МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Отчёт по курсу

«Методология программной инженерии»

Тема: «Техническое задание для разработки

веб-портала поиска видеофайла по исходному изображению»

ИСПОЛНИТЕЛЬ:

студент группы ИУ7-22М,

Акашкин К.В.

ПРЕПОДАВАТЕЛЬ:

Романова Т.Н.

Москва.

2019 г.

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ**

**на разработку распределённой системы обработки информации**

**«Веб-портал поиска видеофайла по исходному изображению»**

***Глоссарий***

|  |  |
| --- | --- |
| **Термин** | **Определение** |
| ТЗ | В текущем контексте – данное техническое задание |
| РСОИ | Распределённая система обработки информации, в данном контексте – веб-портал с набором сервисов, которые могут быть территориально (географически) и функционально (размещение на разных серверах на одной территории) распределены (в т.ч. в части использования различных БД каждым сервисом, а также «по линии» пространственно распределённых frontend-сервиса и остальных сервисов в рамках РСОИ), но имеют внутренние механизмы (стандартные протоколы передачи сообщений и API) для межсервисного взаимодействия как единая система обработки информации. А также используют архитектурный шаблон MVC и, возможно, выделение логики приложения на отдельном слое – сервере приложений (в дополнение к слоям клиента и слою данных) в рамках трёхуровневой архитектуры ПО. |
| БД | База данных, в данном контексте – реляционная (возможно, с наличием ряда возможностей объектной модели БД), или одна из «No-SQL» БД, что оговаривается далее в тексте специально. |
| API | Application Programming Interface – программный интерфейс приложения, в данном контексте – API отдельных сервисов в рамках РСОИ. |
| Архитектурный шаблон MVC | Шаблон Model, View, Controller для разделения представления данных (view) и самих данных (model) на относительно независимые компоненты за счёт введения компонента, управляющего логикой размещения данных в представлении (пользовательском интерфейсе) - controller. |
| Фронтэнд (или Frontend) | В данном контексте – пользовательский интерфейс (слой представления данных) приложения. |
| Бэкэнд (или Backend) | В данном контексте – механизм доступа к данным и бизнес-логика (слои данных и бизнес-логики) приложения. |
| Трёхуровневая архитектура ПО | Архитектурная модель создания ПО, прелполагающая, в данном контексте – разделение создаваемой РСОИ на три компонента: «клиента», «сервера приложений» и «сервера данных». |
| Валидация | Валидация  - это проверка данных на соответствие заданным условиям, ограничениям и регулярным выражениям. |
| UUID | UUID ([англ](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA). universally unique identifier "универсальный уникальный идентификатор") — это стандарт [идентификации](https://ru.wikipedia.org/wiki/%D0%98%D0%B4%D0%B5%D0%BD%D1%82%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F_(%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D1%8B)), используемый в создании [программного обеспечения](https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D0%B5%D1%81%D0%BF%D0%B5%D1%87%D0%B5%D0%BD%D0%B8%D0%B5), стандартизированный [Open Software Foundation](https://ru.wikipedia.org/w/index.php?title=Open_Software_Foundation&action=edit&redlink=1) как часть DCE — среды распределённых вычислений. |

***Введение***

Данное техническое задание составлено для разработки проекта «Веб-портал поиска видеофайлов по исходному изображению**»**

Техническое задание выполнено на основе ГОСТ 19.201—78 «ЕСПД. Техническое задание. Требования к содержанию и оформлению».

***Краткое описание предметной области***

По данным из докладов аналитической фирмы IDC, к 2025 году объем всех данных во всем мире составит 163 зеттабайт. Это в 10 раз больше, чем общий объем данных по состоянию на 2016 год. Каждую минуты на различные видео хостинги загружают более 100 часов видео материала. По данным IMDb, на 16 августа 2015 года количество фильмом снятых во всем мире составило примерно 333 000. При таком росте объёмов информации с каждым годом становится все сложнее отыскать необходимые данные. В настоящий момент существуют различные сервисы, реализующие методы поиска похожих изображений и аудиофайла по входному аудио фрагменту. Ввиду того, что роль визуального контента с каждым годом возрастает, то появляется необходимость в решение задачи поиска видеороликов по изображению.

Данное ТЗ определяет требования к разработке веб-портала для распределённого поиска видеофайлов по изображению и хранения информации о пользователях и видеофайлах.

***Существующие аналоги***

На данный момент не существует общедоступных аналогов портала поиска фильмов / видеофайлов по входному кадру. Существуют сервисы решающие похожие задачи, но для других видов данных:

* Shazam – сервис поиска аудиофайлов по небольшому аудио фрагменту;
* [Яндекс.Картинки – сервис поиска изображений в интернете и поиска по картинке;](https://yandex.ru/images/)
* [TinEye позволяет находить в Интернете изображения, похожие на изображение-образец. Для получения результата в качестве образца достаточно даже миниатюры файла с низким разрешением](https://yandex.ru/images/)[[[3]](https://yandex.ru/images/)](https://ru.wikipedia.org/wiki/TinEye#cite_note-compulenta-3)[.](https://yandex.ru/images/)[[Поисковый индекс](https://yandex.ru/images/)](https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%B8%D1%81%D0%BA%D0%BE%D0%B2%D1%8B%D0%B9_%D0%B8%D0%BD%D0%B4%D0%B5%D0%BA%D1%81)[TinEye насчитывает свыше 35 млрд изображений.](https://yandex.ru/images/)
* [Google картинки - сервис поиска изображений в интернете и поиска по картинке, аналогичен Яндекс.Картинки и другие.](https://yandex.ru/images/)

***Описание системы***

Проект должен представлять собой веб-портал для поиска видеофайлов по исходному изображению с возможностью централизованного хранения данных о файлах и обработки полученной информации от пользователей.

Пользователи должны иметь возможность загружать изображение формата jpeg и png. После загрузки файл обрабатывается и производится поиск необходимого видеофайла в централизованной, специально выделенной для этого БД.

Пользователи должны иметь возможность зарегистрироваться (с разными ролями и правами доступа) на портале для возможности работы со своим функционалом в его рамках.

На рисунке далее отображена схема предметной области.



Система будет состоять из фронтенда и пяти бекендов, что наиболее целесообразно для реализации ее основного назначения.

**Сервис-пользователей** отвечает за сессию пользователей портала и реализует следующие функции:

* регистрация пользователя (модератора или пользователя);
* аутентификация (проверка сессии) пользователя;
* авторизация пользователя (вход, или «логин»);
* выход из сессии («логаут»).

В базе данных пользователей хранится следующая информация о модераторе:

• ФИО;

• пол;

• электронная почта;

• пароль в зашифрованном виде;

• фотография;

• идентификатор пользователя(UUID).

И следующая информация о пользователе:

• ФИО;

• пол;

• электронная почта;

• пароль в зашифрованном виде;

• идентификатор пользователя(UUID).

• фотография.

Сервис-истории запросов отвечает за хранение данных обо всех запросах поиска и обработки видеофайлов.

В базе данных хранится следующая информация:

• дата запроса;

• фотография поиска;

• UUID найденного видеофайла;

Сервис-обработки видеофайла отвечает за обработку видеозаписи и запись в БД результаты обработки.

В базе данных хранится следующая информация:

• UUID видеофайла;

• ссылка на видеофайл;

• название;

• описание;

• фотография;

• массив хешей кадров.

Сервис-поиска видеофайла отвечает за поиск видеозаписи по входному изображению и массива хешей кадров.

Агрегирующий сервис видеофайла отвечает за объединение данных со всех сервисов и отправку данных на сервис фронтенда.

Фронтенд принимает запросы от пользователей по протоколу HTTP и анализирует их. На основе проведенного анализа фронтенд выполняет запросы к агрегирующему сервису и отсылает ответ пользователю. Фронтенд также включает в себя модуль сбора статистики. Модуль сбора статистики считает и визуализирует время загрузки страницы.

***Основания для разработки***

Разработка ведется в рамках выполнения лабораторных работ по курсу «Методология программной инженерии» на кафедре «Программное обеспечение ЭВМ и информационные технологии» факультета «Информатика и системы управления» МГТУ им. Н. Э. Баумана.

***Назначение разработки***

Главное назначение разрабатываемого портала – обеспечение возможности поиска видеофайлов по исходному изображению.

***Требования к системе***

1. Разрабатываемое программное обеспечение не относится к классу критических для миссии организации систем. Поэтому требования к среднегодовому времени доступности и допустимому времени, в течение которого система не доступна в течение года, а также к возможности «горячего» переконфигурирования в данном ТЗ не устанавливаются.

2. Время восстановления системы после сбоя не должно превышать 20 рабочих или 30 календарных дней, в зависимости от того, какой из периодов завершится раньше.

***Требования к функциональным характеристикам***

1. Портал должен обеспечивать возможность запуска в современных браузерах: не менее 85% пользователей Интернета должны иметь возможность пользоваться порталом без какой-либо деградации функционала.

2. По результатам работы модуля сбора статистики медиана времени отклика системы на запросы пользователя (в любой роли) на получение информации не должна превышать 3 секунд без учета латентности географического расположения узла.

3. По результатам работы модуля сбора статистики медиана времени отклика системы на запросы, добавляющие или изменяющие информацию на портале не должна превышать 15 секунд без учета латентности географического расположения узла.

***Функциональные требования к порталу с точки зрения пользователя***

Веб-портал должен обеспечивать реализацию следующих функций:

1. Система должна обеспечивать регистрацию пользователей (для роли каждого типа, т.е. администратор, модератор, пользователь) с валидацией вводимых данных.

2. Система должна обеспечивать аутентификацию указанных в п. 1 пользователей.

3. Система должна обеспечивать разделение вышеуказанных пользователей на три роли:

- пользователь;

- модератор;

- администратор.

4. Система должна предоставлять **пользователю** следующие функции:

- возможность загружать png и jpeg файлы для поиска необходимого видеофайла;

- возможность просмотра списка наиболее популярных видеофайлов;

- возможность просмотра списка ранее выполненных запросов на поиск;

- возможность редактирования данных пользовательского профиля.

5. Система должна предоставлять администратору следующие функции:

- неограниченные полномочия по изменению контента на портале;

- возможность конфигурирования отдельных узлов (сервисов) системы: настройки, удаления, добавления узлов;

- возможность «горячего» (без необходимости «рестарта» всей РСОИ) конфигурирования (включения, отключения, модификации) всех основных узлов системы.

6. Система должна предоставлять модератору следующие функции:

- возможность добавления видеофайла для последующей обработки и добавления в БД;

***Входные данные системы***

Для пользователей при регистрации:

- Имя, фамилия, электронная почта, пароль, максимальная длина каждого из полей – 512 символов (в формате UTF-8).

Для пользователей после регистрации:

- Изображения в формате png и jpeg, максимальный размер файла 5 мегабайт.

Для модератора:

- Наименование видеофайла, описание, ссылки на файл, максимальная длина каждого из полей – 512 символов (в формате UTF-8).

- Видеофайл в формате mp4, максимальный размер файла 2 гигабайта.

***Выходные данные системы***

Выходными параметрами системы являются веб-страницы. Они должны содержать следующую информацию, в зависимости от роли пользователя в системе.

Для пользователей:

- список наиболее популярных видеофайлов,

- детальную информацию по найденному видеофайлу,

- информацию о пользовательском профиле.

Для модератора:

- информацию о загрузке видеофайла.

Для администраторов системы

- информацию обо всех пользователях и видеофайлах.

***Требования к составу и параметрам технических средств***

Каждый экземпляр серверной части данного веб-портала (все back-end сервисы в совокупности) должен потреблять (в т.ч. случае пиковых нагрузок) не более 4 Гб оперативной памяти и работать на сервере с процессором Intel i7 2.4 GHz и жестким диском SATA 100 IOPS 100 Гб.

Конкретные числовые ограничения на системные требования к клиентской части веб-портала не устанавливаются, но при этом они не должны превышать соответствующих величин для серверной части.

***Требования к надёжности***

Разрабатываемая система относится к классу систем «текущей деятельности» («ordinary business»), что снижает требования к её надёжности (а следовательно, также и стоимость) по сравнению с классом «критических для миссии организации» систем: необходимо обеспечить регулярное резервирование («backup») данных из централизованного хранилища (БД) веб-портала.

В системе необходимо использовать штатные промышленные механизмы аварийного завершения работы, обеспечения целостности данных и восстановления после сбоя от вендоров СУБД (например, аналогичные соответствующим механизмам в Oracle СУБД) – автоматический откат до последней успешной операции, резервирование лога текущих операций и его автоматическая загрузка после восстановления системы.

В случае наличия некритических для работы системы ошибок, она должна проинформировать о них пользователя и продолжить работу в части, не затронутой данной ошибочной ситуацией.

***Требования к документации***

Исполнитель должен подготовить и передать Заказчику следующие документы:

- руководство по развертыванию Системы;

- руководство администратора Системы;

- руководство для пользователя по использованию Системы;

- руководство для модератора по использованию Системы.