Мониторинг удовлетворенности качеством жизни горожанина мегаполиса

Вячеслав Гораш (Р4140)

Екатерина Машина (Р4141)

Марк Шерман (Р4140)



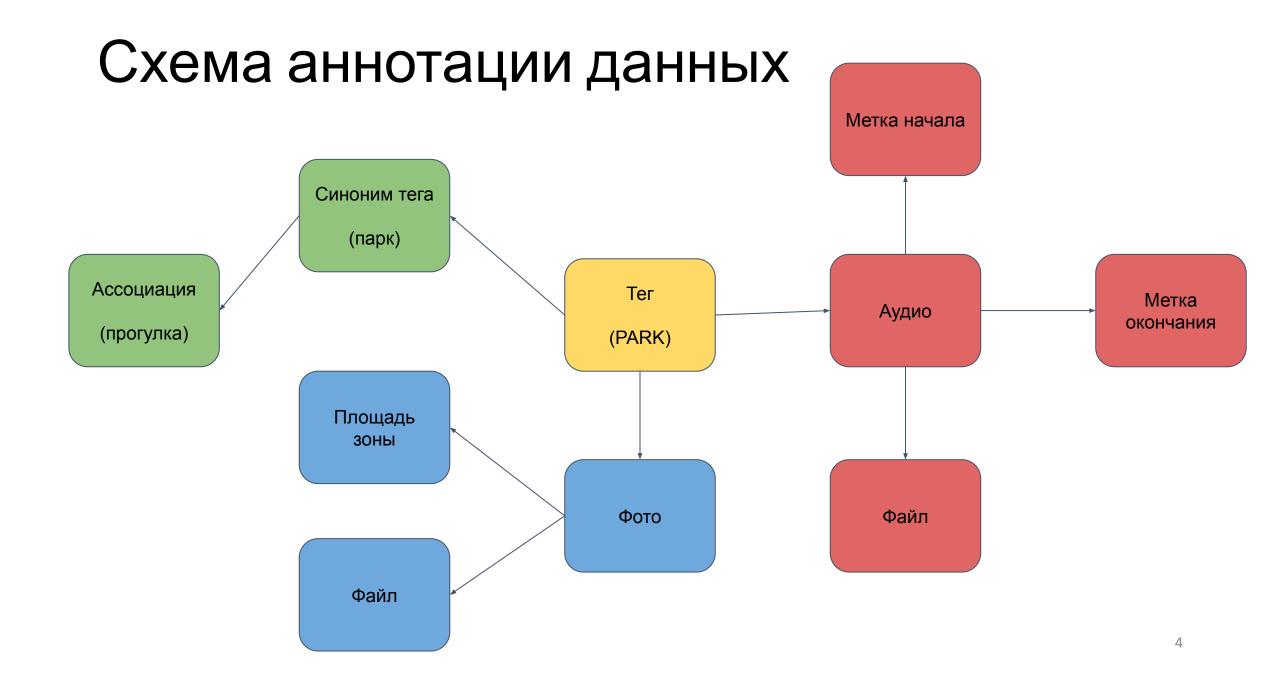
Цель и задачи проекта

- **Целью** работы является автоматизация процесса сбора информации, позволяющей определить качество жизни жителя мегаполиса.
- Задачами работы является:
 - выявление комментариев горожан, связанных с их личной оценкой качества жизни в конкретном мегаполисе,
 - расчет общей площади зеленых насаждений мегаполиса,
 - выявление звукового фона и шума во дворах, полученного из объективных средств фиксации,
 - Объединение трех модальностей данных в один граф знаний.

Методы сбора данных

- Отзывы горожан о качестве жизни в мегаполисе через социальные сети и приложения.
- Спутниковые снимки для расчета площади зеленых насаждений.
- Аудиофайлы с записью звукового фона и шума во дворах.





Сбор и разметка данных текстовой модальности

- Выбрано подходящее сообщество в ВКонтакте
- Использовать специальные инструменты API для сбора данных
- Проведена разметка данных, выделены ключевые сущности и связанные с ними ассоциации
- Для разметки использованы инструменты stanza
- Качество сбора и разметки данных напрямую влияет на результаты исследования

```
<word/дворик> ns1:association <word/Mосква>,
                     <word/mecto> .
  <word/завод> ns1:association <word/Mocква>,
                      <word/HN3>,
                       <word/год>,
                  <word/московский> .
<word/переулок> ns1:association <word/Mocква>,
                      <word/год> .
<word/склад> ns1:association <word/возгорание>,
                   <word/квадратный>,
                      <word/memp>,
                    <word/площадь>,
                     <word/nomap> .
  <word/фабрика> ns1:association <word/MEXA>,
                     <word/Mосква>,
                     <word/Славян>,
                       <word/TU>,
                      <word/банк>,
```

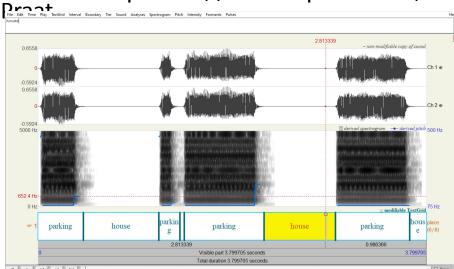


Сбор и разметка данных графической модальности

- Для разметки изображений был выбран инструмент CVAT
- На изображениях были выделены объекты, такие как здания, дороги, зеленые насаждения и т.д.
- Размеченные данные могут быть использованы для обучения моделей компьютерного зрения, например, для распознавания объектов на снимках со спутника
- CVAT позволяет работать в команде над разметкой данных, что ускоряет процесс и повышает качество разметки

Сбор и разметка данных звуковой модальности

- •Данные были собраны самостоятельно авторами исследования в Академическом районе г. Москвы.
- Размер собранного датасета 15 записей
- Разметка производилась при помощи

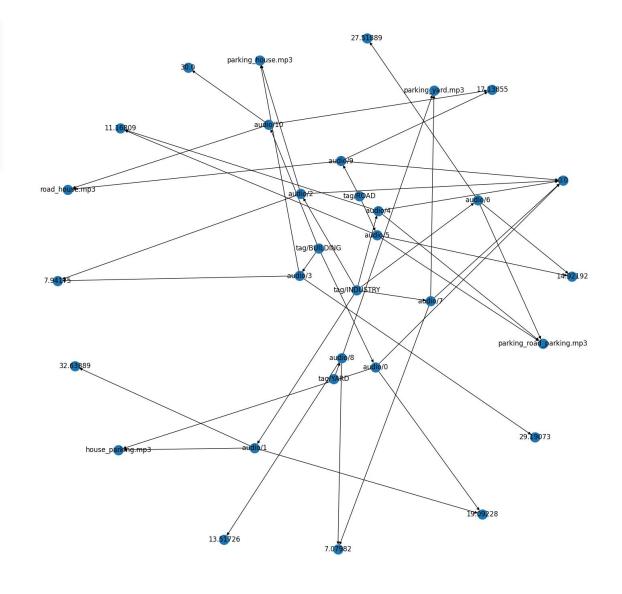


Особенности работы с Praat:

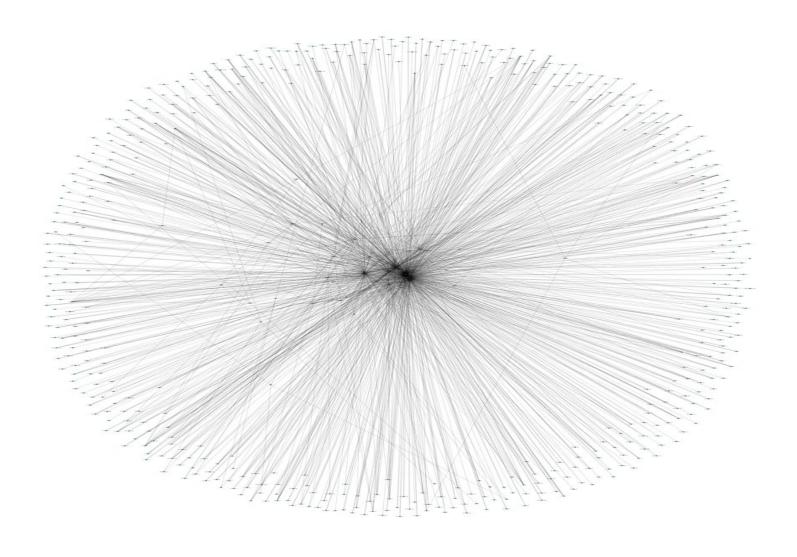
- •возможность разметки сегментов звука по временным интервалам,
- •возможность добавления различных типов меток (например, фоновых звуков, голосовых команд, громких звуков),
- •автоматическая разметка на основе алгоритмов машинного обучения,
- •возможность ручной корректировки автоматической разметки,
- •встроенные инструменты для анализа и визуализации размеченных данных.

Построение аннотационной схемы (графа знаний)

- Объединение трех модальностей данных в один граф знаний.
- Построение зависимостей между качеством жизни жителей, зелеными насаждениями и звуковым фоном в мегаполисе.
- Получение новых знаний о качестве жизни жителей мегаполиса.



Общий вид графа



Вывод

- Автоматизация процесса сбора информации позволяет более точно определить качество жизни жителей мегаполиса.
- Использование графа знаний позволяет получить новые знания о зависимостях между различными факторами, влияющими на качество жизни жителей мегаполиса.

