

Государственное бюджетное общеобразовательное учреждение города Москвы  
"Школа № 171" ш.к.15

«Школа 54»

Проектная работа

**SubSearch**

*Выполнили: Узденов Рустам Шамилевич и  
Петрухин Иван Алексеевич*

Научный консультант:  
Неклеенов Анатолий Николаевич

Москва 2020

## **Оглавление**

<b>Введение.....</b>	<b>3</b>
<b>Теоретические основы.....</b>	<b>3</b>
<b>Описание продукта.....</b>	<b>4</b>
<b>Заключение.....</b>	<b>7</b>
<b>Список литературы.....</b>	<b>7</b>

## Введение

**Актуальность.** В современных условиях возрастает роль самостоятельного изучения людьми вопросов, относящихся к различным сферам общественной жизни. В информационном пространстве размещено большое множество видеоматериалов, имеющих большое значение для самообразования в силу комплексного использования видеоряда и аудиосопровождения. В связи с этим проекты, позволяющие повысить эффективность использования данных материалов в процессе самообразования представляют безусловную актуальность.

**Новизна.** Предлагаемая проектная работа «SubSearch» позволяет эффективным образом осуществлять поиск контекстных слов в субтитрах видеороликов. В мире отсутствуют аналоги, соответствующие по своим возможностям. Есть несколько схожие решения, но существенно ограниченные в сфере своего применения, например:

1) Есть расширение к браузеру, но оно проблемно активируется (мне так и не удалось его активировать, что вызывает у меня сомнения в том, что большинство пользователей сможет воспользоваться этим расширением);

2) среди профессионального сообщества программистов есть ссылка на репозиторий, который теоретически может способствовать в поиске контекстных слов в субтитрах. Однако, это не является единым программным продуктом, которым можно сразу решить поставленные задачи по поиску, а для рядовых пользователей вообще не представляет никакой практической значимости, т.к. как-либо воспользоваться этим они не смогут в силу отсутствия специальных знаний.

«SubSearch» сможет использовать любой пользователь вне зависимости от своего уровня подготовки и на любой платформе.

**Практическая значимость.** Практически всем, кто просматривает видеофрагменты с субтитрами понадобится «SubSearch», который позволит в простой и интуитивно-понятной форме найти необходимые слова в субтитрах с указанием тайм-кода (с указанием временного момента).

## Теоретические основы

Цель проекта заключается в создании программного продукта, который позволяет в простой и интуитивно-понятной форме любому пользователю вне зависимости от своего уровня подготовки и на любой платформе осуществить поиск контекстных слов субтитрах видеофрагментов (например, размещенных в YouTube) с указанием тайм-кода.

Основными компонентами **Front-end** проекта являются:

- библиотека UI-компонентов «Semantic UI»;
- JavaScript фрейворк «React.js»

Весь код написан на JavaScript и находится во front-end. Основными компонентами **Back-end** проекта являются:

- Модуль загрузки субтитров из ссылки будет разработан с помощью библиотеки «axios» (<https://www.npmjs.com/package/axios>);
- Парсинг-модуль будет разработан с помощью специальной библиотеки «subtitle» (<https://www.npmjs.com/package/subtitle>);
- Библиотека JavaScript «Fuse.js»

Используются следующие основные понятия:

**Субтитры** - текстовое сопровождение видеоряда на языке оригинала или переводное, дублирующее и иногда дополняющее, например для зрителей с нарушенным слухом, звуковую дорожку кинофильма или телепередачи.

**Видеосервис** - сайт, на котором хранятся видео в сети «Интернет».

**Контекстное слово** – слово, которое пользователь нашего ПО ищет в субтитрах видео.

**Web-приложение** - клиент-серверное приложение, в котором клиент взаимодействует с веб-сервером при помощи браузера.

**Макет сайта** - графическая схема web-приложения с указанием отступов и прочих параметров в статике.

**Библиотека UI-компонентов** – набор программных компонентов, позволяющих создать макет сайта для использования пользователями сайта.

**Дизайн интерфейса web-приложения** – внешний вид web-приложения.

**Модуль загрузки субтитров** – программное средство, позволяющее загрузить и обработать субтитры.

**Ссылка** - объект, указывающий на определённые данные в сети «Интернет», но не хранящий их.

**Модуль поиска по субтитрам** – программное средство, позволяющее пользователям находить контекстные слова в субтитрах с тайм-кодом.

**Тайм-код** – цифровые данные о временном участке видеофрагмента, в котором используется контекстное слово

**Хостинг** - размещение информации (в частности, web-приложения) на сервере, постоянно имеющем доступ к сети «Интернет».

**Парсинг** - часть программного обеспечения, выполняющая синтаксический анализ.

**Фреймворк** - программная платформа, определяющая структуру программной системы

**“React js”** - JavaScript-библиотека с открытым исходным кодом для разработки web-приложения.

**“GitHub Pages”** – сервис, бесплатно предоставляющий услуги по хостингу web-приложений.

## Описание продукта

### 1. Front-end

#### 1.1. Основные компоненты:

- библиотека UI-компонентов «Semantic UI»;
- JavaScript фрейворк «React.js»

1.2. На главной странице размещается поле для ввода ссылки на видео и поле для ввода поискового слова, совмещенное с полем выбора языка субтитров (см.: *Изображения № 1 - № 3*). Оба поля оснащены индикатором загрузки для отслеживания прогресса поиска.

*Изображение № 1*

*Изображение № 2*

*Изображение № 3*

1.2. Порядок ввода и обработки запроса:

1.2.1. После ввода ссылки идет загрузка субтитров на сайт и весь прогресс загрузки отображается в индикаторе (см.: *Изображения № 1*);

1.2.2. Пользователь выбирает язык субтитров (см.: *Изображения № 2 - 3*);

1.2.3. После загрузки субтитров пользователь может ввести нужное слово в поисковое поле. Прогресс загрузки отображается уже в другом индикаторе (см.: *Изображения № 2 - 3*);

1.2.4. После поиска будет выведена таблица с двумя столбцами: первый столбец будет с тайм-кодами (гиперссылки на нужное видео и время), а второй столбец - с предложением, где было найдено это слово. В данной таблице также

выводятся результаты поиска слов, схожих по написанию в случае ошибочного написания пользователем слова в поисковом окне (см. *Изображение № 4*);

1.2.5. После поиска будет выведена таблица с двумя столбцами: первый столбец будет с тайм-кодами (гиперссылки на нужное видео и время), а второй столбец - с предложением, где было найдено это слово (см. *Изображение № 4*);

*Изображение № 4*

Поле ввода

сто

Таблица

<a href="#">1:10</a>	Сто лет меня не было
<a href="#">11:54</a>	Просто сделать
<a href="#">17:10</a>	Стол черный, небо серое

## 2. Back-end

2.1. Весь код написан на JavaScript и находится во front-end

2.2. Основные компоненты:

➤ Модуль загрузки субтитров из ссылки будет разработан с помощью библиотеки «axios» (<https://www.npmjs.com/package/axios>);

➤ Парсинг-модуль будет разработан с помощью специальной библиотеки «subtitle» (<https://www.npmjs.com/package/subtitle>);

➤ Библиотека JavaScript «Fuse.js»

2.3. Примерное описание работы back-end сайта

2.3.1. Сайт получает ссылку;

2.3.2. Модуль загрузки субтитров делает запрос на [http://www.youtube.com/get\\_video\\_info?video\\_id={id видео}](http://www.youtube.com/get_video_info?video_id={id видео}) (пример ссылки, где в части «id видео» необходимо указать, соответственно, соответствующее значение id видео);

2.3.2. Парсинг-модуль обрабатывает ответ сервера и преобразовывает JSON в JavaScript-объект, с которым удобно работать;

2.3.3. Приложение получает ссылки на файлы субтитров и добавляет параметр vtt-формата, чтобы удобно с ним работать;

2.3.4. Выводится пользователю список языков субтитров (сайт получил языки в первом ответе сервера);

2.3.5. Приложение, при выборе языка, загружает субтитры по соответствующей ссылке;

2.3.6. При помощи метода массива «.filter» отфильтровывается массив отрывков субтитров, чтобы остались только те, в которых содержится текст;

2.3.7. Сайт получает поисковое слово и выводит нужную таблицу.

## Заключение

Проект «SubSearch» (<http://bit.ly/3nSTzOU>) представляет собой реализацию творческого потенциала учеников 10 класса «Н» Рустама Узденова и Ивана Петрухина посредством создания программного продукта, который позволяет в простой и интуитивно-понятной форме любому пользователю вне зависимости от своего уровня подготовки и на любой платформе осуществить поиск контекстных слов субтитрах видеофрагментов (например, размещенных в YouTube) с указанием тайм-кода.

## Список литературы

1. Роберт Мартин. Чистая архитектура. Искусство разработки программного обеспечения. Питер, 2018 г. 352 с. ISBN: 978-5-4461-0772-8
2. React highlight words (технологическая документация) - <https://github.com/bvaughn/react-highlight-words>
3. Технологическая документация к Semantic UI - <https://semantic-ui.com/introduction/getting-started.html>
4. Технологическая документация к «React.js» - <https://ru.reactjs.org/docs/getting-started.html>