# Рабочий кабинет платформы Хронолог-онлайн

Чтобы иметь возможность загружать собственные файлы (изображения, тексты), для пользователя организован личный кабинет – рабочее пространство, где происходит структуризация, накопление и преобразование источников информации с последующим выделением записей, которые и наполняют систему. Данный фрейм представляет собой самое ответственное и важное место всей системы, поскольку содержит в себе обращение к многим сущностям одновременно, позволяя гибко использовать весь функционал системы.

Рабочий кабинет пользователя поэтому должен быть ещё и динамичным в плане вёрстки с применением сплит-технологии (подобно codepen, jsfiddle) для основных областей типа навигации (аналог проводника для файлов и папок – коллекций), изображения, текста и записей. Традиционное расположение этих четырёх блоков предполагает использование правой и левой сторон окна для размещения блоков навигации (слева) и записей (справа) – такая компоновка является для большинства пользователей стандартной, что делает её освоение максимально быстрым.

Блоки изображения и текста размещаются в средней колонке, занимая по умолчанию большую часть всей площади страницы, т.к. являются основными. Два этих блока разделяются блоком управления и навигации, который содержит набор кнопок и объектов выбора для манипуляции текстовыми данными, а также блоками изображения, полученными при помощи OCR-системы либо же выделенными пользователем самостоятельно.

В блоке изображения происходит выделение блоков текста (слова, строки, абзацы), их категоризация (фамилия, имя, отчество, исход и т.п.), а также выделение записей, т.е. по сути это практически полный набор, необходимый для оцифровки и индексации снимков.

Блок текста служит альтернативой и дополнительным подспорьем блоку изображения и также имеет соответствующий функционал в плане категоризации, выделения записей, а также корректировки текста, данного OCR-системой. Бывает, что документ представлен лишь его оцифровкой, т.е. текстовым вариантом, поэтому данный блок является для индексации приоритетным, следовательно должен быть выполнен также качественно, как и блок изображения.

Блок навигации является стандартным и может быть реализован средствами библиотек типа jstree или ей подобных. Блок записей практически не отличается от стандартного табличного вида, используемого для редактирования пользовательских записей за исключением одной лишь особенности – наличия кнопки (иконки) «закрепить/открепить» для удобства работы с записями, занимающими несколько страниц.

Самое главное – организовать взаимодействие всех блоков на стороне клиента через оперирование с входным json (получается после OCR) и полученным из него массивом, а также порождение оного в отсутствии json (м.б. плохой почерк, слабый текст и т.п.).

Сделать всё возможно с использованием библиотек типа select-area (jquery), fabric (Vanila JS). Отличие состоит в использовании директивы canvas (полотно) для последней, а также богатый на примеры сайт с описанием. Из-за canvas могут возникнуть проблемы при работе split – на это нужно обратить внимание при выборе библиотек или же использовании чистого JS.

## Техническое задание

Взаимодействие пользователя с интерфейсом после выбора файла (изображение/текст) в блоке навигации происходит по следующим сценариям в зависимости от результатов автоматического распознавания текста OCR (локализация и детализация блоков) и индексации блоков (категоризации и определения записей):

1. **Машинописный текст на изображении** – хорошие результаты OCR. На изображение накладываются блоки OCR, из которых пользователю необходимо выделить записи (в том числе фон). Делается это путём создания соответствующего блока (прямоугольника) в режиме записей. Блоки OCR, заранее категоризированные пользователем в интерактивном режиме, попавшие во внутрь блока записи/фона считаются относящимися к нему по умолчанию. После этого пользователь нажимает на кнопку создать запись (через панель управления или через контекстное меню рядом с блоком) и данные пополняют таблицу записей.
2. **Набранный текст без изображения**. В блоке изображения ничего не появляется за счёт отсутствия такого файла и работа происходит исключительно в блоке текста. Пользователь категоризирует слова в режиме редактора, после чего переходит в режим записей/фона и путём объединения слов в группы создаёт записи или фон (что не является записью). При нажатии на кнопку создать запись (через панель управления или через контекстное меню рядом с блоком) данные пополняют таблицу записей.
3. **Рукописный текст на изображении** – плохие результаты OCR. Зачастую пользователю приходится всё делать самостоятельно, начиная с выделения словарных блоков (аналог OCR), т.е. выделять их на изображении и «подписывать» (расшифровывать слова). Т.о. ему предоставлен полный доступ к формированию соответствующего массива, который используется как основа для выделения блоков записей/фона по схеме, описанной выше.
4. **Имеются записи без текста, но с изображением.** *Этот вариант касается загруженных ранее из csv-файлов коллекции данных. Здесь возможно работать в режиме «мастера» для выделения блоков записей/фона путём последовательной локализации областей изображения. Блоки OCR могут быть созданы или выделены по желанию пользователя.*
5. **Набранный текст с изображением, загруженным позже.** *Этот вариант также возможен и заключается в том, что блоки OCR, а также блоки записей/фона могут быть не связаны с текстом, из которого ранее выделялись записи. В этом случае в зависимости от текста на изображении (рукописный или машинописный) выделяются блоки слов, возможно, в режиме «мастера», т.е. производится проход по каждому слову текста с тем, чтобы пользователь локализовал его на изображении. После этой процедуры работа с блоками записей/фона является стандартной процедурой.*

Желательным свойством рабочего кабинета является связность всех четырёх массивов – OCR, блоков (записей/фонов), текста и записей. Например, если пользователю удобно работать с текстом через блок текста, то при изменении или корректировке слова там, оно будет скорректировано и в блоке изображения и в блоке записи или если пользователь скорректирует расшифровку слова через блок изображения, изменения должны отобразиться и в блоке текста и в блоке записей. Это позволит производить изменения лишь в одной части и распространять их на другие. Для организации подобного взаимодействия нужно будет поставить триггеры на события изменения и в зависимости от источника этого события оказывать воздействие на остальные компоненты.

**Все действия с массивами выполняются на стороне клиента средствами JavaScript (jQuery). Взаимодействие с сервером происходит только при загрузке и выгрузке данных этих массивов в БД и организовано через соответствующие эндпоинты.**

# Общая компоновка рабочего кабинета, макеты основных панелей и элементов

**УПРАВЛЕНИЕ** (см. ниже)

**ИЗОБРАЖЕНИЕ**

(см. ниже, показаны блоки текста и записей)

**ТЕКСТ** (textarea с html-разметкой и применением span тэгов для обёртки слов)

**ЗАПИСИ**

(см. вправо)

**ФАЙЛЫ**

(стандарт, не показано)

**СОЗДАТЬ**

**Фамилия**

**Имя**

**Место**

🗹

🗹

**ВЫДЕЛЕНО** 2/5

3/7

Категория…

Переходы между файлами

Изменение категории

Управление выделением

Управление видом

Режим…

**OCR**

**INDEX**

Метка…

**СОЗДАТЬ**

Воскобойников (блок OCR)

Фамилия

Изменение категории OCR

Воскобойников П. К.

(блок записи)

Запись

🖉

Изменение категории записи

Изменение категории текста

Фамилия

🖉

Воскобойников