

Видение для сервиса разметки

Описание продукта

Что представляет собой продукт

Продукт представляет собой веб-приложение Сервис Разметки, которое предназначено для предоставления пользователям возможностей по добавлению и разметке датасетов с использованием как простой, так и сложной разметок.

Какую проблему и для кого он решает

Веб-приложение Сервис Разметки предназначено для решения проблемы отсутствия размеченных данных.

Сервис Разметки решает данную проблему для компании, создающей модели на основе машинного обучения.

Цели и задачи проекта

Цель проекта – создание веб-приложения для разметки данных. Это позволит создавать обучающие наборы для прокачки машинного обучения, на основе размеченных данных.

Задачи:

- Создание сервиса с возможностью переключения между рабочим пространством разметчика и заказчика;
- Создание рабочего пространства разметчика со списком доступных и уже взятых задач, которые он может брать, выполнять и покидать;
- Создание личного кабинета заказчика, где тот может создавать, редактировать, закрывать, удалять задания для разметки датасета;
- Создание начальной версии сервиса;
- Дальнейшая доработка веб-приложения в зависимости от результатов внедрения.

Общие требования

Сервис Разметки предназначен для использования внутри компании ее сотрудниками, выступающими в качестве пользователей конечного продукта. Сотрудники могут выступать в ролях как разметчика, так и заказчика:

- Разметчик – сотрудник компании, который выполняет задания заказчика;
- Заказчик – сотрудник компании, который создает задания для разметки данных.

Разрабатываемая система видится такой:

- Сервис должен работать быстро – возможность относительно быстро загрузить данные заказчику для разметки;
- Интерфейс сервиса интуитивно понятный;
- Система будет содержать задания и детальную информацию о каждом из заданий как для роли разметчика, так и для роли заказчика:
 - Название задания;
 - Краткое описание задания;
 - Дедлайн;
 - Тип задания;
 - Пример выполнения задания;
 - Количество оценок.
- Система будет содержать следующую информацию о задании для роли разметчик:
 - Кто добавил задание;
 - Сколько данных осталось разметить (в процентах).
- Система будет содержать следующую информацию о задании для роли заказчик:
 - Прогресс выполнения задания в процентах;
 - Датасет.
- Система должна отслеживать, кто из разметчиков плохо разметил данные и блокировать его;
- Не размеченные задания, картинки и текст подкидываются разметчикам в первую очередь;
- После дедлайна проект становится неактивным или невидимым для остальных;

- На главной странице в виде каталога размещены разные датасеты;
- Система должна проверять качество оценки на основе следующих критериев:
 - Если есть тренировочный набор, то система должна сравнивать ответы разметчика с правильными ответами заказчика. Если из тренировочных данных разметчик ошибся более, чем в половине, то его следует больше не допускать к разметке;
 - Если тренировочного набора нет, то система должна сравнивать количество ответов среди разметчиков;
 - Если более 50% одинаково разметили данные, то система будет принимать так размеченные данные за правильный ответ;
 - Если разметчики одинаково разметили данные, то система должна спорные размеченные данные подсовывать другим разметчикам в первую очередь. Если в таком случае заказчик закрывает задание, то система должна выбрать наугад.
- Сервис должен позволять осуществлять разметку следующих типов:
 - Простой – классификация;
 - Сложный – теги.
- Сервис должен работать с данными следующих типов:
 - Изображения;
 - Текст.
- Заказчик должен иметь возможность добавить минимальный коэффициент;
- Количество разметчиков, задействованных для работы над заданием, не ограничено;
- Дедлайн контролируется заказчиком, система предупреждает разметчиков и заказчиков о надвигающемся дедлайне;
- Если в течении какого-то промежутка времени разметчик плохо проявлял себя, то его следует больше не допускать к разметке;
- Если набран лимит по количеству оценок, то задание из каталога не пропадает. Закрывать задание может только заказчик;

- Разметчик может взять максимум 5 заданий;
- Пользователь по умолчанию заходит в систему как разметчик. Любой сотрудник компании может быть как заказчиком, так и разметчиком;
- Пользователь сервиса разметки может поменять пространство – стать заказчиком или разметчиком;
- Система регистрирует то, как разметчик работает и есть ли необходимость убрать его из платформы;
- Для разработки веб-приложения будем использовать Spring Framework, компонент Spring MVC. Он упрощает взаимодействие с большим количеством объектов и управление ими. Предоставляет удобный и эффективный доступ к базе данных;
- Загружаемые и выгружаемые данные имеют формат json;
- Сервис не имеет ограничений по объему загружаемых данных;
- Минимальное количество оценок для разметки данных: 2;
- Человек вне компании не может быть заказчиком;
- Сервис Разметки предназначен для некоммерческого использования.

Анализ конкурентов

1. [Толока](#);
2. [LabelBox](#).

Что умеют конкуренты

- Визуализация данных разного типа из разных источников; Например, если у файлов датасета разное расширение;
- Поиск, сортировка и фильтрация заданий, возможность поделиться датасетом (как ссылка-приглашение присоединиться к заданию, так и ссылка на обзор, просмотр датасета), группировать датасеты;
- Определение эффективности датасета по встроенным моделям ИИ;
- Легко настраиваемые заказчиком задания для разметчиков;
- Настраиваемый план проверки состояния датасетов;
- Защита интеллектуальной собственности;
- Множество инструментов по улучшению качества разметки, например, Auto-Segment 2.0 – инструмент автоматической разметки;

необходимо взять в рамку область, на которой находится предмет разметивания, а инструмент обведет все объекты внутри рамки.



У конкурентов реализованы версии, портированные на мобильные устройства.

Что обязательно должен уметь наш продукт

- Позволять менять рабочее пространства между ролями заказчика и разметчика;
- Позволять заказчику создавать задание с датасетом;
- Позволять заказчику устанавливать критерий приемки датасета – качество разметки;
- Позволять заказчику добавлять, редактировать информацию о задании, в том числе и дедлайн;
- Позволять разметчику присоединяться к заданию и выполнять его;
- Позволять разметчику выбирать не более 5 заданий;
- Позволять разметчику покидать задание;
- Вычислять разметчиков, которые плохо выполняют задание и блокировать их.

Что пока не должен уметь наш продукт?

- Иметь ограничение по объему загружаемых данных;

- Иметь пользователя с ролью админ, который мог бы разблокировать разметчиков;
- Иметь возможность смены языка интерфейса;
- Проводить онбординг разметчиков;
- Предоставлять возможность взаимодействия с корпоративной почтой;
- Иметь всплывающие подсказки (добро пожаловать, если кликните сюда - увидите это, фильтрация...) для первого входа пользователя.

Описание разрабатываемой системы

Сервис Разметки предназначен для использования внутри компании ее сотрудниками, выступающими в качестве пользователей конечного продукта. Сотрудники могут выступать в ролях как разметчика, так и заказчика.

По умолчанию пользователь сервиса для разметки входит в систему в роли разметчика. Разметчик – сотрудник компании, который выполняет задания заказчика. Разметчик заходит в систему и выбирает не более 5-ти датасетов из каталога заданий. Выбранные датасеты попадают в раздел “Мои задания”. В данном разделе расположены взятые разметчиком датасеты, где тот может либо продолжить размечать задание, либо покинуть его. Если разметчик полностью сделал какое-либо из заданий, то оно автоматически удаляется из раздела “Мои задания”.

Заказчик – сотрудник компании, который создает задания, добавляя описание, датасет, количество оценок, дедлайн, тип задания (картинки / текст, классификация / теги), пример выполнения задания. Заказчик может видеть статистику по своим заданиям и закрыть задание в любой момент.

Пользователь сервиса имеет возможность “сменить пространство”, т.е. перейти из роли разметчика в роль заказчика и наоборот.

Пользовательские истории / сценарии

Пользовательские истории и критерии приемки	
Пользовательская история	Критерий приемки

1. Я, как сотрудник компании, хочу иметь возможность войти в систему.	У сотрудников компании есть свои корпоративные логин и пароль для входа в сервис.
2. Я, как разметчик, хочу видеть каталог со списком всех заданий.	Разметчик по умолчанию появляется на странице каталога со списком доступных заданий.
3. Я, как разметчик, хочу иметь возможность выбрать и разметить любой датасет из каталога.	Нажав кнопку Взять, находясь на странице каталога, разметчику отобразится окно с подробной информацией о задании. Нажав кнопку Взять, находясь на странице с подробной информацией о задании, задание добавится в список Моих заданий и не будет больше отображаться в каталоге.
4. Я, как разметчик, хочу иметь возможность фильтрации заданий по типу данных и типу задания.	Окно для фильтрации отобразится только после нажатия на кнопку Фильтрации. Фильтрация представлена в виде двух RadioButton со значениями “Картинки”, “Текст”; двух Checkbox со значениями “Классификация”, ”Теги”, находящимися под RadioButton со значением “Картинки” Фильтрация выполняется сразу после выбора соответствующей RadioButton или Checkbox.
5. Я, как разметчик, хочу иметь возможность удобного перемещения по каталогу.	Для удобства перемещения по странице каталога, в сервисе реализованы кнопки вверх и вниз. Реализована пагинация для удобного перемещения по страницам каталога.
6. Я, как разметчик, хочу иметь возможность покинуть задание.	Покинуть задание можно будет только после нажатия на кнопку Покинуть, находясь в меню Мои

	задания.
7. Я, как разметчик, хочу иметь возможность продолжить размечать задание.	Вернуться к выполнения задания можно будет только после нажатия кнопки Продолжить, находясь в меню Мои задания.
8. Я, как заказчик, хочу иметь возможность добавлять задание, загружать датасет и устанавливать дедлайн.	<p>Окно добавления задания отобразится только после нажатия на кнопку Добавить.</p> <p>В окне добавления задания должны присутствовать поля для добавления со значениями: “Название”, “Описание”, “Количество оценок”, “Дедлайн”, “Данные”, “Тип задания”. Причем последние два поля реализованы в виде двух RadioButton со значениями “Картинки”, “Текст” и двух RadioButton со значениями “Классификация”, “Теги” соответственно.</p> <p>В добавления задания должны присутствовать область для загрузки примера в форматах png, jpg, svg и область для загрузки датасета формата json.</p> <p>Задание добавится только после нажатия кнопки Добавить задание.</p>
9. Я, как заказчик, хочу иметь возможность переносить дедлайн задания или закрывать задание раньше срока.	Система должна предоставить возможность переноса дедлайна и закрытия задания раньше срока.
10. Я, как заказчик, хочу иметь возможность редактировать информацию задания.	<p>Редактировать задание можно нажав на кнопку “Редактировать” в Обзоре задания. Заказчик имеет возможность редактировать все поля задания:</p> <ul style="list-style-type: none"> ● название ● описание

	<ul style="list-style-type: none"> ● дедлайн ● количество оценок ● тип данных, тип задания ● пример ● датасет <p>Нажав на кнопку “Сохранить” все изменения сохраняются.</p>
11. Я, как заказчик, хочу иметь возможность скачивать размеченные данные.	<p>Кнопка скачать данных находится в каждом окне для подробного просмотра задания.</p> <p>Скачивание данных начнется только после нажатия на кнопку Скачать.</p>
12. Как система, я должна уведомлять пользователей о приближении дедлайна и невозможности добавить задание в список своих заданий (разметчик).	<p>После нажатия разметчиком на кнопку взять при уже имеющихся пяти заданиях на странице Мои задания, система отправляет предупреждение о невозможности добавить задание.</p> <p>В ситуации приближения дедлайна Пользователь уведомляется за неделю, три дня и один день до закрытия задания.</p>
13. Как система, я должна блокировать разметчиков, которые ошибаются более чем в 50% заданий.	<p>Система блокирует разметчиков, которые ошибаются более чем в 50% заданий.</p> <p>Заблокированный разметчик больше не может войти в систему как разметчик.</p> <p>Заблокированный пользователь входит как заказчик без возможности смены пространства.</p>
14. Как система, я должна предоставлять возможность скачать размеченные данные.	<p>После размечивания единицы данных разметчиком, система обновляет данные в датасете.</p> <p>После нажатия кнопки Скачать, система предоставляет данные для скачивания.</p>
15. Как система, я должна	Система обновляет данные в

вносить в датасет задания для тегов, размеченные по принципу совпадения областей.	датасете для каждого случая, когда одна из областей содержит другую или другие область хотя бы на 50%.
---	--

Основные пользовательские сценарии:

Разметчик	<ul style="list-style-type: none"> ● Публичная страница с призывом войти в систему (переход по реферальной ссылке). ● Просмотр главного экрана с каталогом всех активных заданий. ● Фильтрация заданий по типу задания (классификация, теги) и типу данных (картинки, текст). ● Получение уведомлений о приближении дедлайна задания. ● Получение уведомлений о том, что лимит по заданиям набран. ● Просмотр взятых заданий в разделе “Мои задания”. ● Просмотр полного описания задания с информацией о количестве оценок, дедлайне, примере выполнения, имени автора, сколько процентов осталось разметить. ● Продолжить разметку задания с того момента, на котором остановился. ● Покинуть задание. ● Сменить роль разметчика на роль заказчика.
Заказчик	<ul style="list-style-type: none"> ● Просмотр главного экрана добавленных заказчиком заданий с описанием задания, дедлайном, статистикой и возможностью более подробного просмотра. ● Подробный просмотр задания с названием, описанием, количеством оценок, дедлайном, типом задания, прогрессом. ● Заккрыть задание. ● Скачать размеченный датасет. ● Редактировать задание. ● Добавить новое задание с загрузкой датасета, примера (по желанию), названием, описанием, количеством оценок, дедлайном, типом данных,

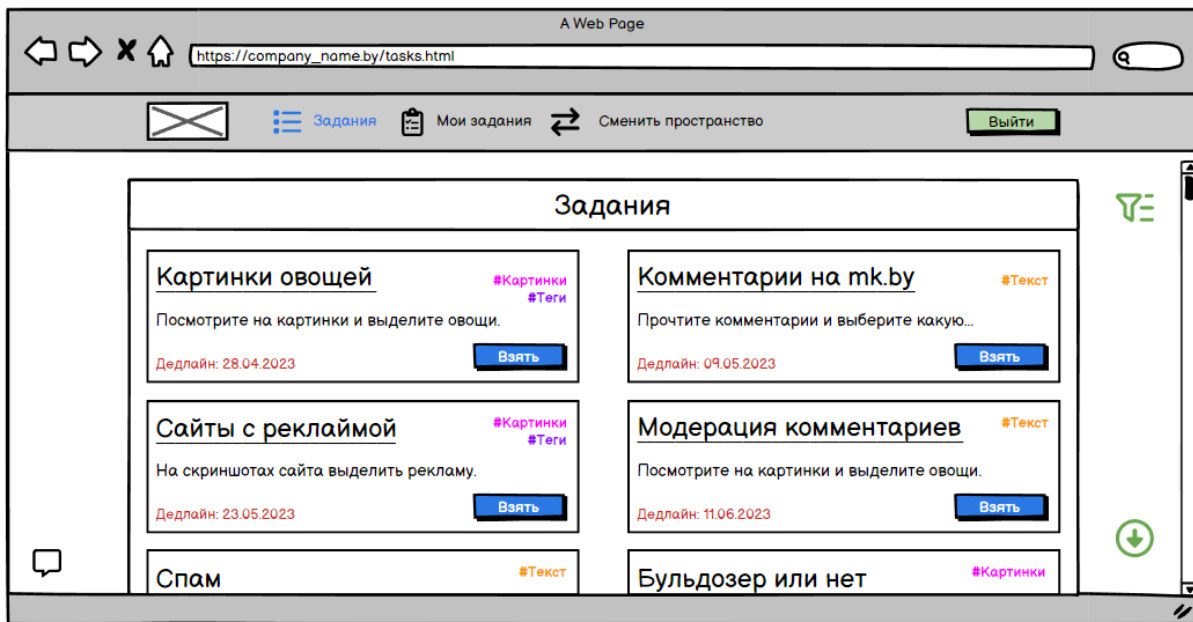
	<p>типом задания.</p> <ul style="list-style-type: none"> ● Получение уведомлений о приближении дедлайна задания. ● Получение уведомлений о выполненных заданиях. ● Просмотр архива заданий. ● Сменить роль разметчика на роль заказчика.
Система	<ul style="list-style-type: none"> ● Контроль выполнения заданий разметчиками. ● Блокировка разметчиков. ● Контроль оценки заданий.

Пользовательский интерфейс

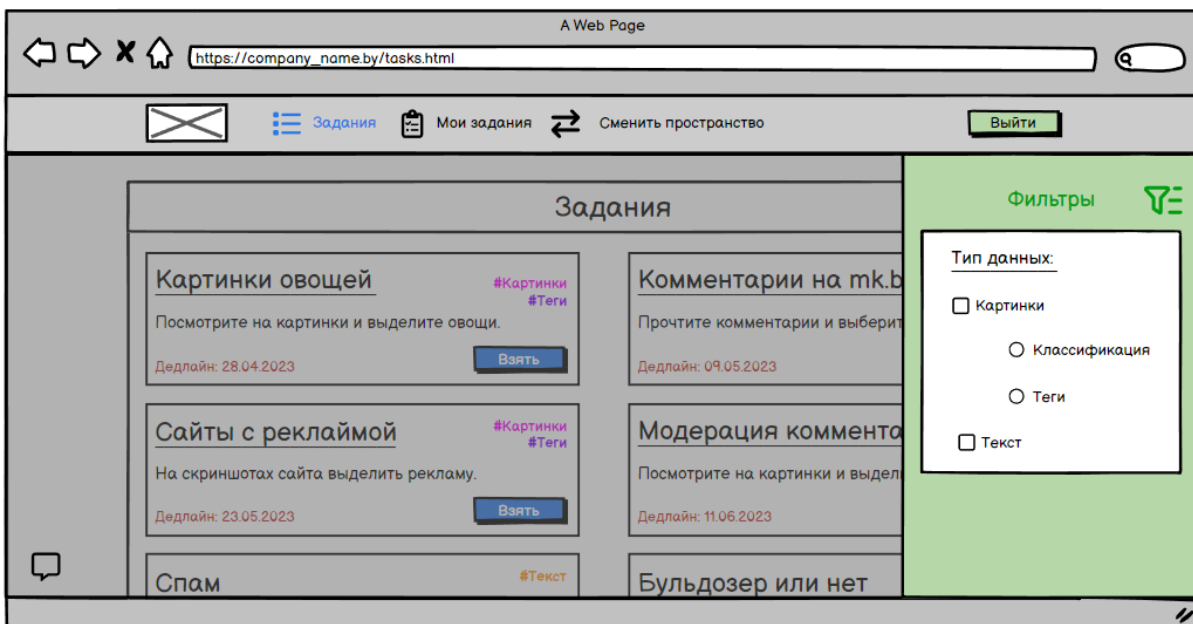
Пользователь открывает сервис по реферальной ссылке и попадает на страницу входа, где может ввести свои корпоративные логин и пароль для входа в сервис.



The image shows a web browser window with the title 'A Web Page'. The address bar contains the URL 'https://company_name.by/log-in.html'. The main content area has a gray background with the word 'Вход' (Login) centered. Below it is a white rectangular form with two input fields: 'Логин' (Login) and 'Пароль' (Password). To the right of the password field is a circular button with a right-pointing arrow. The browser window has standard navigation buttons (back, forward, home, search) and a search icon in the top right corner.

По умолчанию пользователь входит в сервис в роли разметчика. После входа разметчик попадает на главную страницу, где располагается каталог доступных заданий, которые можно взять на выполнение.

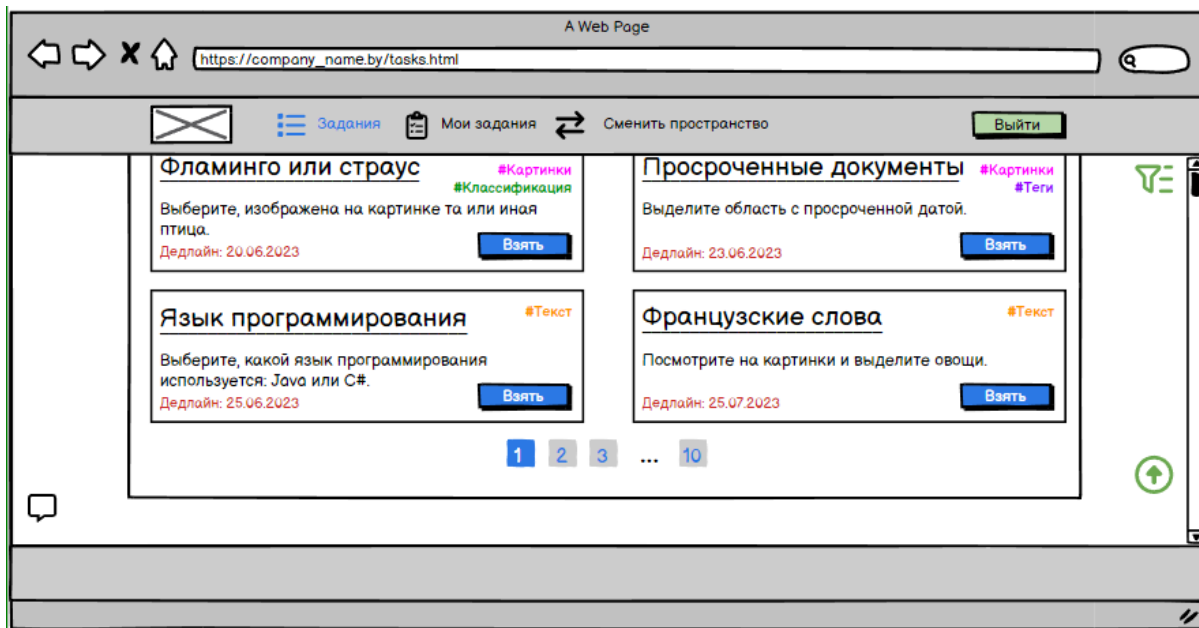


По каталогу можно провести фильтрацию заданий по типу данных и типу самих заданий (классификация, теги).

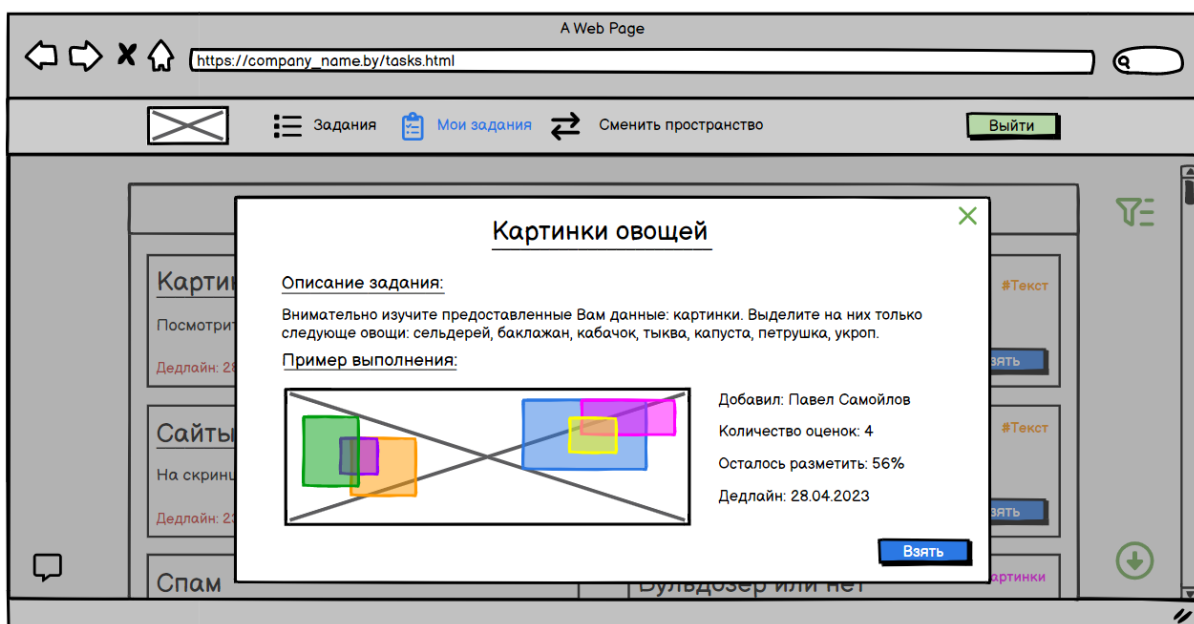


Для удобства перемещения по странице каталога, в сервисе реализованы кнопки вверх  и вниз .

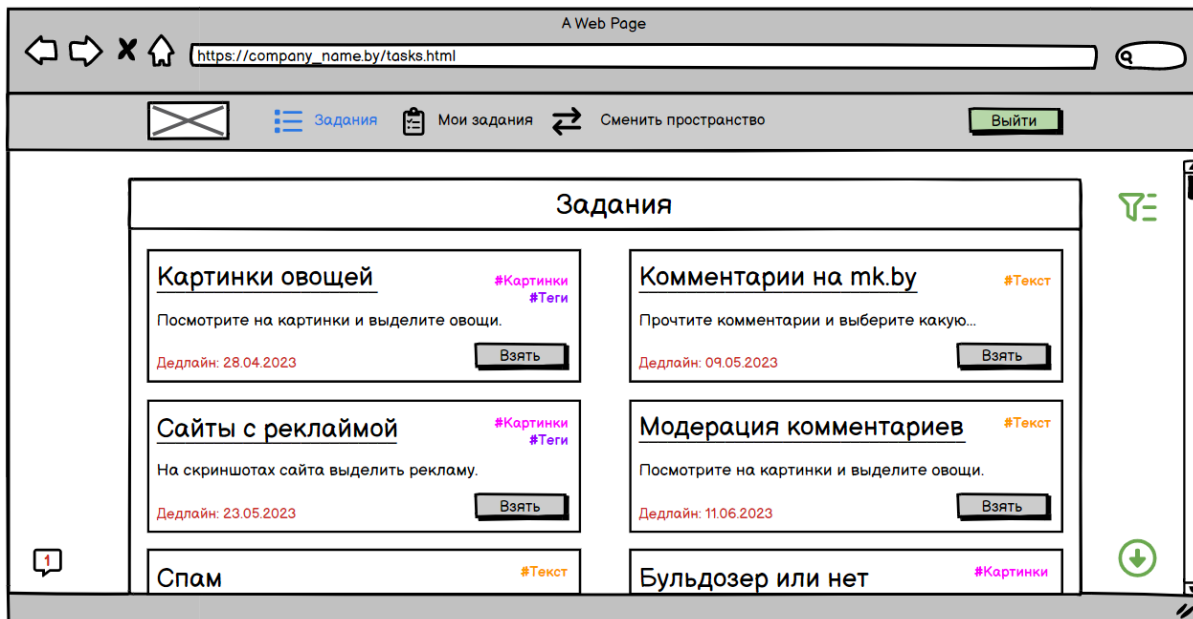
Реализована пагинация для удобного перемещения по страницам каталога.



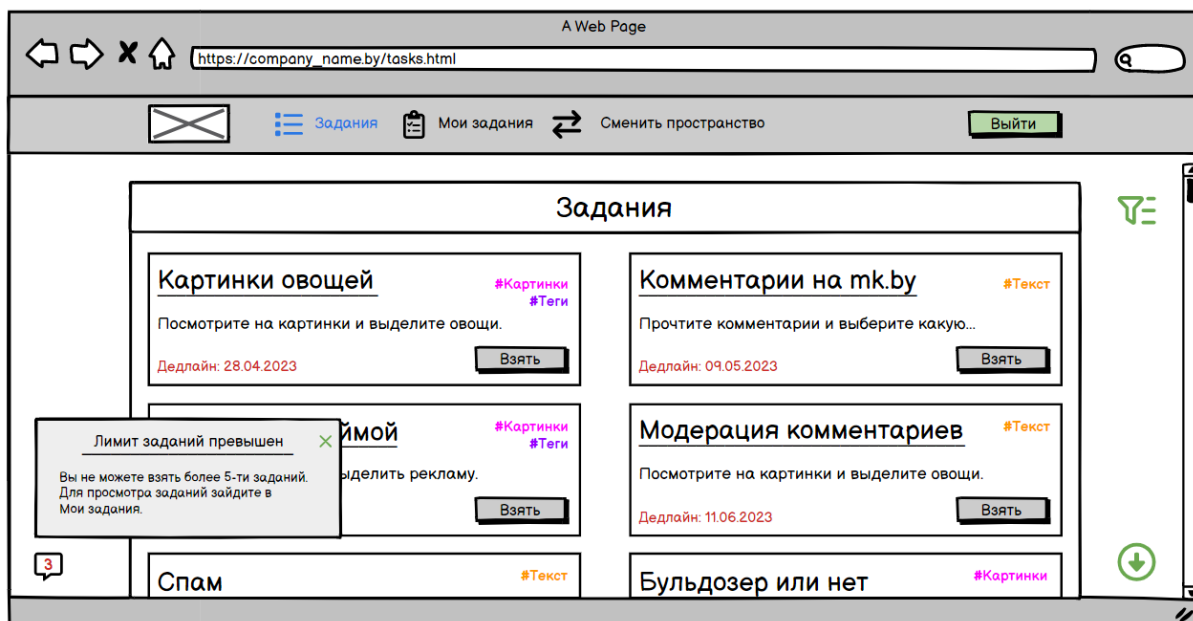
Перед тем как взять задание разметчик просматривает его описание, пример, и другие данные, в том числе дедлайн и минимальное количество оценок для единицы данных.



Приняв участие в разметке 5-ти заданий, разметчик больше не может взять новые. В этом случае кнопки Взять будут неактивны, а при попытке нажатия, разметчик будет уведомлен о том, что лимит заданий превышен.

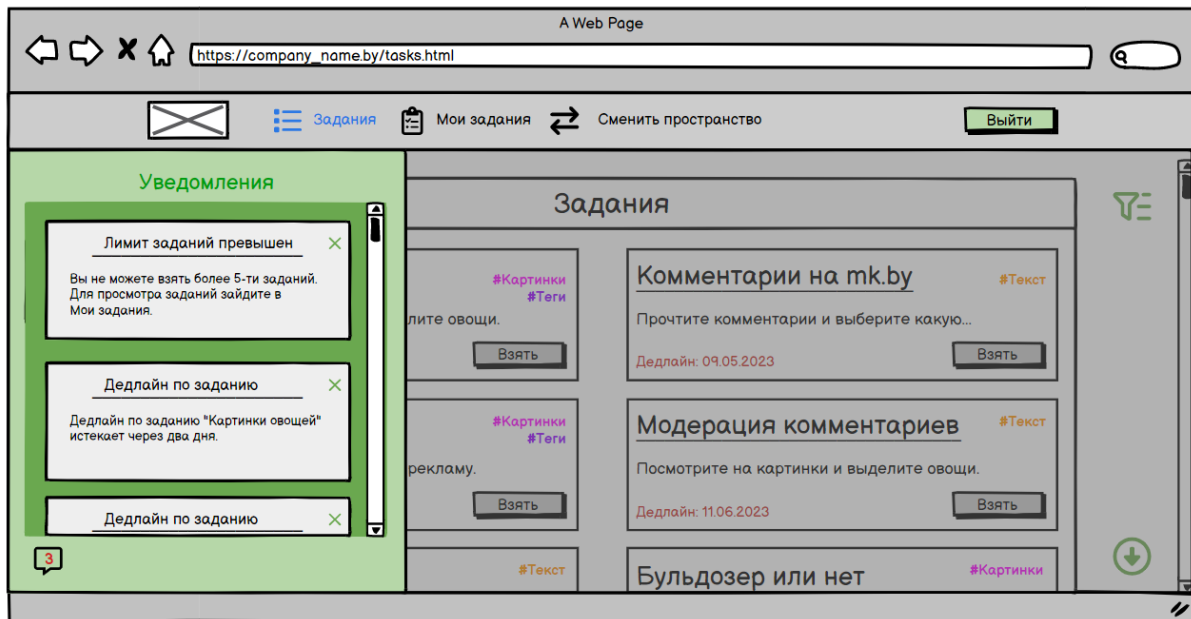


Уведомление осуществляется посредством всплывающих сообщений в левом нижнем углу экрана. Количество таких сообщений отображается на кнопке Сообщения ³.



Нажав на кнопку Сообщения ³ открывается окно уведомлений, в котором располагаются все сообщения, начиная с последнего. Каждое такое сообщение остается в окне уведомлений, пока пользователь не решит закрыть его, нажав на кнопку Закреть ^X. В этой начальной версии веб-приложения разметчику приходят сообщения двух видов: предупреждение о лимите заданий и предупреждение о предстоящем

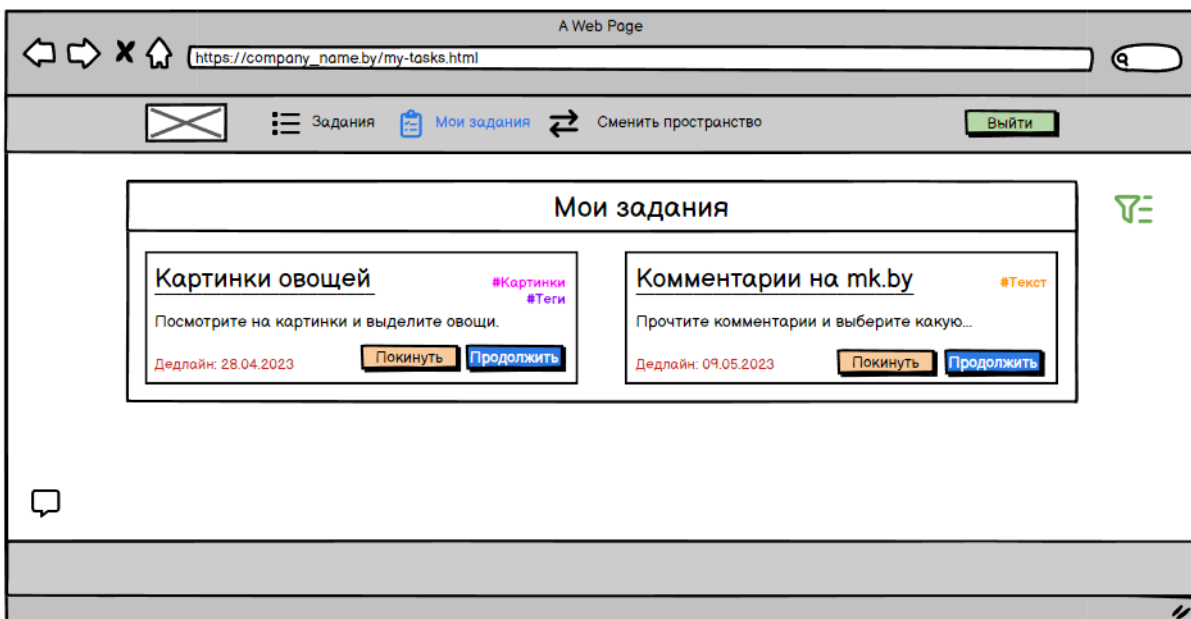
дедлайне.



В роли разметчика с главной страницы можно совершать переход между разделами-меню сервиса:

- Задания;
- Мои задания;
- Сменить пространство.

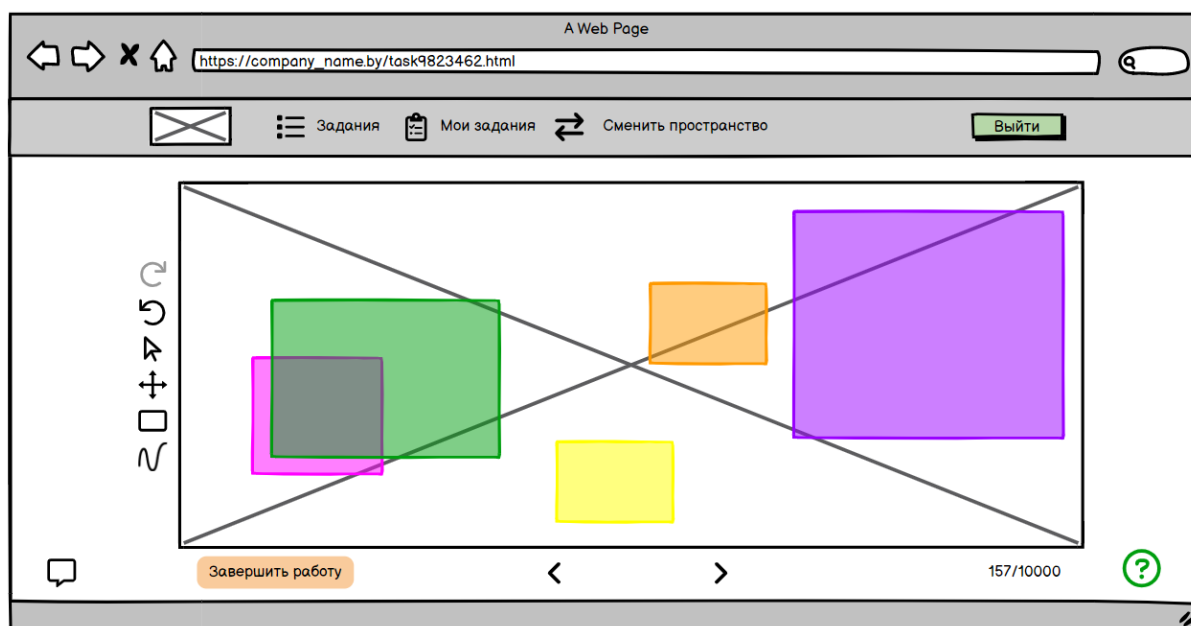
В разделе мои задания находятся задания, которые были взяты разметчиком. Разметчик может покинуть задание или продолжить работу в выбранном задании.

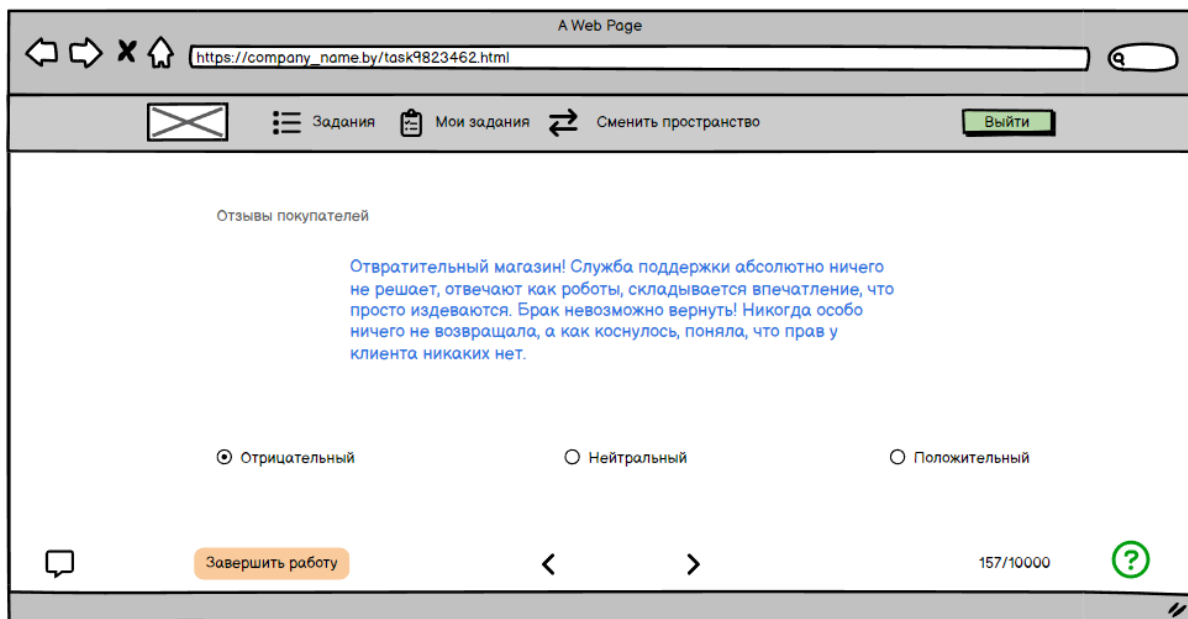


После нажатия кнопки Продолжить разметчик попадает на страницу выполнения задания, где ему доступны инструменты для работы с данными:

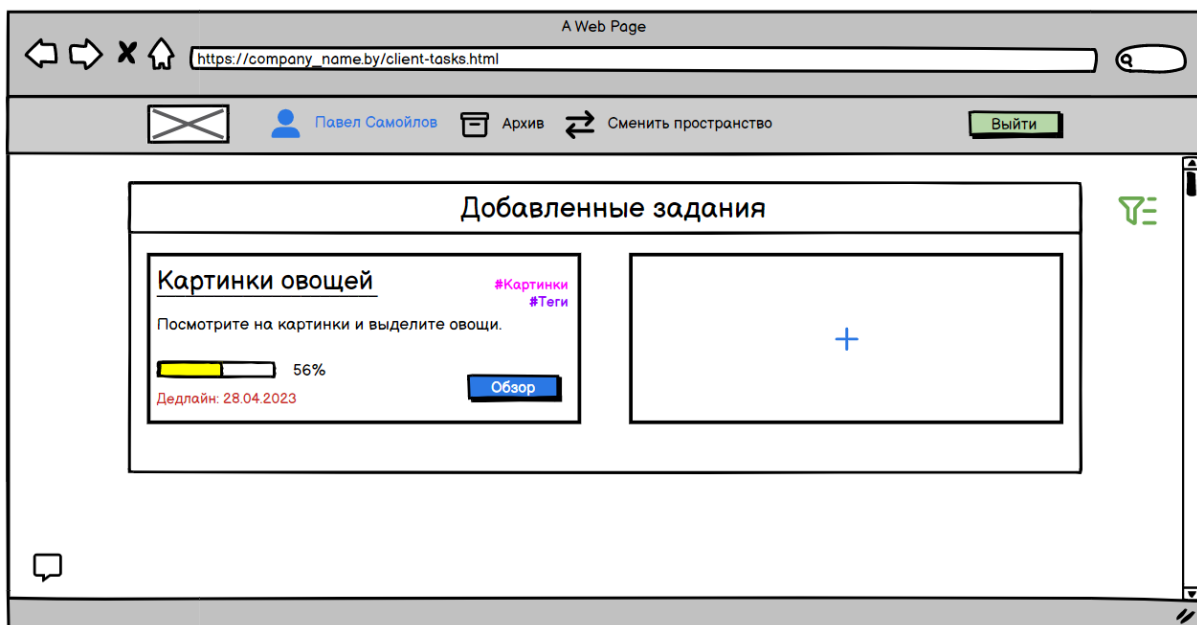
- ↶ – вернуть отмененное действие;
- ↷ – отменить последнее действие;
- ➦ – навигация;
- ➦ – перемещение выделенных областей;
- □ – выделение области прямоугольниками;
- ~ – выделение области по ломаной.


Разметчик может переключаться между единицами данных посредством кнопок Назад < , Вперед > . Справа на том же уровне находится информация о текущем местоположении разметчика в датасете 157/10000 . Нажав на кнопку Завершить работу, разметчик прекращает выполнение задания на текущем этапе, возвращаясь в меню Мои Задания.





Нажав на кнопку Сменить пространство разметчик меняет свою роль на заказчика. На главном экране отображаются активные, добавленные им задания.



Нажав на кнопку Создать задание , заказчик оказывается в окне добавления задания. Здесь он добавляет всю обязательную информацию о задании, пример и, конечно же, датасет.

A Web Page
https://company_name.by/add-task9823462.html

Павел Самойлов
 Архив
 Сменить пространство
Выйти


Добавление задания

Название:	<input type="text" value="Введите название"/>	Пример:	<input type="text" value="Выберите файл"/>
Описание:	<input type="text" value="Введите описание"/>		Загрузить данные Очистить
Количество оценок:	<input type="text" value="Введите количество оценок"/>	Данные:	<input type="text" value="Выберите файл"/>
Дедлайн:	<input type="text" value="Введите дедлайн"/>		Загрузить данные Очистить
Данные:	<input type="radio"/> Картинки <input type="radio"/> Текст		
Тип задания:	<input type="radio"/> Классификация <input type="radio"/> Теги		

Отмена
Сохранить

Вернуться к списку
Добавить задание

Каждое задание можно посмотреть, отредактировать. В каждом задании отображается текущий прогресс разметки и вся основная информация.

Нажав на кнопку **Скачать** , заказчик запустит скачивание актуальных на данный момент размеченных данных.

A Web Page
https://company_name.by/client-view-task9823462.html

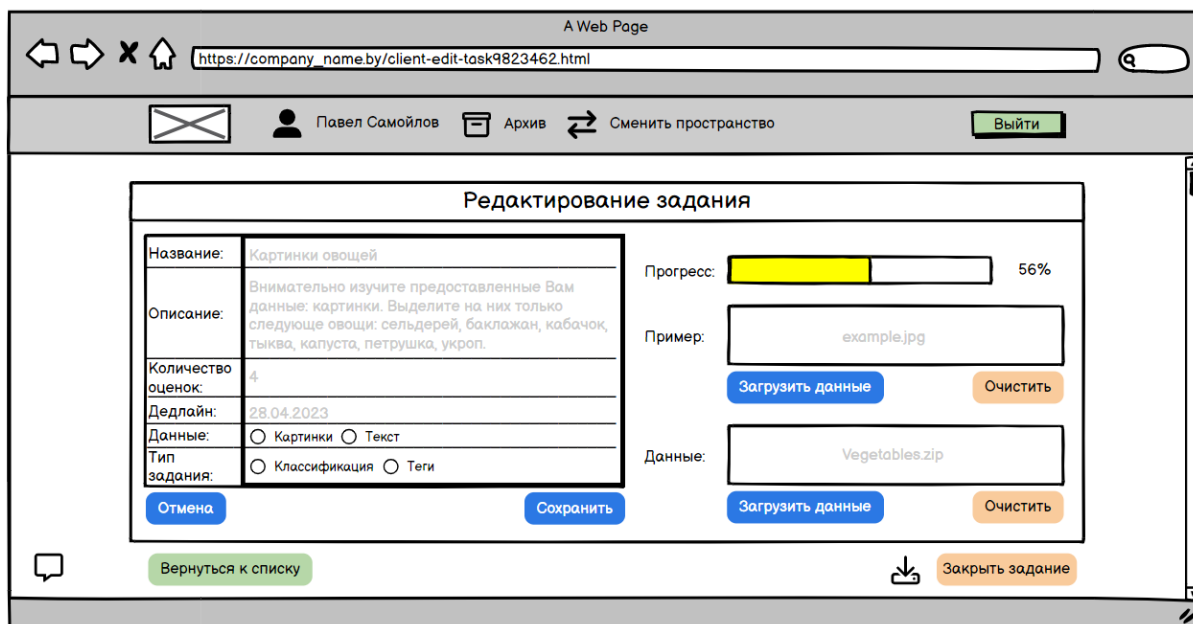
Павел Самойлов
 Архив
 Сменить пространство
Выйти

Обзор задания

Название:	Картинки овощей	Прогресс:	<div><div></div></div> 56%
Описание:	Внимательно изучите предоставленные Вам данные: картинки. Выделите на них только следующие овощи: сельдерей, баклажан, кабачок, тыква, капуста, петрушка, укроп.	Пример:	<input type="text" value="example.jpg"/>
Количество оценок:	4	Данные:	<input type="text" value="Vegetables.zip"/>
Дедлайн:	28.04.2023		
Тип задания:	#Картинки #Теги		

Редактировать

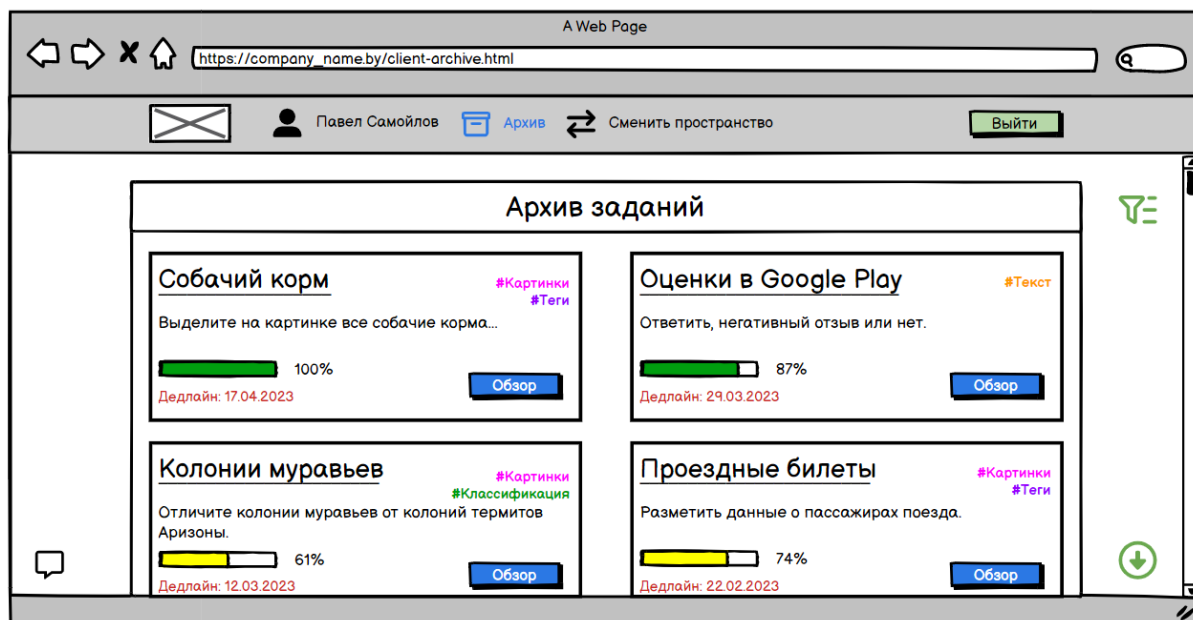
Вернуться к списку
 Заккрыть задание



В роли заказчика с главной страницы можно совершать переход между разделами-меню сервиса:

- Задания;
- Архив;
- Сменить пространство.

Нажав на кнопку Архив заказчик попадает на страницу со списком закрытых им заданий, которые также можно посмотреть.



Нажав на кнопку Сменить пространство, заказчик меняет свою роль на разметчика. Нажав на кнопку Выйти, пользователь выходит из сервиса.