Документация к проекту «Unknown name»

Санкт-Петербург

2021

Оглавление

[Документация к проекту «Название придумаю потом» 1](#_Toc85762839)

[1. Структура программы 3](#_Toc85762840)

[1.1. Древо классов для обработки информации (демо) 3](#_Toc85762841)

[2. Описание программных компонентов 4](#_Toc85762842)

[2.1. Class Information (демо) 4](#_Toc85762843)

[2.1.1. Class Text (демо) 4](#_Toc85762844)

[2.1.1.1. Class Article (демо) 4](#_Toc85762845)

[2.1.1.2. Class Document (демо) 4](#_Toc85762846)

[2.1.1.3. Class Book (демо) 4](#_Toc85762847)

[2.1.1.4. Class Table (демо) 4](#_Toc85762848)

[2.1.1.5. Class Social (демо) 4](#_Toc85762849)

[2.1.2. Class Media (демо) 4](#_Toc85762850)

[2.1.2.1. Class Image (демо) 4](#_Toc85762851)

[2.1.2.2. Class Video (демо) 5](#_Toc85762852)

[2.1.2.3. Class Audio (демо) 5](#_Toc85762853)

[2.2. File Control (демо) 5](#_Toc85762854)

[2.3. File Find (демо) 5](#_Toc85762855)

[3. Графический интерфейс 6](#_Toc85762856)

[4. Файловая система 7](#_Toc85762857)

[4.1. Формат хранения различных данных 7](#_Toc85762858)

[4.1.1. Научные, развлекательные и новостные статьи. (демо) 7](#_Toc85762859)

[4.1.2. Документы (демо) 7](#_Toc85762860)

[4.1.3. Книги (демо) 7](#_Toc85762861)

[4.1.4. Посты из социальных сетей(демо) 7](#_Toc85762862)

[4.1.5. Таблицы (демо) 7](#_Toc85762863)

[4.1.6. Изображения (демо) 7](#_Toc85762864)

[4.1.7. Видео (демо) 8](#_Toc85762865)

[4.1.8. Аудио (демо) 8](#_Toc85762866)

[4.2. Структура файловой системы 8](#_Toc85762867)

1. Структура программы

# Древо классов для обработки информации (демо)

Тут будет изображение древа. Сделаю позже.

1. Описание программных компонентов

# Class Information (демо)

Абстрактный класс, являющийся родительским для всех остальных информационных классов. Позже появится указания на методы.

## Class Text (демо)

Абстрактный класс, наследуемый от класса Information, является родительским для классов, которые обрабатывают все возможные виды текстовой информации. Позже появится указания на методы.

### Class Article (демо)

Класс, наследуемый от класса Text. Создан для работы со всевозможными статьями. Функционал состоит из превращения ссылок в удобочитаемые человеком документы (pdf or markdown). Позже появятся указания на методы.

### Class Document (демо)

Класс, наследуемый от класса Text. Создан для работы со всевозможными документами различных форматов. Функционал состоит из превращения ссылок в удобочитаемые человеком документы. Позже появятся указания на методы.

### Class Book (демо)

Класс, наследуемый от класса Text. Создан для работы с литературой, цитатами, учебниками и т.д. Позже появятся указания на методы и функционал.

### Class Table (демо)

Класс, наследуемый от класса Text. Создан для работы с таблицами, конвертации их в единый формат (csv) и сбора графиков и информации в понятном человеке виду. Позже появятся указания на методы и функционал.

### Class Social (демо)

Класс, наследуемый от класса Text. Создан для работы с постами в различных социальных сетях. Отделен от класса Article т.к. предположительно потребует других способов изъятия и сохранения информации. Позже появятся указания на методы и функционал.

## Class Media (демо)

Абстрактный класс, наследуемый от класса Information. Является родительским для классов, которые обрабатывают все возможные виды мультимедиа информации, такие как звук, видео, изображение и др. Позже появится указания на методы.

### Class Image (демо)

Класс, наследуемый от класса Media. Создан для работы с изображениями и превращения их в один вид (png). Позже появятся указания на методы.

### Class Video (демо)

Класс, наследуемый от класса Media. Создан для работы с видео и превращения их в один вид (AVI). Позже появятся указания на методы.

### Class Audio (демо)

Класс, наследуемый от класса Media. Создан для работы с аудиофайлами и превращения их в один вид (FLAC or WAV), сбора необходимой информации в текстовый формат. Позже появятся указания на методы.

# File Control (демо)

Файл с управляющим кодом, обеспечивающий связь кода и интерфейса. На вход от пользователя должен получать указание на входящий формат данных (Статья, картинка и т.д.), указание на формат получения доступа к файлу (ссылка или документ), непосредственно сама ссылка или документ. Должен запоминать полученные файлы, не получившие дополнительные данные для хранения (аннотации, краткое содержание, анализ и т.д.) и давать пользователю возможность их заполнить после добавления главного файла. Должен вести контроль дерева тегов, сохранять файлы и предоставлять доступ к ним в формате, который будет задан алгоритмом поиска по тегам. Связь с другими файлами и интерфейсом предположительно будет осуществляться посредством передачи данных в формате json.

Позже появится более подробное описание

# File Find (демо)

Файл с кодом, обеспечивающим поиск файлов по заданным тегам. Необходимо определиться с алгоритмом поле чего в описание добавить формат данных, которые должен получать н вход.

Позже появится более подробное описание

1. Графический интерфейс

Тут будет описание и инструкции по использованию графического интерфейса.

P.S. Предлагаю на первых парах остановиться на реализации телеграм бота. Быстро, просто, доступно из любой точки мира, дарит кроссплатформенность. В общем и целом, плюсов много.

1. Файловая система

# Формат хранения различных данных

## Научные, развлекательные и новостные статьи. (демо)

Предполагаемый формат хранения - директория, содержащая:

* Файл со статьей в формате pdf или markdown
* Аннотация к данной статье
* Файл со статьями и документами, рекомендуемыми к прочтению после и перед данной статьей (связные статьи, способствующие лучшему поймаю тематики и идей автора. Создан для того случая, когда такие статьи не связаны совпадением по тегам)
* Текстовое представление статьи, создаваемого для дальнейшей программной обработки (не обязательно человекочитаемый формат)

Ожидание более подробного описания

## Документы (демо)

Предполагаемый формат хранения – директория, содержащая:

* Непосредственно сам документ
* Текстовый эквивалент документа, созданный для дальнейшей программной обработки (необязательно человекочитаемый формат)

Ожидание более подробного описания

## Книги (демо)

Предполагаемый формат хранения – директория, содержащая:

* Саму книгу в формате pdf или epub
* Краткое содержание книги
* Рецензия на книгу (наиболее понравившаяся или написанная самостоятельная)
* Файл с выбранными цитатами (Придется либо подбирать удобный формат либо сохранять каждую отдельно как текстовые документы)

Ожидание более подробного описания

## Посты из социальных сетей(демо)

Формат хранения совпадает с форматом хранения, описанным в пункте 4.1.1

## Таблицы (демо)

Предполагаемый формат хранения – директория, содержащая:

* Сама таблица в формате csv
* Файл с краткими выводами, аналитикой таблицы и другими ее характеристиками

Ожидание более подробного описания

## Изображения (демо)

Предполагаемый формат хранения – директория, содержащая:

* Изображение в формате png
* Файл с информацией и описанием картинки (Создан для тех случаев, когда на изображении могут быть представлены графики или таблицы. В таком случае могут потребоваться аналитика и выводы в текстовом формате)

Ожидание более подробного описания

## Видео (демо)

Предполагаемый формат хранения – директория, содержащая:

* Непосредственно видео в формате AVI

## Аудио (демо)

Предполагаемый формат хранения – директория, содержащая:

* Аудиофайл в формате FLAC или WAV
* Расшифровка аудио (текст содержимого)

P.S. Для пунктов 4.1.6, 4.1.7, 4.1.8 стоит задуматься о делении форматов на разные тематики т.к. файлы с дополнительной информацией для данных одного формата. Но разной тематики могут очень сильно отличаться. В таком случае классы Audio, Video, Image станут абстрактными, и уже от них будем наследоваться дальше.

# 4.2. Структура файловой системы

Предположительно все данные представлены в виде директорий, лежащих в определенном заранее месте. Все названия представляют из себя сгенерированные hash коды. Hash коды используются для того что бы избежать повторений в названиях. Однако это сделает невозможным ручное управление данными. Стоит подумать над этим моментом.