**Московский Авиационный Институт**

**(Национальный исследовательский университет)**

Институт №8 «Информационные технологии и прикладная математика»

Кафедра вычислительной математики и программирования

**Лабораторная работа №1**

**по курсу «Информационный поиск»**

|  |
| --- |
| Студент: Зайцев Н.В.  группа М8О-208М-20  Преподаватель: Кухтичев А.А. |
|
|
|
|
|
|
|
|
|

Москва, 2021

# Лабораторная работа № 1

Необходимо подготовить корпус документов, который будет использован при выполнении остальных лабораторных работ. Требуется скачать его к себе на компьютер, ознакомиться с ним и его характеристиками, разбить на документы.

**Подготовка к работе**

Источником данных является Википедия на русском языке. В качестве корпуса документов была выбрана категория Википедии «Живопись».

Вся категория довольно легко поддалась скачиванию и выделению текста, но процесс занял много времени из-за большого количества статей.

**Выделенный текст. Пример статьи**

Альфрейная живопись — разновидность интерьерной росписи, имитирующая различные типы и виды отделки внутреннего помещения (например, ценные породы дерева, гипсовую лепнину, шёлковые драпировки, золотые или серебряные поверхности, растительные узоры и т.п.).

Альфрейная живопись отличается от фресковой способом нанесения рисунка на стену: фресковая роспись предполагает нанесение изображения на стену с сырой штукатуркой, а альфрейная, напротив, — на высохшую оштукатуренную поверхность. В настоящее время под термином «альфрейная живопись» подразумеваются следующие виды орнаментально-декоративной росписи:

монохромная (выполняемая в одном цвете и тоновых градациях этого цвета);

полихромная (многоцветные орнаменты, стилизованные цветочные композиции, фантастические фигурки, геральдические элементы и т.д.);

гризайль (раздвигающие пространство картины-обманки, имитация архитектурных элементов, лепнины и т.д.)Альфрейная живопись зародилась в III тысячелетии до н.э. у древних египтян при росписи стен пирамид. Альфрейная живопись по сухой штукатурке стала вытеснять классическую фреску по сырой штукатурке в эпоху Возрождения, когда искусство постепенно возвращалось к идеалам античности и становилось всё более светским. Наибольшего расцвета альфрейная живопись достигла к концу XVIII века в Венеции. В России образцы альфрейной живописи можно увидеть в дворцах Санкт-Петербурга: Зимний, Александровский, Мраморный, Юсуповский и т.д.

Ссылки

Альфрейная живопись // «Карл Павлович Брюллов» : сайт художника (bryullov.ru) (Проверено 10 апреля 2017)

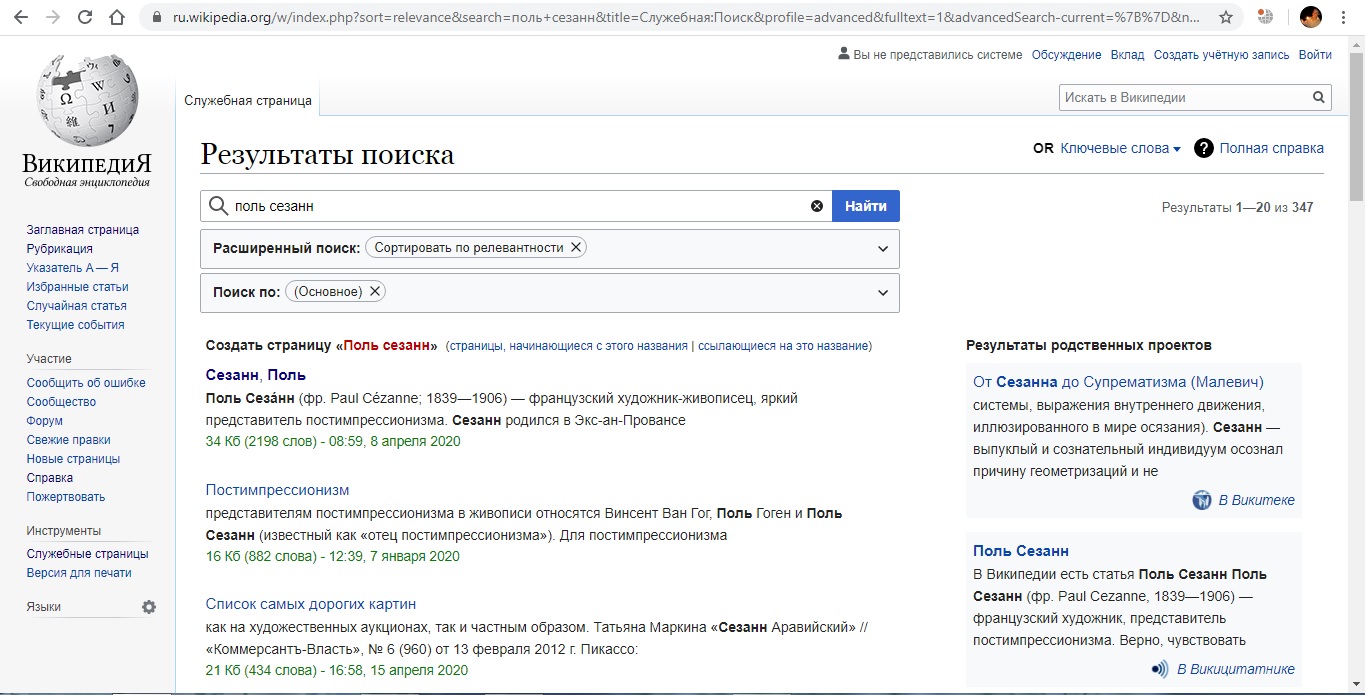
**Ход работы**

Программа получения корпуса документов была написана на языке Python. Также была использована библиотека wikipedia-api, позволяющая работать с Википедией и средствами языка для записи документов.

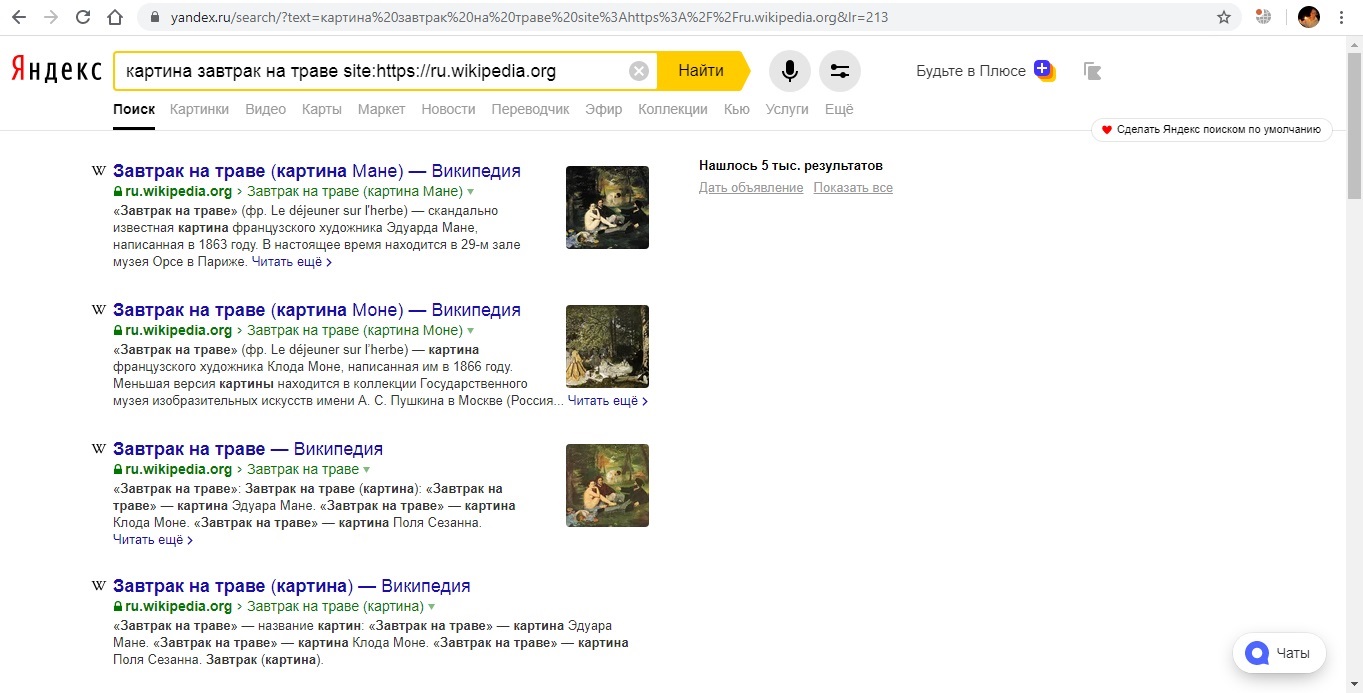
Для поиска по выбранному корпусу документов можно использовать как встроенный поиск Википедии, так и поиски Google, Mail или Яндекс с ограничением на сайт Википедия.

**Примеры запросов к существующим поисковикам:**

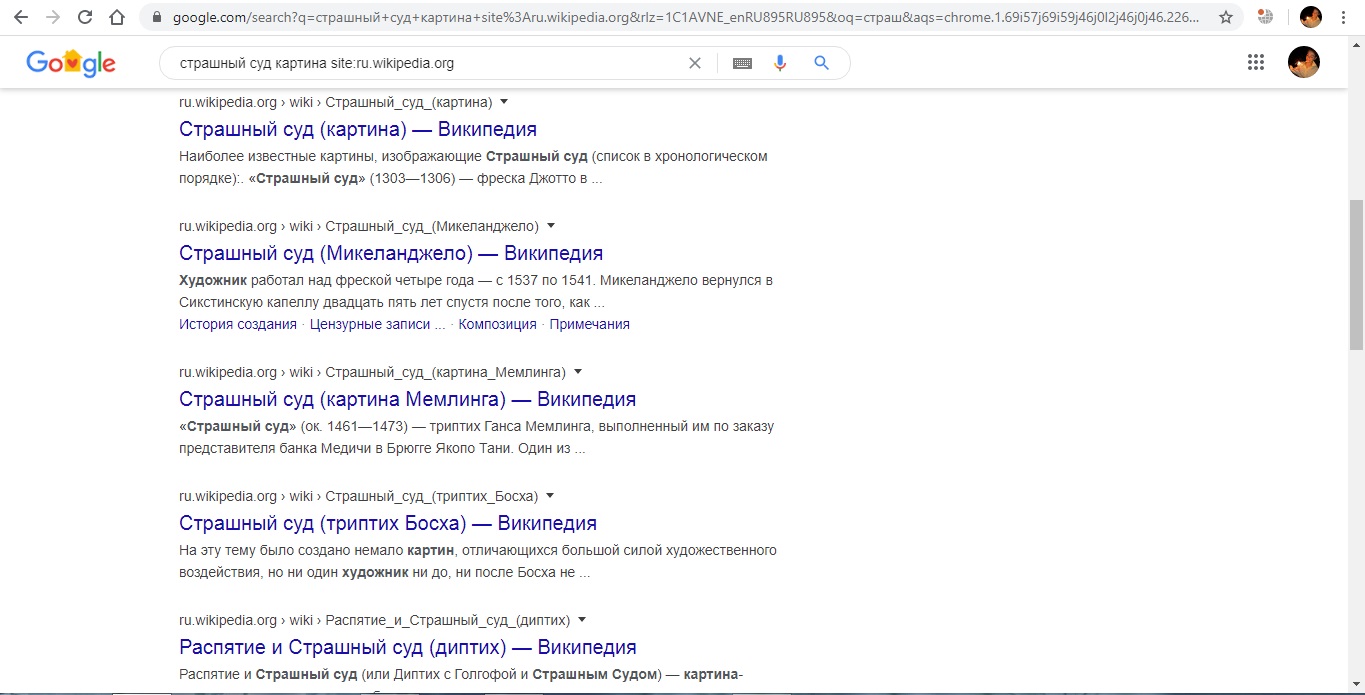
**Википедия**

****

**Яндекс**

****

**Google**

****

**Статическая информация о корпусе**

«Сырые» данные не сохранялись.

Количество документов – 132938.

Размер «чистого» текста – 975 МБ.

Средний размер документа – 7.3 Кб.

Средний объем текста в документе – 4000 знаков с пробелами.

**Вывод**

В ходе выполнения лабораторной работы был получен корпус документов по категории «Живопись» с русскоязычной Википедии. Я познакомился с библиотекой Python – wikipedia-api. Для работы с Википедией эта библиотека подходит как можно лучше и позволяет производить импорт категорий и отдельных страниц, их полный текст, разделы страницы и ее перевод на другие языки. В целом время скачивания корпуса документов приблизительно составило 19 часов.