**«Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова»**

**(БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | |  | | ДОПУСКАЕТСЯ К ЗАЩИТЕ: | | | | | | | | | |
| **Факультет** | | И | | Заведующий кафедрой | | | | |  | | И5 | | |
|  | |  | |  | | | | |  | | шифр кафедры | | |
| **Выпускающая кафедра** | | И5 | | Скулябина О.В. | | |  |  | | | | | |
|  | |  | | Фамилия И.О. | | |  | подпись | | | | | | |
| **Группа** | | И976 | | «\_\_\_\_\_» | |  | | | | | | 2021 г. | |
|  |  | |  | |  | |  | | | | |  | | |

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА**

|  |  |
| --- | --- |
| Кравца Марка Анатольевич | |
| Фамилия, имя, отчество обучающегося | |
| **На тему** | Разработка приложения для конвертации файлов формата DOCX |
| в файлы формата ODT | |
|  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Направление подготовки** | 09.03.04 |  | Программная инженерия | |
|  | индекс направления |  | | полное наименование направления |
|  | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Консультант:** | | | | |  | | | |  | **Руководитель:** | | |  | | | | | | |
| при необходимости | | | | | подпись | | | |  |  | | | подпись | | | | | | |
|  | | |  | |  | | | |  | ст. преподаватель | | | | |  | | Бармина А.А. | | |
| ученая степень, ученое звание | | | | | Фамилия ИО | | | |  | ученая степень, ученое звание | | | | | | Фамилия ИО | | | |
| « » |  |  | | | | |  | 2021 г. |  | « » |  |  | | | | | |  | 2021 г. |
|  | | | | | |  | | |  | **Обучающийся:** | | | |  | | | | | |
|  | | | |  | |  | | |  |  | | | | |  | | Кравец М.А. | | |
|  | | | | | |  | | |  | подпись | | | | |  | | Фамилия ИО | | |
|  |  |  | | | | |  |  |  | « » |  |  | | | | | |  | 2021 г. |

**РЕФЕРАТ**

Пояснительная записка с., рис., табл., источн., прил.

ZZZZ, ZZZZ, ZZZZ

Целью выпускной квалификационной работы является разработка приложение для преобразования файлов формата DOCX в файлы формата ODT.

Объектом исследования является процесс преобразования файлов.

Предметом исследования является структура файлов для преобразования.

В ходе работы была проанализирована предметная область преобразования файлов формата DOCX в файлы формата ODT, изучены требования пользователей к системе, рассмотрены проблемы предметной области, а также выдвинуты функциональные требования к системе.

Помимо этого, были выбраны шаблоны проектирования и инструменты разработки и определены системные требования.

На основании принятых проектных решений разработана программное обесепечения для преобразования файлов формата DOCX в файлы формата ODT.

**СОДЕРЖАНИЕ**

[ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ 4](#_Toc66540937)

[ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ 5](#_Toc66540938)

[ВВЕДЕНИЕ 6](#_Toc66540939)

[1 Анализ предметной области 8](#_Toc66540940)

[1.1 Open Office XML 8](#_Toc66540941)

[1.1.1 Структура Open Office XML файла 9](#_Toc66540942)

[1.2 Open Document Text 11](#_Toc66540943)

[1.2.1 Структура OpenDocument файла 12](#_Toc66540944)

[1.3 Описание методов решения подобных задач 15](#_Toc66540945)

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящей пояснительной записке применяют следующие термины и определения:

ZIP — формат архивации файло или файла и сжатие данных без потерь

Метаданные — это совокупность данных описывающие первоначальное состояние данных.

XML (англ. eXtensible Markup Language) — расширяемый язык разметки.

Фреймворк — программное обеспечение, облегчающее разработку и объединение разных компонентов большого программного проекта.

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящей пояснительной записке применяют следующие сокращения и обозначения:

АЦП – аналогово-цифровой преобразователь

АЭС – атомная электростанция

ВКР – выпускная квалификационная работа

ОС – операционная система

ПК – персональный компьютер

ПО – программное обеспечение

ТЭС – тепловая электростанция

ЭДС – электродвижущая сила

# ВВЕДЕНИЕ

В наше время большинство компаний и государственных учереждений переходят на формат файлов ODT для своих документов, и нужно быстрое решение для ковертации файлов формата DOCX в форматы ODT.

В рамках выпускной квалификационной работы будет описано создание приложения для преобразовании файлов форматов DOCX в ODT.

Существующие аналоги распространяются либо по платной модели, либо по модели пробного использования(free-trial), где возможности приложения до покупки продукта сильно ограничены.

В данной работе объектом исследования является изменение формата файла, а предметом исследования является процесс преобразования файлов.

Цель выпускной квалификационной работы — разработать программу для преобразования файлов форматов DOCX в ODT.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать проблему предметной области;
2. Провести сравнение программ-аналогов;
3. Спроектировать систему преобразования файлов;
4. Разработать пользовательский интерфейс;
5. Реализовать готовый продукт (приложение);
6. Провести тестирование программного продукта.

В результате выполнения вышеописанных задач должно быть разработано готовое приложение которое осуществляет преобразование файлов формата DOCX в файлы формата ODT.

Выпускная квалификационная работа состоит из введения, четырех разделов и заключения.

В первом разделе описана постановка задачи, и изучены существующие программы-аналоги, а также рассмотрены их достоинства и недостатки.

Во втором разделе содержит информацию о процессе проектирования программного продукта и программной реализации продукта.

Третий раздел описывает процесс разработки пользовательского интерфейса. Будут рассмотрен принцип работы созданного программного обеспечения.

В четвертом разделе описано тестирование программного продукта с использование unit-тестов.

Анализ предметной области

В ВКР требуется реализовать программу для преобразования файлов из формата DOCX в формат ODT. Программа должна иметь удобный пользовательский интерфейс, возможность преобразования файлов без доступа к сети интернет. Основной идеей данной задачи являлся переход с платного пакета Microsoft Office на бесплатный OpenOffice.

## Open Office XML

Open Office XML (DOCX) — это новая версия формата файла DOC, выпущенная в 2007 году, которая была модифицырованна старая версия формата .doc добавленнием языка XML. Обычный файл DOCX является архив состоящий из XML файлов. Формат файла DOCX является одним из самых используемых во всем мире, т.к в мире большинство компаний используют Microsoft Office. Одно из преимуществ формата DOCX над DOC является размер файла рисунок 1.Так же формат DOCX стал стандартом для Word в версии программного обеспечения 2007 года.

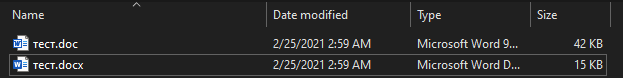


Рисунок 1 — Сравнение форматов DOC и DOCX

Базовый пакет файла DOCX представленная на рисунке 2.

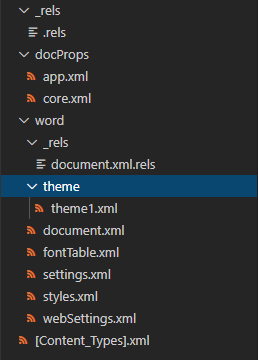


Рисунок 2 — Базовый пакет DOCX файла

### Структура Open Office XML файла

Структура DOCX является ZIP архивом состоящим из XML файлов. XML файлы делять на:

— файлы метаданных,

— документ.

Файлы мета-данных содержат информацию о других файлах содержашихся в этом архиве. В документе находится информация о фактическом содержании документа.

Так же схема пакета DOCX включает в себя два элемента:

— Компонент(англ. Part),

— Отношения(англ. Relationships).

Компонент (англ. Part) — является основной еденицей хранения любого содержимого документа. Например :

— XML,

— VIDEO,

— IMG,

— META-DATA,

— и т.п.

Отношения (англ. Relationships) — определяют назначение каждой связи между компонентами и пакетом в целиком. Пример кода отношения представлен на рисунке 4. Ссылка на связи документа находится в файле DOCUMENT.XML.RELS.

На рисунке 3 представленна сама структура контейнеров документа DOCX.

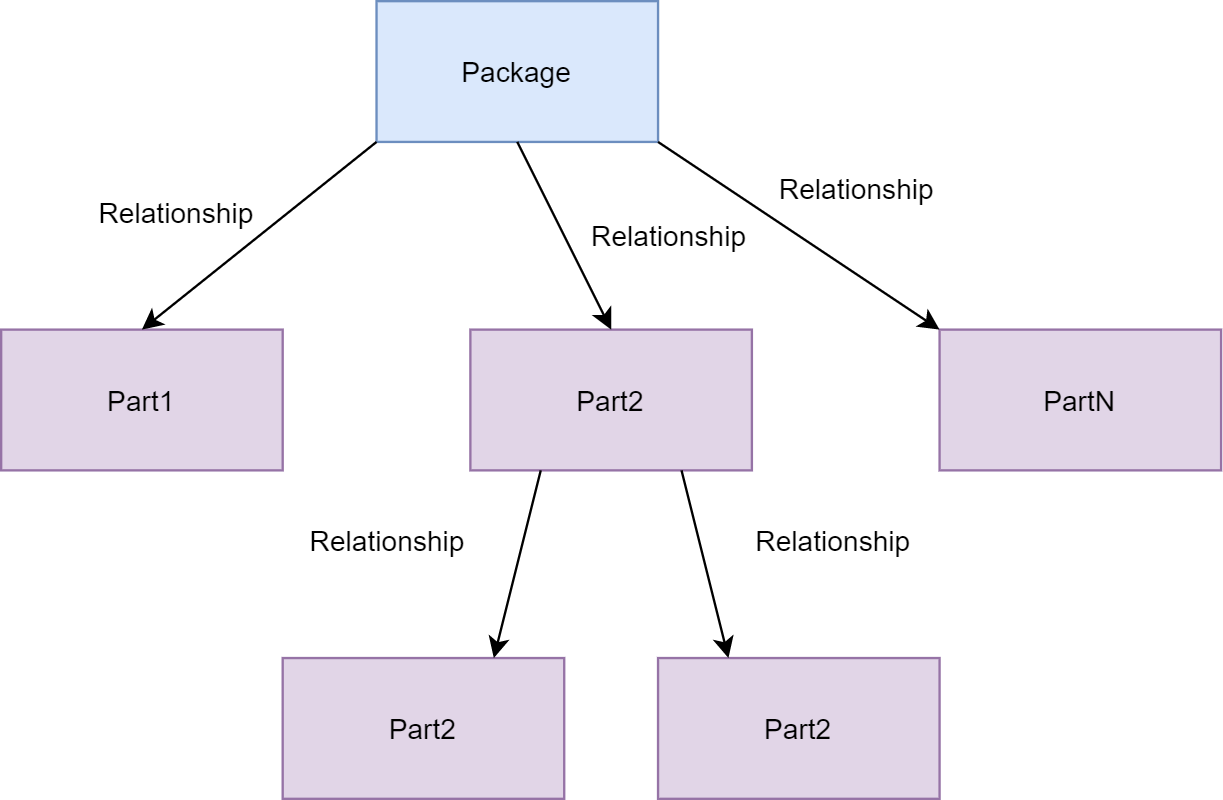


Рисунок 3 — Структура документа DOCX

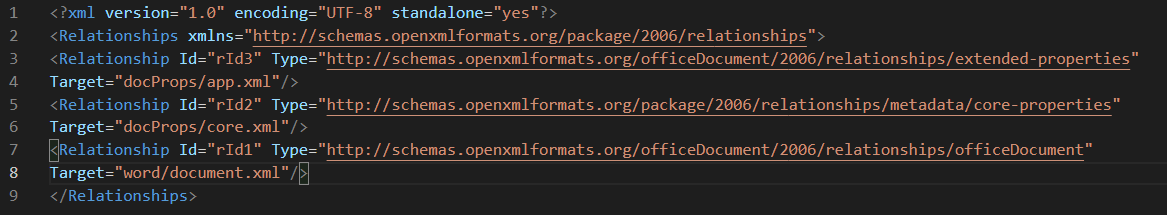


Рисунок 4 — Код отношения

В основном файле DOCUMENT.XML представленны все элементы данного документа рисунок 5.

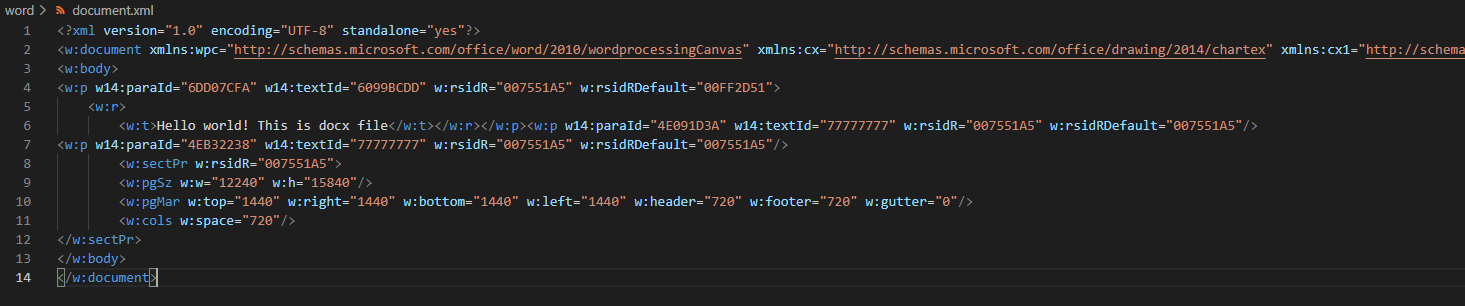


Рисунок 5 — Код файла DOCUMENT.XML

Главный элемент <w:document> показывает основное содержание данного файла. Элемент <w:body> является частью скелета документа, который может в себе содержать параграфы, таблицы и разделы. Элемент <w:p> освновной контейнер контента документа. Элемент <w:r> содержит техт параграфа.

## Open Document Text

OpenDocument Text (ODT) — это тип документов, созданных с помощью текстовых редакторов, основанных на формате текстового файла OpenDocument. Они создаются с помощью приложений для текстовых редакторов, таких как бесплатный OpenOffice Writer. Сравнение размера файла .odt и .docx представленно на рисунке 6.



Рисунок 6 — Сравнение форматов DOCX и ODT

Базовый пакет ODT файла представлен на рисунке 7.

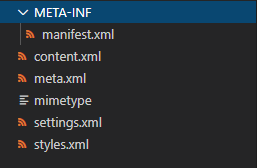


Рисунок 7 — Базовый пакет ODT файла

Формат ODT имеет открытый исходный код, что позволяет разобраться в его работе. Одним из главных плюсов формата ODT является безопасность. При использовании программы Microsoft Word, идет сбор данных во время работы пользователя, тогда как пр ограммы, использующие формат ODT, обычно этим не занимаются. Перечень дополнительных отличий приведен ниже:

— ODT возможно, более безопасен, посколько все больше разработчиков исправляют дыры в безопасности на регулярной основе,

— ODT можно использовать в любой операционной системе, в то время как .docx доступный только на компьютерах использующий Microsoft Word,

— программы использующие ODT формат обычно бесплатны, в отличии от программ использующий DOCX,

— документы ODT обновляются сообществом, а не одной определенной коропорацией.

### Структура OpenDocument файла

OpenDocument имеет два метода построения документа:

— одним XML файлом,

— набором файлов в архиве, каждый из которых хранит часть полного документа.

Элемент <office:document> является основным элементом документа в OpenDocument который представлен одним XML файлом, в котором находится все содержимое документа.

Когда OpenDocument документ представлен архивом, он будет состоять из четырех основных элементов <office:document-content>, <office:document-styles>, <office:document-meta> и <office:document-settings>. Каждый из них хранится в отдельном файле. Так же в архиве могут хранится изображения, встроенные объекты, видео и файлы зависимости исполнения.

Элемент <office:document-content> хранит в себе содержимое и автоматические стили используемые в документе. Данный элемент находится в файле CONTENT.XML. Содержимое данного файла представленно на рисунке 8.

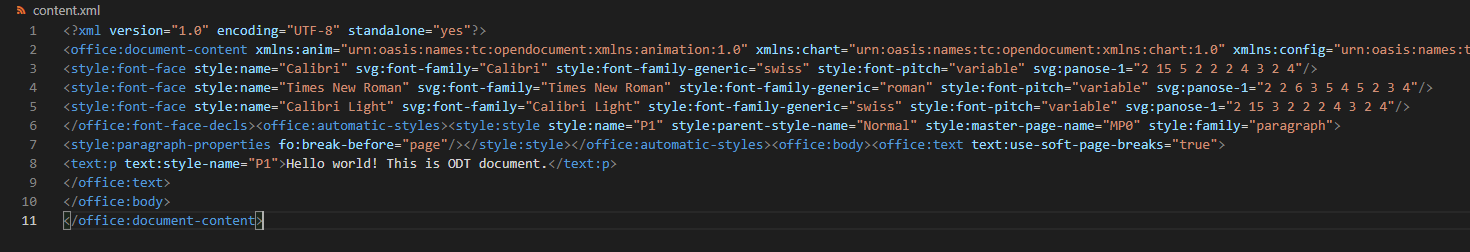
****

Рисунок 8 — Содержимое файла CONTENT.XML

Элемент <office:document-styles> хранит в себе стили используемые в контенте документа и автоматические стили используемые в документе. Данный элемент находится в файле STYLES.XML. Содержимое данного файла представненно на рисунке 9.

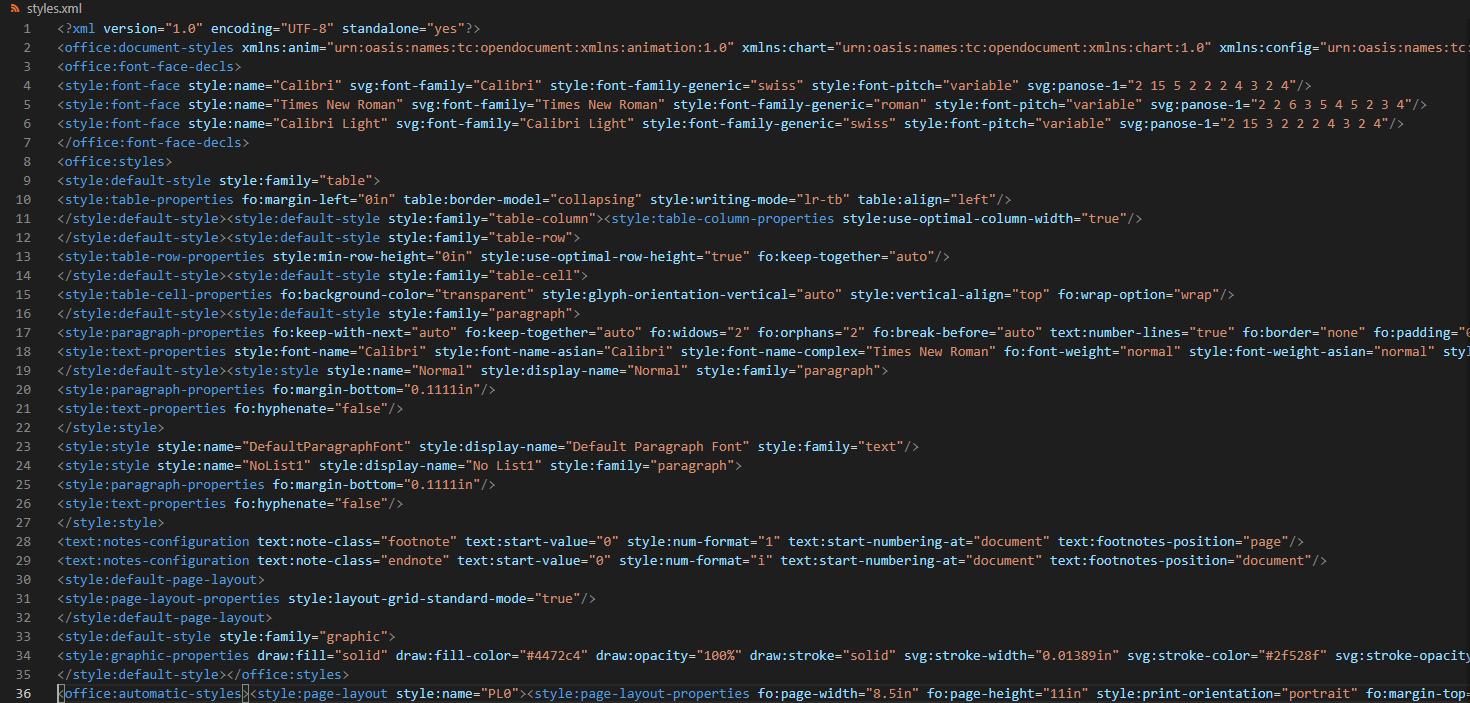


Рисунок 9 — Содержимое файла STYLES.XML

Элемент <office:document-meta> хранит в себе метаданные документа. Данный элемент находится в файле META.XML. Содержимое данного файла представленно на рисунке 10.

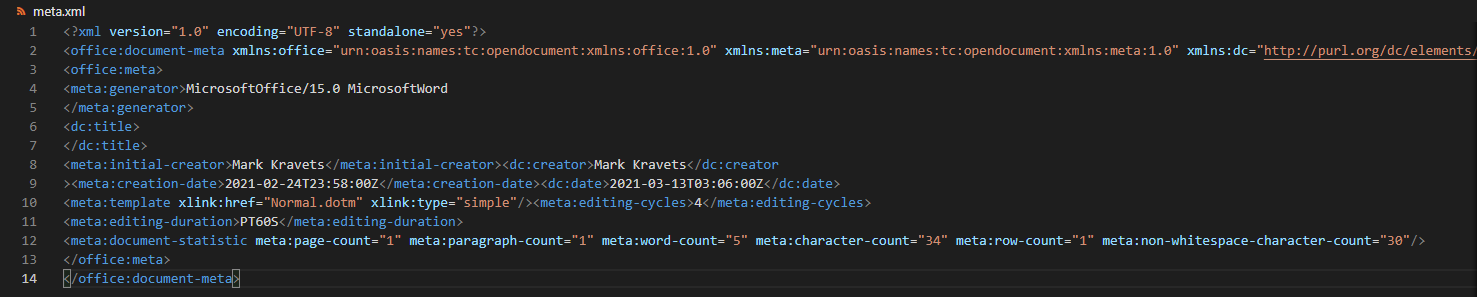


Рисунок 10 — Содержимое файла META.XML

Элемент <office:document-settings> хранит в себе зависимости исполнения настроек. Данный элемент находиится в файле SETTINGS.XML. Содержимое файла представленно на рисунке 11.

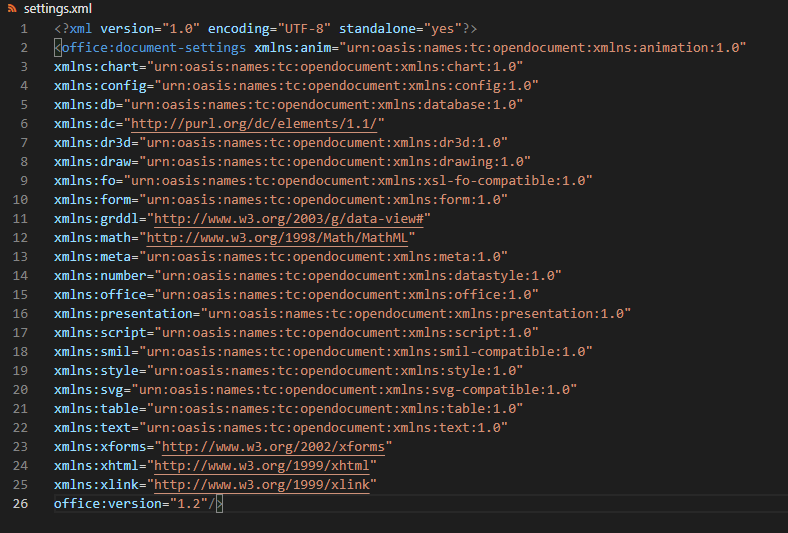


Рисунок 11 — Содержимое файла SETTINGS.XML

Модель представляющая архив файла OpenDocument представленна в таблице 1.

Таблица 1 — Модель архива файла OpenDocument

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Основной элемент | <office: meta> | <office: settings> | <office:scripts> | <office: font-face-decls> | <office:styles> | <office:automatic-styles> | <office:master-styles> | <office:body> | |
| <office:document > | X | X | X | X | X | X | X | X |
| <office:document-content > |  |  | X | X |  | X |  | X |
| <office:document-styles |  |  |  | X | X | X | X |  |
| <office:document-meta> | X |  |  |  |  |  |  |  |
| <office:document-settings> |  | X |  |  |  |  |  |  |

## Описание методов решения подобных задач

«Doxilion Document Converter»‎ — является программным обеспечением распространяемым по методу free-trial, и следовательно имеет ограниченные функции для использования. Для того что бы использовать ее в больших компаниях понадобится покупка лицензии на использование программного обеспечения. Программа может конвертировать файлы .docx, .doc и .odt. Еще одним недостатком данного ПО является осутствие изменение названия выходного файла.

На рисунке 12 представелен графический интерфейс ПО Doxilion Document Converter. Так же на рисунке 13 представлено меню выбора формата файла, пути и открытия.

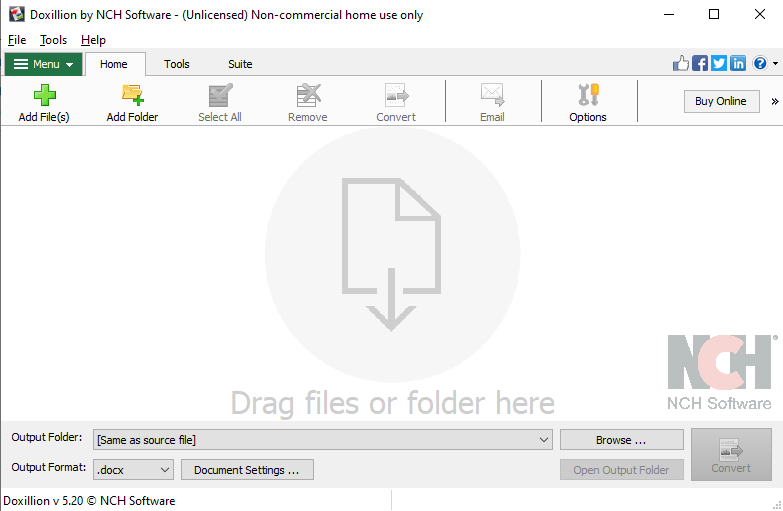


Рисунок 12 — Графический интерфейс ПО Doxilion Document Converter

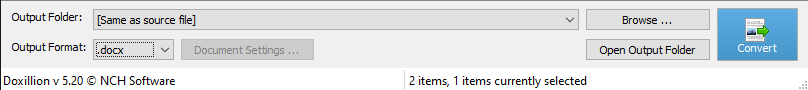


Рисунок 13 — Меню выбора формата файла, пути и открытия

«Pandoc» это утилита использующая командную строку для конвертации файлов из множества форматов. Является бесплатным программным обеспечение, возможно редактирование названия файла. Недостатком является отстутсвие встроеного пользовательского интерфейса и руководства пользователя. На рисунке 14 показан пример команд для конвертации файлов. На рисунке 15, 16 и 17 показан результат конвертации и содержимое файлов.

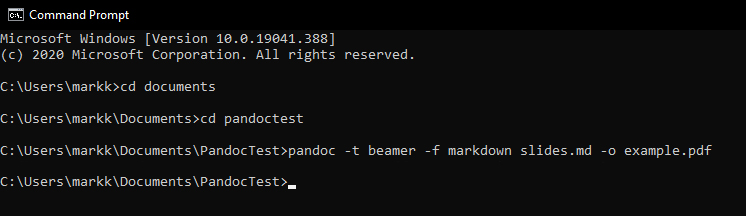


Рисунок 14 — Пример команд Pandoc для конвертации файлов



Рисунок 15 — Результат конвертации Pandoc

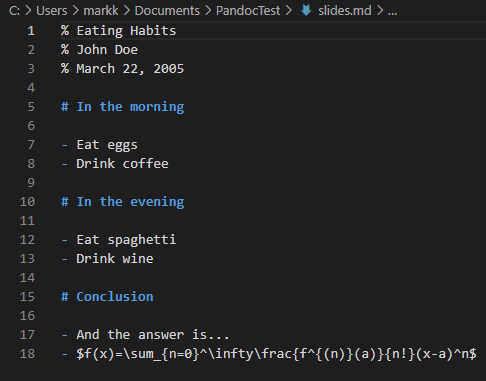


Рисунок 16 — Содержимое файла SLIDES.MD



Рисунок 17 — Содержимое преобразованного файла EXAMPLE.PDF

Онлайн-конверторы являются самым популярным средством конвертации файлов среди пользователей. Множество вариаций именения файлов, так же являются бесплатным решением. Из недостатков:

— нужен постоянный доступ к интернету,

— документы которые пользователь конвертирует использую сайт, остаются в базе данных сайта, что не целесообразно для некоторых компаний в плане безопасности.

На рисунке 18 представлен интерфейс сайта для преобразования файлов. На рисунке 19 представлен интерфейс скачивания преобразованного файла.

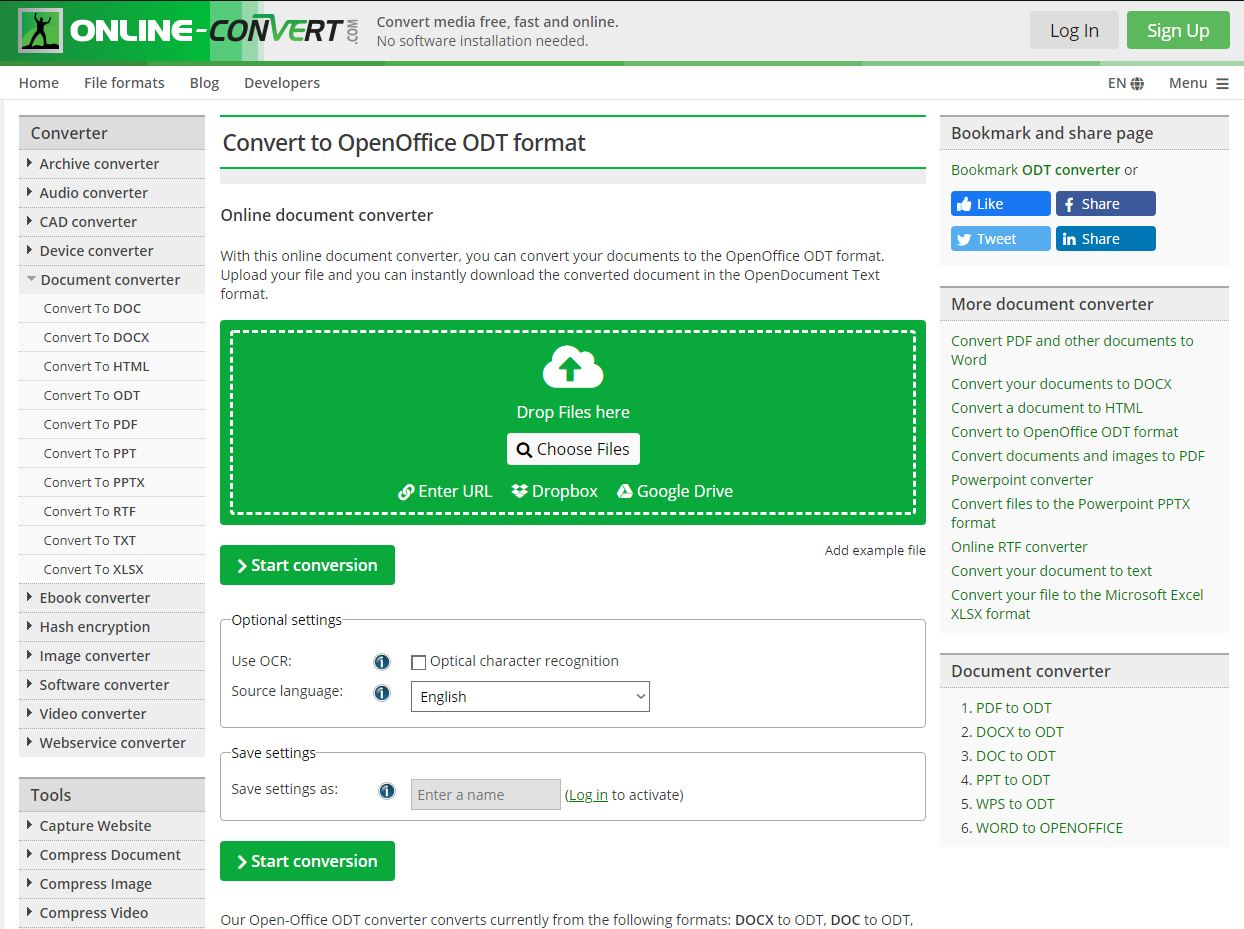


Рисунок 18 — Интерфейс сайта для преобразования файлов

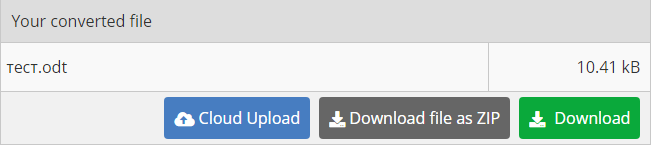


Рисунок 19 — Интерфейс скачивания преобразованного файла

«Universal Document Converter» тоже распространяется по методу free-trial, следовательно чтобы получить доступ к всем опциям недоступных в бесплатной версии. Так же одним из больших минусов является, то что последнее обновления ПО было в 2017 году. Так же невозможна установка данного ПО из-за ошибки в инсталяции программы на компьютер. Ошибка инсталяции представленна на рисунке 20.

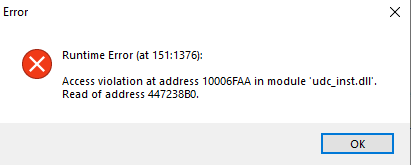


Рисунок 20 — Ошибка инсталяции Universal Document Converter