

Техническое задание на создание сервиса для полуавтоматической разметки товаров

[Заказчик](#)

[Описание проекта](#)

[Функциональные требования](#)

[Загрузка и отображение данных:](#)

[Предоставление вариантов соответствия:](#)

[Выбор соответствия оператором:](#)

[Сохранение разметки:](#)

[Статистика и отчетность:](#)

[Описание данных](#)

[Описание кросскомандного взаимодействия](#)

[Взаимодействие и последовательность выполнения работ](#)

[Используемый стек технологий](#)

[User stories](#)

[Просмотр всех записей](#)

[Разметка товаров](#)

Заказчик

ООО «ПРОСЕПТ» — российская производственная компания, специализирующаяся на выпуске профессиональной химии. В своей работе используют опыт ведущих мировых производителей и сырье крупнейших химических концернов. Производство и логистический центр расположены в непосредственной близости от Санкт-Петербурга, откуда продукция компании поставляется во все регионы России.

Сайт: <https://prosept.ru/>

Описание проекта

Заказчик производит несколько сотен различных товаров бытовой и промышленной химии, а затем продаёт эти товары через дилеров. Дилеры, в свою очередь, занимаются розничной продажей товаров в крупных сетях магазинов и на онлайн площадках.

Для оценки ситуации, управления ценами и бизнесом в целом, заказчик периодически собирает информацию о том, как дилеры продают их товар. Для этого они парсят сайты дилеров, а затем сопоставляют товары и цены. Зачастую описание товаров на сайтах дилеров отличаются от того описания, что даёт заказчик. Например, могут добавляться новые слова ("универсальный", "эффективный"), объём (0.6 л -> 600 мл). Поэтому сопоставление товаров дилеров с товарами производителя делается вручную.

Цель этого проекта - разработка решения, которое отчасти автоматизирует процесс сопоставления товаров. Основная идея - предлагать несколько товаров заказчика, которые с наибольшей вероятностью соответствуют размечаемому товару дилера. Предлагается реализовать это решение, как онлайн сервис, открываемый в веб-браузере. Выбор наиболее вероятных подсказок делается методами машинного обучения.

Функциональные требования

Загрузка и отображение данных:

Оператор должен иметь возможность загрузить данные для разметки из существующей БД заказчика.

Система должна отображать загруженные данные оператору.

Предоставление вариантов соответствия:

Система должна использовать рекомендательную модель, созданную ML специалистами, для предоставления наиболее вероятных вариантов соответствия товара.

Для каждого товара система должна предложить несколько (настраиваемое количество) вариантов соответствия.

Выбор соответствия оператором:

Оператор должен иметь возможность выбрать один из предложенных вариантов, который соответствует необходимому товару. Или ни одного.

Должна быть предусмотрена возможность снятия выбора или замены выбранного варианта.

Сохранение разметки:

Система должна сохранять выбранный оператором вариант соответствия для каждого товара.

Статистика и отчетность:

Сервис должен предоставлять оператору статистику по размеченным товарам (количество и т.д.).

Оператор должен иметь доступ к отчетам и аналитике по работе.

Система сохраняет статистику выбора оператором предложенных вариантов (какой по счёту из предложенных вариантов выбран, как часто ни один вариант не выбран).

Интерфейс для оператора:

Интерфейс должен быть интуитивно понятным и удобным для оператора.

Описание данных

Заказчик предоставил несколько таблиц (дамп БД), содержащих необходимые данные:

1. *marketing_dealer* - список дилеров;
2. *marketing_dealerprice* - результат работы парсера площадок дилеров:
 - *product_key* - уникальный номер позиции;
 - *price* - цена;
 - *product_url* - адрес страницы, откуда собраны данные;
 - *product_name* - заголовок продаваемого товара;
 - *date* - дата получения информации;
 - *dealer_id* - идентификатор дилера (внешний ключ к *marketing_dealer*)
3. *marketing_product* - список товаров, которые производит и распространяет заказчик;
 - *article* - артикул товара;
 - *ean_13* - код товара (см. EAN 13)
 - *name* - название товара;
 - *cost* - стоимость;
 - *min_recommended_price* - рекомендованная минимальная цена;
 - *recommended_price* - рекомендованная цена;
 - *category_id* - категория товара;
 - *ozon_name* - названиет товара на Озоне;
 - *name_1c* - название товара в 1C;
 - *wb_name* - название товара на Wildberries;
 - *ozon_article* - описание для Озон;
 - *wb_article* - артикул для Wildberries;
 - *ym_article* - артикул для Яндекс.Маркета;

4. *marketing_productdealerkey* - таблица матчинга товаров заказчика и товаров дилеров
- *key* - внешний ключ к *marketing_dealerprice*
 - *product_id* - внешний ключ к *marketing_product*
 - *dealer_id* - внешний ключ к *marketing_dealer*

Описание кросс-командного взаимодействия

Выполнение проекта подразумевает кросскомандное взаимодействие:

- ML специалисты - создают сервис для рекомендательной системы;
- Front-end - создают интерфейс, в котором работает оператор;
- Back-end - работают над основой, которая объединяет сервис и интерфейс;

Взаимодействие и последовательность выполнения работ

Специалисты работают параллельно, что позволяет ускорить процесс разработки и быстрее выпустить работающее приложение.

ML специалисты:

- Разработка рекомендательной модели на основе данных о товарах.
- Интеграция модели с бэкендом для предоставления вариантов соответствия.
- Тестирование и оптимизация модели для достижения высокой точности.

Backend разработчики:

- Создание бэкенда для обработки запросов от фронтенда и интеграции с ML моделью.
- Реализация системы сохранения и управления разметкой товаров.
- Обеспечение безопасности и авторизации пользователей.
- Создание API для взаимодействия с фронтендом.

Frontend разработчики:

- Создание интерфейса для оператора, включая загрузку данных, отображение вариантов соответствия и выбор разметки.
- Интеграция с бэкендом для отправки запросов на разметку и получения статистики.
- Реализация функциональности отображения и редактирования разметки.

Тестирование и оптимизация

- Проведение тестирования на различных этапах разработки для выявления и устранения ошибок.
- Тестирование с участием операторов (с привлечением специалистов заказчика) для оценки удобства использования.

Деплой и мониторинг:

- Развертывание сервиса на серверах - на стороне заказчика (мы предоставляем скрипт и инструкцию)
- Компоненты системы пишут логи событий и ошибок в стандартные потоки.

Используемый стек технологий

Фронтенд:

Для фронтенд-разработчиков: React

Бэкенд:

Язык программирования: Python

Фреймворк для веб-приложений: Django

Фреймворк для API: FastAPI

База данных: PostgreSQL или MySQL

Машинное обучение:

Библиотеки для ML: Pandas, Sklearn, градиентный бустинг (CatBoost, LightGBM, XGBoost), Keras, PyTorch

Другие инструменты:

Git, Docker, NGINX

User stories

Просмотр всех записей

1. Пользователь открывает страницу сервиса через веб-браузер и видит главную страницу. На этой странице отображена таблица товаров продавцов, которые есть в БД. В эту же таблицу выводится дополнительная информация по товарам продавцов - дата получения записи, статус (есть сопоставление с товаром производителя или нет), с каким товаром производителя сопоставлена, др. полезную информацию.

2. У пользователя есть возможность отфильтровать записи таблицы по дате, статусу, др.

Разметка товаров

1. Пользователь может перейти в режим разметки товаров. В этом режиме он видит один товар продавцов и несколько вариантов соответствия с товарами производителя.
2. Пользователь может выбрать один из предложенных товаров, либо попытаться найти во всём списке товаров производителя, либо указать что ничего не подходит.
3. После этого пользователь переходит к разметке следующего товара, либо переходит на страницу просмотра всех записей.