

**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

← → ↻ 🏠 ru.wikipedia.org/w/index.php?sort=relevance&search=поль+сезанн&title=Служебная:Поиск&profile=advanced&fulltext=1&advancedSearch-current=%7B%7D&n... ☆ 🌐 👤

Вы не представились системе Обсуждение Вклад Создать учётную запись Войти

Искать в Википедии 🔍

OR Ключевые слова ? Полная справка

Результаты поиска

🔍 поль сезанн ✕ **Найти** Результаты 1—20 из 347

Расширенный поиск: Сортировать по релевантности ✕

Поиск по: (Основное) ✕

Создать страницу «Поль сезанн» (страницы, начинающиеся с этого названия | ссылающиеся на это название)

Сезанн, Поль
Поль Сезанн (фр. Paul Cézanne; 1839—1906) — французский художник-живописец, яркий представитель постимпрессионизма. **Сезанн** родился в Экс-ан-Провансе
34 Кб (2198 слов) - 08:59, 8 апреля 2020

Постимпрессионизм
представителям постимпрессионизма в живописи относятся Винсент Ван Гог, **Поль Гоген** и **Поль Сезанн** (известный как «отец постимпрессионизма»). Для постимпрессионизма
16 Кб (882 слова) - 12:39, 7 января 2020

Список самых дорогих картин
как на художественных аукционах, так и частным образом. Татьяна Маркина «Сезанн Аравийский» // «Коммерсантъ-Власть», № 6 (960) от 13 февраля 2012 г. Пикассо:
21 Кб (434 слова) - 16:58, 15 апреля 2020

Результаты родственных проектов

От **Сезанна** до Супрематизма (Малевич) системы, выражения внутреннего движения, иллюзорного в мире осознания). **Сезанн** — выпуклый и сознательный индивидум осознал причину геометризации и не

Поль Сезанн
В Википедии есть статья **Поль Сезанн** **Поль Сезанн** (фр. Paul Cézanne, 1839—1906) — французский художник, представитель постимпрессионизма. Верно, чувствовать

В Викицитатнике

Яндекс

← → ↻ 🏠 yandex.ru/search/?text=картина%20завтрак%20на%20траве%20site%3Ahttps%3A%2F%2Fru.wikipedia.org&lr=213 ☆ 🌐 👤

Яндекс картина завтрак на траве site:https://ru.wikipedia.org ✕ **Найти** 🔊 🔄 Будьте в Плюсе + 📄

Поиск Картинки Видео Карты Маркет Новости Переводчик Эфир Коллекции Кью Услуги Ещё

♥ Сделать Яндекс поиском по умолчанию

Нашлось 5 тыс. результатов
[Дать объявление](#) [Показать все](#)

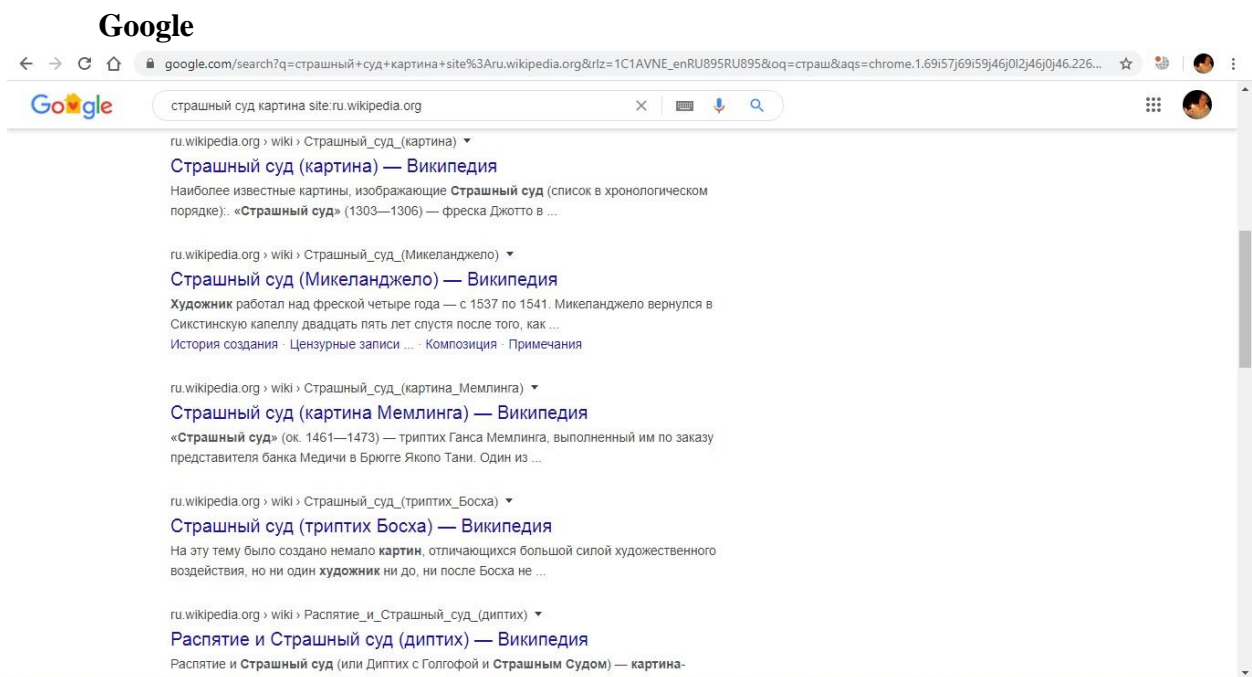
📄 **Завтрак на траве (картина Мане) — Википедия**
ru.wikipedia.org > Завтрак на траве (картина Мане) > «Завтрак на траве» (фр. Le déjeuner sur l'herbe) — скандально известная картина французского художника Эдуарда Мане, написанная в 1863 году. В настоящее время находится в 29-м зале музея Орсе в Париже. Читать ещё >

📄 **Завтрак на траве (картина Моне) — Википедия**
ru.wikipedia.org > Завтрак на траве (картина Моне) > «Завтрак на траве» (фр. Le déjeuner sur l'herbe) — картина французского художника Клода Моне, написанная им в 1866 году. Меньшая версия картины находится в коллекции Государственного музея изобразительных искусств имени А. С. Пушкина в Москве (Россия... Читать ещё >

📄 **Завтрак на траве — Википедия**
ru.wikipedia.org > Завтрак на траве > «Завтрак на траве»: Завтрак на траве (картина): «Завтрак на траве» — картина Эдуарда Мане. «Завтрак на траве» — картина Клода Моне. «Завтрак на траве» — картина Поля Сезанна. Читать ещё >

📄 **Завтрак на траве (картина) — Википедия**
ru.wikipedia.org > Завтрак на траве (картина) > «Завтрак на траве» — название картин: «Завтрак на траве» — картина Эдуарда Мане. «Завтрак на траве» — картина Клода Моне. «Завтрак на траве» — картина Поля Сезанна. Завтрак (картина).

Чаты



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.