

Проектирование сервиса для разметки данных

Лабораторная работа #4

Работу выполнили Атрушкевич А., Буценко С., Радионова Е.

Описание проекта

Цель проекта - создание внутренней системы компании для разметки данных. Это позволит компании самостоятельно получать данные для обучения нейросетей. Проект экономически выгоден компании, ведь ей не придется покупать датасеты для обучения.

Требования клиента

В результате интервью с клиентом были выявлены следующие требования:

- В сервис может войти любой работник компании
- Небольшой бюджет
- Большая скорость разметки
- Простой интерфейс
- Разметка текста и картинок по классам и тэгам
- Тренировочные тесты / примеры
- Определение качества разметки
- Рейтинг разметчика
- Админ для контроля проектов

Анализ конкурентов

- SuperAnnotate - самый популярный сервис на данный момент. Предоставляет мощные инструменты, например, семантическую сегментацию, ключевые точки и рисование линий, которые позволят сэкономить время и повысить точность. Однако интерфейс сервиса не кажется интуитивным.
- Supervisely - имеет более восьмидесяти вариантов аннотирования, упорядоченных в категории, обеспечивает высокую гибкость, предоставляет собственные датасеты.
- Hive Data — имеет удобные функции наподобие возможностей облачного масштабирования.

Все из вышеперечисленных сервисов обеспечивают высокую скорость разметки, имеют огромный функционал, и не только в сфере аннотирования, но на всех этапах машинного обучения.

Описание разрабатываемой системы

Предполагаются следующие участники будущей системы:

- Заказчик
- Разметчик
- Администратор

Разрабатываемая система видится такой:

- Простой сервис, вход в который осуществляется с использованием рабочих аккаунтов сотрудников.
- Интеграции с другими сервисами не предполагается
- Система будет обеспечивать разметчикам доступ к данным, которые нужно разметить и инструменты разметки, а заказчикам - доступ к уже размеченным данным.
- Для разметчика и заказчика интерфейс одинаков. Пользователь может как размечать чужие проекты, так и публиковать свои. Администратор - такой же пользователь, который имеет некоторые привилегии для модерации проектов и пользователей.
- Какое-либо взаимодействие между пользователями внутри системы не предусмотрено, так как это внутренний сервис компании.

Наш продукт должен предоставить пользователям возможность загрузить неразмеченный датасет (в формате текста или картинок), настроить и опубликовать задание, установить желаемое качество разметки (сколько раз каждое задание отображается разным разметчикам), добавить дедлайн, скачать размеченный датасет, просмотреть рейтинг разметчика, убрать разметчика с проекта. С другой стороны должна быть возможность зайти в сервис и быстро начать разметку, по желанию предварительно потренировавшись на 2-3 примерах или просмотрев подсказку. Процесс разметки должен быть максимально быстрым и простым. Также у админа должна быть возможность модерации всех проектов и разметчиков.

Out of scope:

- Более чем 4 необходимых типа разметки
- Самостоятельная оценка разметчиков
- Открытый доступ к заданиям (вне компании)
- Встроенные датасеты
- Интеграция с другими сервисами

Пользовательские истории

С пользовательскими историями и критериями приемки можно ознакомиться на доске [Kanban Flow](#).

Пользовательское взаимодействие и дизайн

Проекты в draw.io и balsamiq приложены.