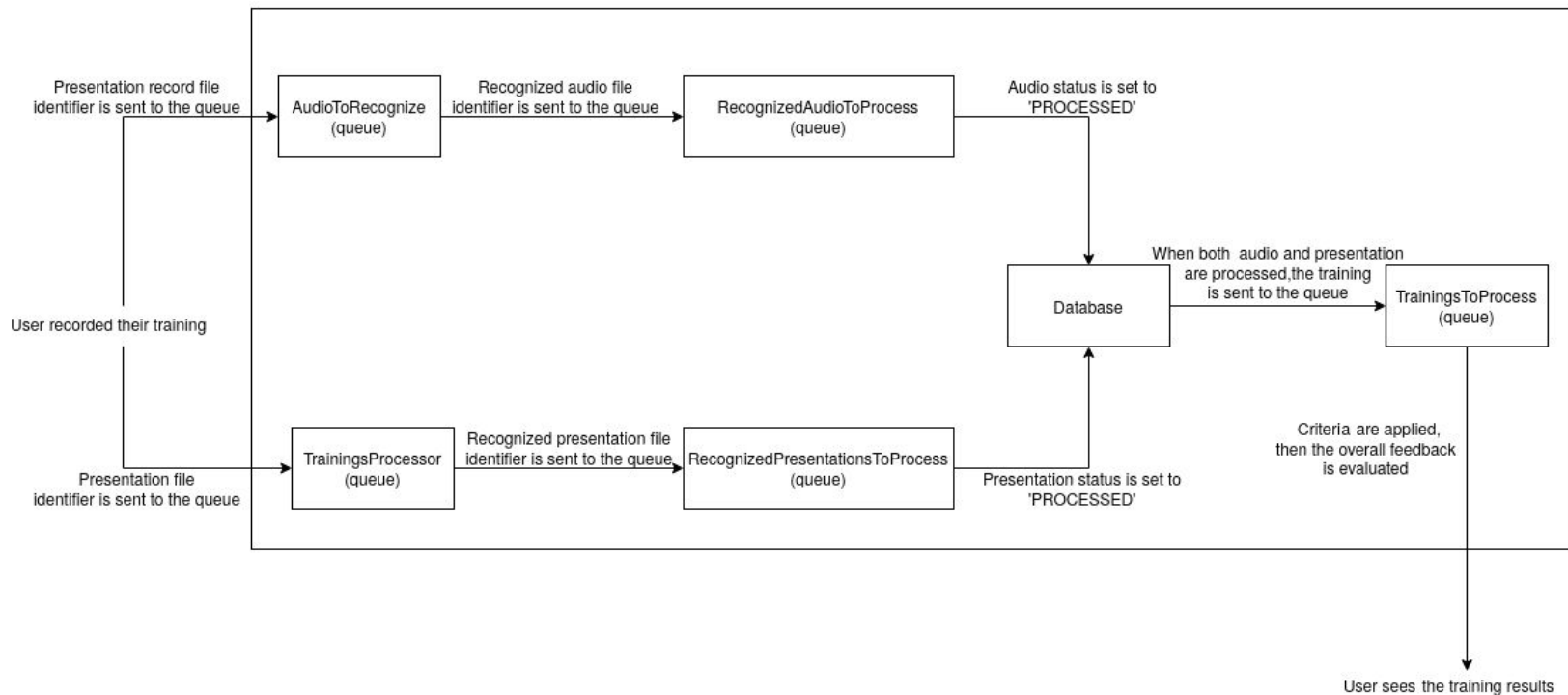
Результат критерия -- дробное число.

Пример параметризованного критерия: длительность доклада не превышает 7 минут.

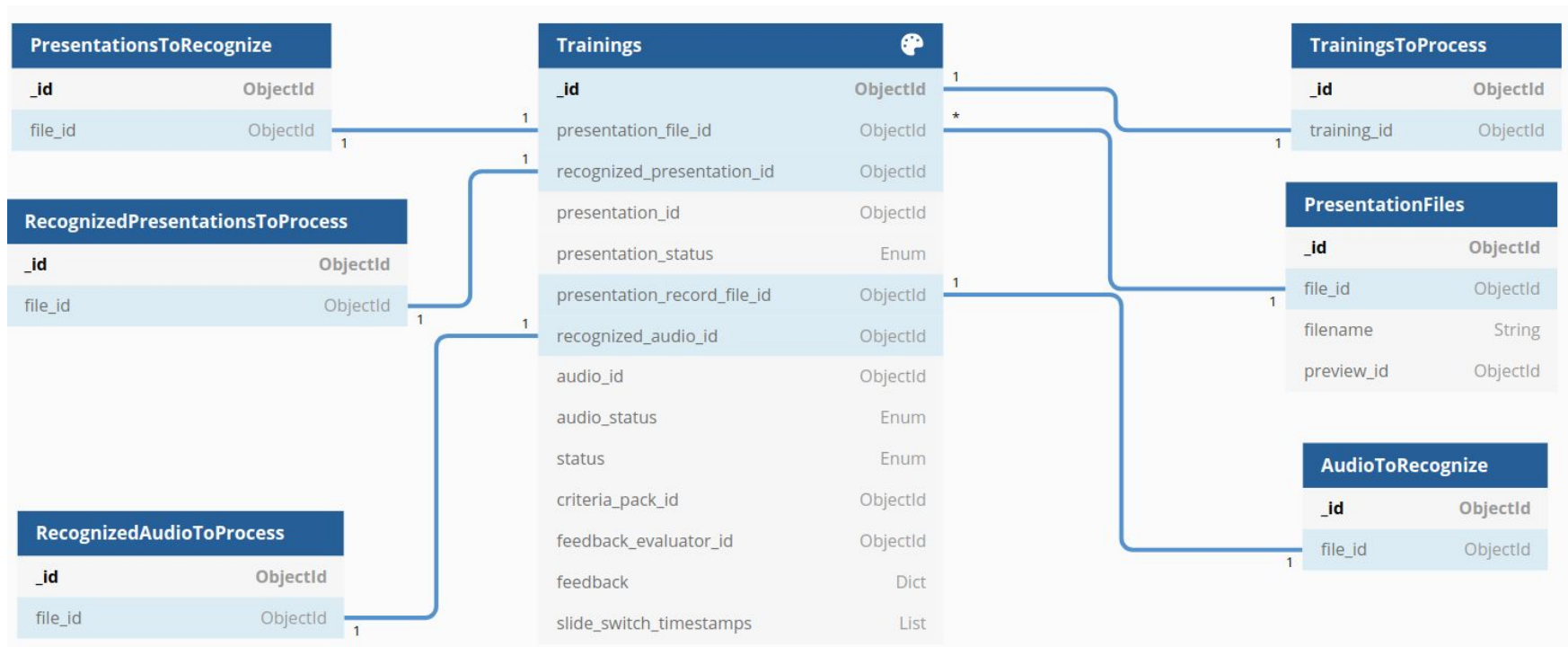
Объединяем параметризованные критерии в наборы с информацией о зависимости критериев друг от друга, вычисляем в нужном порядке (топологическая сортировка или просто заданный порядок)

Используем функцию оценки, которую можем менять. Функция определяет итоговый результат в зависимости от значимости каждого критерия

Обработка тренировки



Структура базы данных



Интерфейс



Record Next Done Page: 2 / 16

Введение

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

К примеру, (пред-) защиты курсовых и дипломов, выступления на конференциях

Им также приходится тренироваться перед выступлением

Чтобы оценить выступление, нужны квалификация в предметной области и время для прослушивания, анализа и оценки доклада

Статистика тренировки с ID: 5fff2f0b92abcad3c0cac199

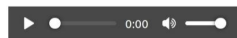
`feedback.score = 1`

Название презентации: Плющенко(1).pdf


id презентации: 5fff2f0b92abcad3c0cac192

id тренировки: 5fff2f0b92abcad3c0cac199

id аудиозаписи: 5fff2f4992abcad3c0cac1a0



Интерфейс

Choose File	No file chosen	Добавить презентацию
Превью	id	
	5fe28bb2729b7390444accf1	Начать тренировку.
Разработка веб-тренажёра публичных выступлений Даниил Площенко (руководитель: Марк Маркович Заславский, JetBrains Research)	5fff2f1a92abcd3c0cac19a	Начать тренировку.

id	Время	Результат
5fd7b05b8b4d1b2c50eead14	Mon, 14 Dec 2020 18:35:07 GMT	1
5fd80b8f0449c02971b56bb8	Tue, 15 Dec 2020 01:04:15 GMT	0.95
5fd929639b5c0c270f254c36	Tue, 15 Dec 2020 21:23:47 GMT	1
5fd92e4177657cc275f978	Tue, 15 Dec 2020 21:44:33 GMT	0.87
5fd934335bbc9880ec5a7c94	Tue, 15 Dec 2020 22:09:55 GMT	1
5fd9362d5bbc9880ec5a7ca3	Tue, 15 Dec 2020 22:18:21 GMT	...

Производительность

На данный момент самая долгая часть -- распознавание аудиозаписи, обработка презентации идёт параллельно, а критерии очень простые и вычисляются очень быстро, поэтому длительность обработки всей тренировки почти совпадает с длительностью распознавания аудиозаписи.

Длительность аудиозаписи, секунды	Длительность распознавания аудиозаписи, секунды	Отклонение
459	144.09	0.49
541	154.87	0.51
436	139.74	0.42
568	168.77	0.68
614	226.37	0.82
492	155.40	0.63
810	232.43	3.14
1151	324.33	5.55

Количество слайдов	Длительность парсинга презентации, секунды	Отклонение
10	1.52	0.03
10	1.55	0.09
17	2.48	0.12
18	2.59	0.10
20	2.73	0.04
15	1.97	0.12
9	1.29	0.12
11	1.50	0.12

Результаты

- Проведен обзор библиотек распознавания речи и аналогов
- Предложена архитектура приложения
- Разработан прототип веб-приложения

Будущая работа

- Поддержка LTI
- Добавить больше критериев
- Улучшить интерфейс, в т.ч. преподавательский
- Более осмысленно измерить производительность, в т.ч. при нагрузке