Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

«Уральский федеральный университет

имени первого Президента России Б. Н. Ельцина»

Институт радиоэлектроники и информационных технологий - РТФ

Школа бакалавриата

**К ЗАЩИТЕ ДОПУСТИТЬ**

Руководитель образовательной программы

09.03.04 Программная инженерия

И. Н. Обабков

« 02 » июня 2022 г.

**Разработка приложения, определяющего соответствие документа заданному ГОСТу**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Пояснительная записка

Руководитель: к. т. н., доцент \_\_\_\_\_\_\_\_\_ А. В. Корелин

Нормоконтролер: \_\_\_\_\_\_\_\_\_ И. С. Колмогорцева

Обучающийся: группа РИ-480012 \_\_\_\_\_\_\_\_\_ Г. С. Емельянов

Екатеринбург – 2022

Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

«Уральский федеральный университет

имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий – РТФ

Школа бакалавриата

Направление (специальность) 09.03.04 Программная инженерия

УТВЕРЖДАЮ

Руководитель образовательной программы 09.03.04 Программная инженерия

И. Н. Обабков .

( подпись) (Ф.И.О.)

« 20 » декабря 2021 г.

**ЗАДАНИЕ**

на выполнение выпускной квалификационной работы

обучающегося \_\_Емельянова Глеба Сергеевича\_\_\_\_\_\_\_\_\_\_\_группы РИ-480012

(фамилия, имя, отчество)

1 Тема ВКР Разработка приложения, определяющего соответствие документа заданному ГОСТу

Утверждена распоряжением по институту от « 15 » декабря 2021 г. № 33.02-05/377

2 Руководитель Корелин Андрей Викторович, доцент, канд. техн. наук.

(Ф.И.О., должность, ученое звание, ученая степень)

3 Исходные данные к работе Материалы, полученные в ходе преддипломной практики, ГОСТы, техническая литература \_\_\_\_\_\_\_\_\_\_\_\_

4 Содержание пояснительной записки (перечень подлежащих разработке вопросов)

1. Обзор предметной области

2. Анализ структуры DOCX документа

3. Проектирование прототипа приложения

4. Разработка прототипа приложения

5 Перечень демонстрационных материалов Презентация в MS Powerpoint

6 Календарный план

|  |  |  |
| --- | --- | --- |
| Наименование этапов  выполнения работы | Срок выполнения этапов работы | Отметка  о выполнении |

|  |  |  |
| --- | --- | --- |
| Обзор предметной области | 15.02.22 | выполнено |
| Анализ структуры DOCX документа | 15.03.22 | выполнено |
| Проектирование прототипа приложения | 15.04.22 | выполнено |
| Разработка прототипа приложения | 17.05.22 | выполнено |

Руководитель . . Корелин Андрей Викторович

(подпись) Ф.И.О.

Задание принял к исполнению . .

(подпись обучающегося)

7 Выпускная квалификационная работа закончена « 18 » мая 2022 г.

Пояснительная записка и все материалы просмотрены

Считаю возможным допустить Емельянова Глеба Сергеевича к защите его выпускной квалификационной работы в экзаменационной комиссии.

Руководитель . (подпись)

8 Допустить Емельянова Глеба Сергеевича к защите выпускной квалификационной работы в экзаменационной комиссии (распоряжение по кафедре ИИТ № 100.063-03/10 от « 02 » июня 2022 г.)

Руководитель образовательной программы

09.03.04 Программная инженерия . И. Н. Обабков

(подпись) Ф.И.О.

**РЕФЕРАТ**

Выпускная квалификационная работа бакалавра 53 с., 9 рис., 3 табл., 24 источника, 10 прил.

OPEN XML, WEB ПРИЛОЖЕНИЕ, OPEN XML SDK, ПРОЕКТ ПО РАЗРАБОТКЕ ПРОГРАММНОГО ПРОДУКТА, ФОРМАТ DOCX, ПРОЕКТИРОВАНИЕ ПРОГРАММНОГО ПРОДУКТА, ГОСТ 7.32–2017.

Цель работы – разработать приложение, которое помогало бы обычным пользователям, не знакомым с государственными стандартами оформления документов, корректно оформить необходимые им бумаги.

Объект исследования формат Office Open XML.

Методы исследования: анализ, системный подход, метод индукции, метод абстрагирования, метод наблюдений.

Результат работы: разработан прототип приложения, проверяющего соответствие документа выбранному стандарту.

Выпускная квалификационная работа выполнена в текстовом редакторе Microsoft Word и представлена в электронном формате.

**ОГЛАВЛЕНИЕ**

[ВВЕДЕНИЕ 5](#_Toc103790351)

[1 ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ 7](#_Toc103790352)

[1.1 Обзор основных форматов электронных документов 7](#_Toc103790353)

[1.2 Обзор Open XML и Open XML SDK 11](#_Toc103790354)

[1.3 Анализ конкурентов 13](#_Toc103790355)

[2 АНАЛИЗ СТРУКТУРЫ DOCX ДОКУМЕНТА 15](#_Toc103790356)

[2.1 Файлы входящие в состав простейшего docx документа 15](#_Toc103790357)

[2.2 Структура файла document.xml 17](#_Toc103790358)

[3 ПРОЕКТИРОВАНИЕ ПРОТОТИПА ПРИЛОЖЕНИЯ 21](#_Toc103790359)

[3.1 Выбор формата взаимодействия с пользователем 21](#_Toc103790360)

[3.2 Проектирование серверной части приложения 23](#_Toc103790361)

[3.3 Проектирование клиентской части приложения 28](#_Toc103790362)

[4 РАЗРАБОТКА ПРОТОТИПА ПРИЛОЖЕНИЯ 30](#_Toc103790363)

[4.1 Создание класса редактора 30](#_Toc103790364)

[4.2 Создание класса ошибки 33](#_Toc103790365)

[4.3 Создание класса стандарта 34](#_Toc103790366)

[4.4 Разработка серверных контроллеров 41](#_Toc103790367)

[4.5 Разработка клиента приложения 45](#_Toc103790368)

[4.6 Тестирование приложения 47](#_Toc103790369)

[ЗАКЛЮЧЕНИЕ 49](#_Toc103790370)

[БИБЛИОГРАФИЧЕСКИЙ СПИСОК 51](#_Toc103790371)

[ПРИЛОЖЕНИЕ А 54](#_Toc103790372)

[ПРИЛОЖЕНИЕ Б 55](#_Toc103790373)

[ПРИЛОЖЕНИЕ В 56](#_Toc103790374)

[ПРИЛОЖЕНИЕ Г 57](#_Toc103790375)

[ПРИЛОЖЕНИЕ Д 58](#_Toc103790376)

[ПРИЛОЖЕНИЕ Е 59](#_Toc103790377)

[ПРИЛОЖЕНИЕ Ж 60](#_Toc103790378)

[ПРИЛОЖЕНИЕ И 61](#_Toc103790379)

[ПРИЛОЖЕНИЕ К 62](#_Toc103790380)

[ПРИЛОЖЕНИЕ Л 64](#_Toc103790381)

**ВВЕДЕНИЕ**

В наше время всем людям так или иначе приходится сталкиваться с оформлением документов. Это, не говоря уже о представителях профессий, непосредственно связанных с работой с официальными бумагами, таких как секретари, юристы или же научные сотрудники.

Обычным людям наиболее часто приходится работать с документами во времена студенчества, когда появляется необходимость в оформлении рефератов, отчётов, научно-исследовательских работ и т. д. В то же время стиль оформления таких бумаг строго формализован и описан во множестве государственных стандартов, поэтому обывателям зачастую бывает сложно разобраться во всех тонкостях написания официальных документов.

В то же время на рынке программных продуктов практически отсутствуют решения данной проблемы. Немногочисленные же сервисы проверки документов на соответствие ГОСТу являются платными и недоступны большинству обучающихся.

Поэтому была поставлена цель – разработать приложение, которое помогало бы обычным пользователям, не знакомым с государственными стандартами оформления документов, корректно оформить необходимые им бумаги.

Для достижения назначенной цели были выделены следующие задачи:

* проанализировать структуру docx документов;
* спроектировать архитектуру приложения;
* реализовать прототип приложения;
* протестировать работоспособность приложения.

Объектом исследования является формат Office Open XML.

В качестве теоретической и информационно аналитической базы была использована официальная документация к формату Open XML, официальная документация к Open XML SDK, государственные стандарты Российской Федерации диктующие правила оформления документов, статьи российских исследователей по этой теме, информационные материалы, размещённые на сайтах в сети Интернет, а также методические указания Уральского федерального университета по оформлению научно-исследовательских работ.

1. **ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ**
   1. **Обзор основных форматов электронных документов**

Существует множество различных форматов хранения электронных документов. Каждый из них обладает своим назначением, областью применения, своими плюсами и минусами. Для того чтобы понять почему большинство людей в своей работе с документами на компьютере используют формат DOCX, необходимо рассмотреть ключевые особенности самых популярных из них.

В рамках данной работы интерес представляют те форматы хранения электронных документов, которые можно легко редактировать и создавать, это является крайне важным критерием при написании электронных документов, а также поддерживают стили форматирования текста.

Таким образом в данной работе не будут рассмотрены форматы по типу PDF, который используется в первую очередь для хранения и передачи документов и сложно поддаётся редактированию [1]. А также самые примитивные текстовые форматы по типу TXT.

Рассмотрим несколько самых популярных форматов электронных документов:

* DOC, проприетарный формат, созданный Microsoft для своего текстового процессора Microsoft Word;
* DOCX, открытый формат, созданный Microsoft на смену устаревающего DOC;
* ODT, открытый формат, разработанный сообществом OASIS.

Для них сразу можно отметить, что DOC на данный момент является устаревшим форматом, и несмотря на то, что он пользуется значительной популярность среди пользователей, используют его чаще всего для работы со старыми электронными документами.

Общую информацию касающуюся каждого из форматов можно увидеть в таблице 1.

Таблица 1 – Сравнение форматов DOCX и ODT

|  |  |  |  |
| --- | --- | --- | --- |
| Критерий | DOCX | ODT | DOC |
| Создатель | Microsoft | OASIS | Microsoft |
| Год создания | 2006 | 2006 | 1990-е |
| Открытый формат | да | да | нет |
| Формат файла | Office Open XML | Open Document Format | бинарный |
| Стандарт | ECMA-376 | ISO/IEC 26300:2006 | – |

Главное отличие форматов DOCX и ODT от DOC заключается в формате файлов. DOCX и ODT основываются на XML и представляют из себя ZIP архивы, содержащие внутри файлы разметки и медиафайлы. DOC же в свою очередь является бинарным, что крайне негативно сказывается на размере файлов данного формата, и возможностях сжатия [2]. Сравнительный график, показывающий разницу в занимаемом документами месте на диске, можно увидеть на рисунке 1.

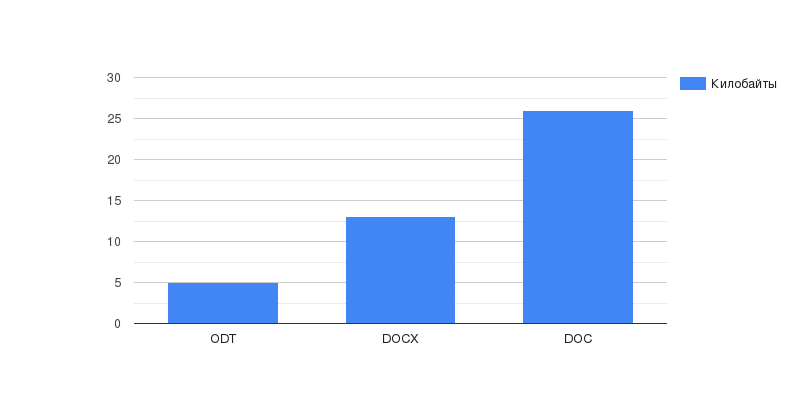


Рисунок 1 – Занимаемое место на диске

На представленном выше рисунке видно, что файл в формате DOC занимает более чем в пять раз больше дискового пространства чем, сохранённый в формат ODT, и в два раза больше места, чем сохранённый в формате DOCX. В качестве текста в рассматриваемых файлах была представлена панграмма «съешь ещё этих мягких французских булок, да выпей чаю» без каких-либо настроек форматирования.

Для того чтобы определить, насколько популярен каждый из форматов, узнаем, как много файлов, содержащих в своём названии цифру 1, удастся найти при помощи поиска Google в разных временных промежутках, используя команду filetype [3]. Полученные данные отображены в таблице 2.

Таблица 2 – Популярность форматов в Google поиске

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Запрос | Всего | За последний год | За последние 5 лет | За последние 10 лет |
| 1 filetype:docx | 17,300,000 | 1,310,000 | 3,130,000 | 3,560,000 |
| 1 filetype:doc | 43,000,000 | 954,000 | 3,130,000 | 7,280,000 |
| 1 filetype:odt | 367,000 | 3,860 | 11,800 | 17,000 |

Из данных таблицы можно заметить, что наибольшей популярностью на данный момент всё ещё пользуется формат DOC, однако как не сложно заметить DOCX в последние годы занимает лидирующую позицию по показателям скорости распространения. ODT при этом показывает результаты значительно меньшие чем его конкуренты.

Благодаря тем же поисковым запросом можно также получить статистику о том, как много файлов каждого формата было за гружено в каждый отдельный год. В качестве временного промежутка используем 2013–2018 года. Полученную статистику можно увидеть на рисунке 2.

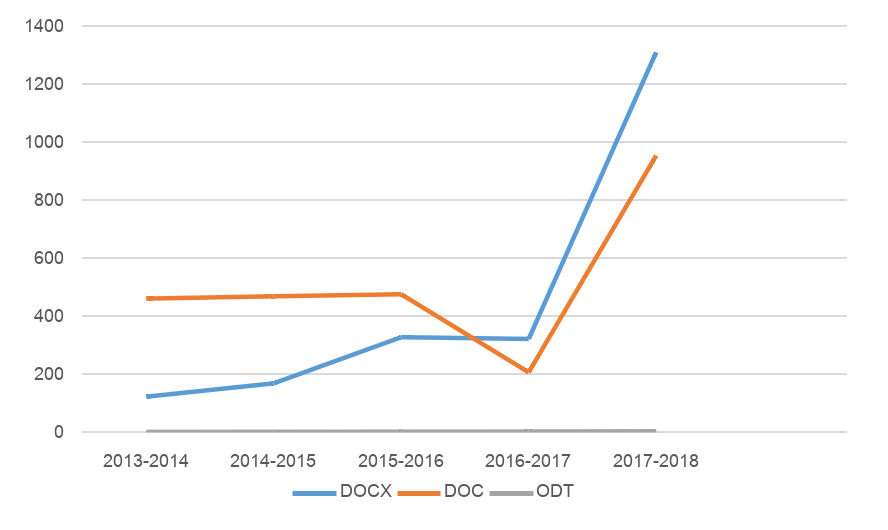


Рисунок 2 – Статистика использования форматов за 2013–2018 годы

Так как количество файлов в формате ODT загруженных в данные промежутки времени несоизмеримо меньше его рассматриваемых конкурентов, был составлен ещё один график отдельно для этого формата, его можно увидеть на рисунке 3.

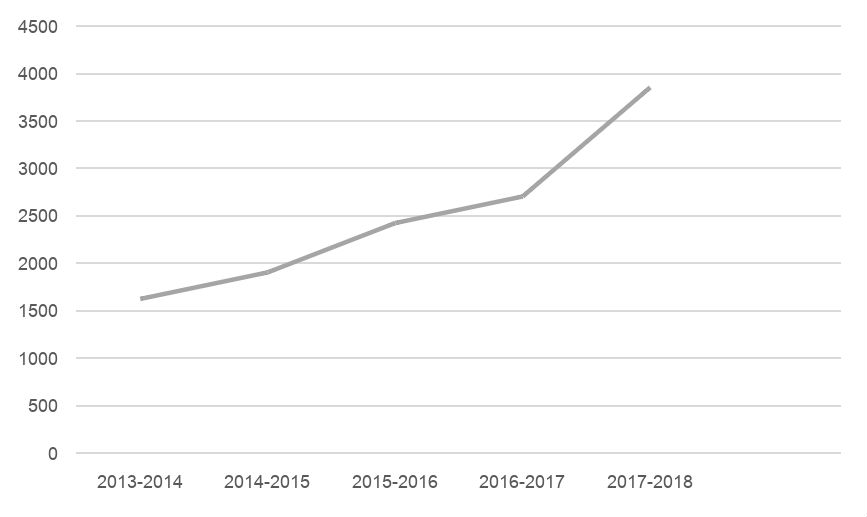


Рисунок 3 – Статистика использования формата ODT за 2013–2018 годы

Как видно из представленных графиков, наибольший рост популярности наблюдается у формата DOCX. DOC хоть и остаётся на данный момент крайне популярным, всё же начинает уступать своему более продвинутому наследнику. Формат же ODT не может конкурировать с другими двумя, ни в распространённости, ни в темпах роста популярности.

На основании полученных данных для работы был выбран формат DOCX как наиболее современный и распространённый.

* 1. **Обзор Open XML и Open XML SDK**

Так как основным форматом хранения электронных документов в данной работе был выбран DOCX следует более подробно рассмотреть формат файла, лежащий в его основе, а также существующие инструменты для работы с ним.

Формат Open XML был разработан Microsoft как ответ набирающему популярность открытому формату Open Document Format, разрабатываемому консорциумом OASIS. На момент 2005 начала набирать популярность концепция применения в документообороте госструктур открытых форматов файлов, на то время, в распоряжении Microsoft были лишь проприетарные форматы DOC, XLS и т. д.

Формат Microsoft оказался более конкурентноспособным, что было продемонстрировано ранее, однако он обладает рядом недостатков, затрудняющих работу с ним в области разработки программного обеспечения, а именно:

* сертификация стандарта, как отмечают юристы из Software Freedom Law Center, даёт право на свободное использование кода только в рамках стандарта, однако не позволяет его использовать в работах, не связанных с ним напрямую [4];
* документация к формату Open XML содержит в себе более 5000 страниц, что крайне затрудняет его изучение и использование в написании приложений, базирующихся на нём [5];
* формат использует устаревшую систему расчёта времени, не поддерживающую часовые пояса, и переход на летнее время.

В связи с перечисленными фактами работать напрямую с форматом Open XML представляется затруднительным. Поэтому Microsoft в 2008 году выпускает SDK для формата Open XML, позволяющий проводить различные операции с документами его поддерживающими. С помощью Open XML SDK можно создавать документы, редактировать их, проверять их отдельные части и многое другое [6].

Open XML SDK определяет форматы текстовых документов, электронных таблиц и презентаций в форме определённых языков разметки, таких как [7]:

* WordprocessingML, использующийся для работы с текстовыми документами;
* SpreadsheetML, использующийся для работы с электронными таблицами;
* PresentationML, использующийся для работы с презентациями.

В рамках данной работы будет использоваться WordprocessingML так как именно он отвечает за работу с документами в формате DOCX. Данный язык разметки представляет XML элементы в соответствии с DOM моделью, что позволяет легко работать с ними в коде.

В данном разделе был проведён краткий обзор формата Open XML и пакета разработки программного обеспечения для работы с ним. Так как спецификация формата, ввиду своего объёма, лишь затруднила бы предстоящую разработку, было принято решение об изучении структуры документа и поиска необходимых для работы элементов на практике.

* 1. **Анализ конкурентов**

В рамках выбранного формата хранения электронных документов, а также формата файла, на котором он базируется, имеется возможность рассмотреть существующие работы каким-либо образом, связанные с рассматриваемой темой.

По результатам поиска среди работ российских авторов с использованием сервиса «КиберЛенинка», мною не было найдено аналогов, соответствующих или частично соответствующих заявленной теме работы [8]. Однако было найдено несколько статей, посвящённых генерации и редактированию текстовых документов с использованием файлового формата Open XML [9][10].

В статье Харлова И. А. приводится пример разработки модуля формирования отчётов [9]. Автор описывает создание веб приложения на основе языка C#, с использованием Open XML SDK. Предложенная в статье блок-схема взаимодействия компонентов модуля может оказаться полезной в построении архитектуры прототипа приложения. Так же автор отмечает эффективность работы с использованием пакета Open XML SDK.

В статье Миронова В. В. и Шакировой Г. Р. приводятся данные о структуре DOCX документа, а также даются комментарии касательно разбиения внутренней структуры файла, на более мелкие компоненты, для точечной работы с рассматриваемыми компонентами текста [10]. Что может быть полезно исходя из специфики проверки документов.

Поиск среди готовых программных продуктов, решающих ту же проблему, что рассматривается в работе, дал только один результат. Им стал сервис «Стандарты и нормы» платформы Системы Торговых Аналитических Решений [11]. Данный сервис позволяет произвести проверку документов на соответствие выбранному стандарту, однако является платным. В бесплатной же версии возможности воспользоваться данной услугой нет, что в свою очередь делает его практически не пригодным для большинства студентов, являющихся основной целевой аудиторией разрабатываемого продукта.

По результатам поиска и анализа существующих аналогов разрабатываемого приложения, был сделан вывод об отсутствии существующих некоммерческих решений в российском сегменте рынка разработки программного обеспечения.

1. **АНАЛИЗ СТРУКТУРЫ DOCX ДОКУМЕНТА**
   1. **Файлы входящие в состав простейшего docx документа**

Так как для работы был выбран формат DOCX как самый распространённый формат для редактирования документов, работу я начал с разбора внутренней структуры этого формата.

Расширение docx пришло на смену устаревающему формату DOC в 2006 году. Основным отличием этих двух типов файлов было то, что docx, расшифровываемый как Office Open XML Document, базировался на Extensible Markup Language (XML), что в свою очередь давало ему ряд преимуществ, таких как [12]:

* уменьшение занимаемого файлами пространства на диске;
* простота распространения;
* общедоступность и открытость формата.

Внутри docx документ представляет из себя zip-архив, в состав которого входят XML файлы и встроенные изображения. При этом состав входящих в docx документ XML файлов варьируется и расширяется в зависимости от используемых в документе функций формата Office Open XML и стилей текста. Состав файлов, входящих в простейший docx документ можно увидеть на рисунке 4.

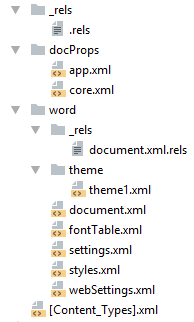


Рисунок 4 – Состав простейшего docx документа

Разберём что из себя представляет каждая из частей [13]:

* [Content\_Types].xml – XML файл, содержащийся в корне документа, хранит в себе MIME-типы содержимого документа;
* \_rels/.rels – XML файл, представляющий из себя главный список связей всего документа, в рассматриваемом простейшем docx документе приводит в соответствие идентификатор rId1 и файл word/document.xml, основное тело документа;
* docProps/app.xml – XML файл, содержащий общую информацию о документе: количество слов, страниц, символов а также информация о приложении в котором был создан файл;
* docProps/core.xml – XML файл, содержащий информацию об основных метаданных документа в соответствии с Open Packaging Conventions и Dublin Core;
* word/\_rels/document.xml.rels – XML файл, содержащий список связей в части word/document.xml;
* word/theme/theme1.xml – XML файл, описывающий тему документа: цветовую схему, шрифты, форматирование;
* word/document.xml – XML файл, содержащий основную разметку документа;
* word/fontTable.xml – XML файл, содержащий список используемых в документе шрифтов;
* word/settings.xml – XML файл, содержащий относящиеся к документу настройки;
* word/styles.xml – XML файл, содержащий информацию о стилях применимых к документу;
* word/webSettings.xml – XML файл, содержащий информацию о том как именно следует конвертировать документ в HTML формат.

В работе наиболее подробно будет рассматриваться файл document.xml так как он содержит исчерпывающую информацию о вёрстке документа и именно благодаря ему можно понять как будут отображаться в текстовом редакторе все структурные элементы документа.

* 1. **Структура файла document.xml**

Для анализа внутренней структуры основного файла разметки docx документа необходимо использовать текстовый редактор, имеющий возможность форматирования XML документов, так как по умолчанию Microsoft Word не проставляет никаких разделителей в XML элементах и весь текст внутри файла представлен одной строкой, что делает его нечитаемым для человека.

Для чтения XML файлов была использована программа Notepad++ – это свободно распространяемый, бесплатный текстовый редактор с открытым исходным кодом, поддерживающий подсветку синтаксиса и разметки. По умолчанию программа не поддерживает возможности автоматического форматирования XML текста, поэтому для реализации поставленной передо мной задачи был установлен плагин XML Tools, средствами которого и проводилось форматирование.

Государственные стандарты, диктующие нормы оформления документов, описывают конкретные правила для определённых элементов, содержащихся в тексте, таких как [14]:

* таблицы;
* заголовки;
* изображения;
* колонтитулы;
* абзацы.

Так же ГОСТ регламентирует вид самого шрифта, использующегося в документе, а именно:

* цвет;
* название;
* начертание;
* размер;

В изучении структуры файла document.xml я придерживался следующего алгоритма:

* создание DOCX документа, содержащего исследуемую часть текста;
* приведение его к виду ZIP архива;
* анализ изменений в структуре document.xml.

На основании полученных данных, были выделены следующие XML теги, на основе которых можно судить о форматировании текста и абзацев, смотри таблицу 3.

Таблица 3 – XML элементы, задающие форматирование текста и абзацев

|  |  |  |
| --- | --- | --- |
| Родительский элемент | Элемент, задающий форматирование | Назначение элемента |
| <w:rPr>  …  </w:rPr> | <w:rFonts /> | содержит информацию о названии шрифта |
| <w:b /> | содержит информацию о полужирном начертании шрифта |
| <w:i /> | содержит информацию о курсивном начертании шрифта |
| <w:sz /> | содержит информацию о размере шрифта |
| <w:color /> | содержит информацию о цвете шрифта |
| <w:pPr>  …  </w:pPr> | <w:spacing /> | содержит информацию о межстрочном интервале |
| <w:ind /> | содержит информацию об абзацном отступе |
| <w:numPr>  …  </w:numPr> | содержит информацию о том является ли абзац элементом перечисления, а также хранит параметры этого перечисления |
| <w:jc /> | содержит информацию о выравнивании текста абзаца |

Как видно из таблицы 3, все элементы, задающие форматирование абзацев имеют один и тот же родительский элемент, а именно Paragraph properties, который в свою очередь является дочерним элементом тега Paragraph. Весь текст DOCX документа поделён на абзацы, представленные в формате Open XML элементом Paragraph.

В таблице 3 так же видно, что все элементы, задающие форматирование шрифта имеют один родительский элемент ­– Run properties. Этот элемент является дочернем элементом Run. В формате Open XML Run является дочерним элементом Paragraph, и если Paragraph представляет собой абзац, то Run – это фрагмент текста.

Форматирование таблиц, подписей к ним, а также изображений и подписей к ним можно так же определить, опираясь на описанные выше XML элементы.

Параметры полей документа, а также колонтитулов можно найти в XML элементе Section properties.

1. **ПРОЕКТИРОВАНИЕ ПРОТОТИПА ПРИЛОЖЕНИЯ**
   1. **Выбор формата взаимодействия с пользователем**

На этапе проектирования приложения необходимо было определить, как оно будет взаимодействовать с пользователем. Размышляя над форматом взаимодействия с пользователями были выдвинуты следующие идеи:

1. реализовать с нуля функционал для отображения docx документов с дополнительным функционалом для указания на проблемные места в документе;
2. воспользоваться интегрированным в формат Open XML инструментом рецензирования документов, представленным комментариями к конкретным частям текста.

Говоря о плюсах первой идеи, можно отметить следующее:

* наглядность, пользователь мог бы взаимодействовать с документом и видеть свои ошибки непосредственно в интерфейсе приложения;
* возможность тонкой настройки каждого комментария, реализуя функционал комментирования собственными силами можно было бы добавить возможность для каждого комментария настраивать его собственный стиль, чтобы таким образом категорировать их, дополнительно подчёркивать важность той или иной ошибки.

Однако нельзя не отметить и очевидных минусов такого подхода:

* реализация с нуля на клиентской части приложения функционала для отображения docx документов требует куда как более глубокого изучения документации для реализации всех тонкостей функционала, предоставляемого форматом Open XML;
* реализация подобного функционала является слишком затратной по времени даже для команды разработчиков, так как объём всевозможных настроек форматирования документа, предоставляемый форматом Open XML, слишком велик;
* для успешной реализации столь крупного программного продукта следовало бы прибегнуть к групповой разработке, чтобы избежать возможных ошибок проектирования и иметь возможность разделить обязанности меду разработчиками.

Рассуждая над преимуществами второй идеи, было получено следующее:

* универсальность, любой текстовый редактор, поддерживающий работу с docx документами, реализует функционал по их рецензированию, так как он является частью формата Open XML;
* схожее направление разработки, реализация взаимодействия с пользователем, через встроенный механизм комментирования формата Open XML, сводится к работе с внутренней структурой docx документа [14];
* время разработки, в открытом доступе имелись материалы с примерами работы со встроенным механизмом рецензирования формата Open XML, что должно было существенно ускорить разработку.

Из недостатков второго подхода можно отметить:

* малую наглядность, по результатам работы приложения пользователь получит docx документ с комментариями, но у него не будет возможности изучить их непосредственно в клиенте приложения;
* отсутствие настроек для комментариев, внешний вид рецензий, оставляемых программой будет неизменным в рамках текстового редактора, которым пользователь пользуется для открытия полученного файла.

По результатам анализа преимуществ и недостатков этих двух методов был сделан вывод что в рамках работы по созданию прототипа приложения для определения соответствия документа заданному ГОСТу следует обратиться именно ко второму способу взаимодействия с пользователем, через встроенные средства комментирования формата Open XML.

* 1. **Проектирование серверной части приложения**

Проектируя серверную часть приложения, необходимо было определится с основным функционалом приложения, архитектурой приложения и базы данных, а также выбрать язык программирования и метод разработки программного обеспечения, которому можно было бы следовать процессе работы.

Размышляя над основным функционалом, был сделан вывод, что пользователю может потребоваться регистрация и авторизация в приложении с целью защиты его личных данных, которые могут содержаться в загруженных им документах. Так же было решено добавить возможность хранения и отображения ранее загруженных и ранее проверенных на корректность документов. Основным же функционалом приложения должна являться возможность проверки загруженного документа на соответствие стандарту, выбранному пользователем. В рамках работы по созданию прототипа приложения, мною было принято решение реализовать функционал проверки только на один стандарт, ГОСТ 7.32–2017, в связи с ограничениями по времени разработки. Use-case диаграмму, отражающую взаимодействие пользователя с приложением можно увидеть на рисунке 5.

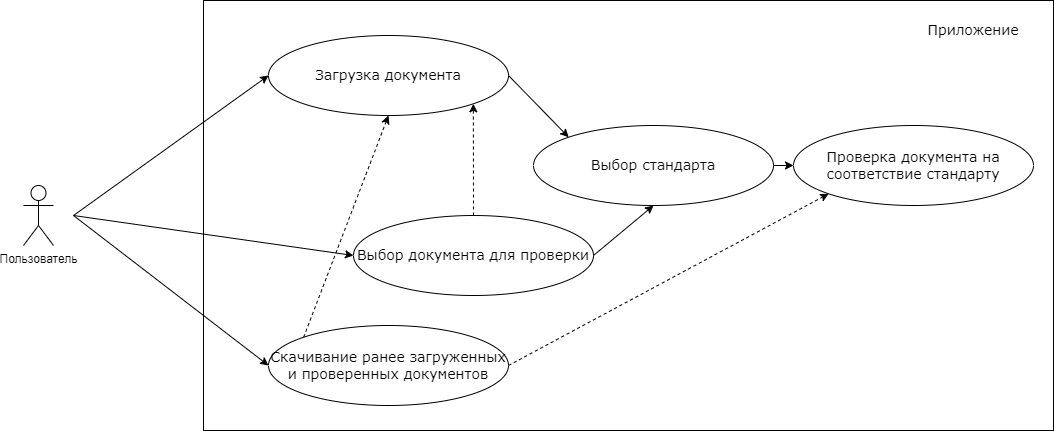


Рисунок 5 – Взаимодействие пользователя с приложением

Архитектура базы данных тесно связана с функционалом. Исходя из обозначенных ранее возможностей приложения, были спроектированы следующие таблицы:

* Users, таблица пользователя хранит его логин и пароль для авторизации;
* UploadedDocuments, таблица для хранения документов загруженных пользователем, ставит в соответствие путь к документу на сервере, его название и идентификатор пользователя, загрузившего документ;
* EditedDocuments, таблица для хранения документов уже проверенных программой, ставит в соответствие путь к обработанному документу на сервере, его название и идентификатор пользователя, которому принадлежит этот документ.

Описанных выше таблиц достаточно для реализации функционала, который предусматривает проектируемая серверная часть приложения.

Продумывая архитектуру приложения, был сделан акцент на её расширяемости, чтобы в будущем сохранялась возможность добавлять новые стандарты, расширяя тем самым функционал приложения. Поэтому было важно сделать объекты в приложении слабо связными и легко заменимыми.

Одной из основных сущностей приложения должен стать стандарт, в соответствии с которым должна проводиться проверка. Для того чтобы сохранить возможность для дальнейшего расширения приложения мною было принято решения об описании интерфейса IDocumentStandard с которым и взаимодействовали компоненты приложения, вместо конкретной реализации стандарта.

IDocumentStandard должен содержать типовые методы «FindErrorIn…» которые должны проверять конкретную часть документа, отражённую в названии метода, и в случае несоответствия данной части проверяемому стандарту возвращать ошибку.

Основной функционал ошибки заключается в возможности конвертации её в комментарий формата Office Open XML. При этом для сохранения возможности внутренней настройки ошибки в зависимости от реализации, так же было принято решение выделить отдельный интерфейс IDocumentError который требовал бы от класса его реализующего наличия метода GetComment(), который возвращал бы комментарий в формате Open XML для дальнейшего добавления его в проверяемый документ. При этом также были выделены следующие свойства характерные для любого типа ошибок вне зависимости от их реализации, а именно:

* текст ошибки, в котором содержалась бы исчерпывающая информация о причине возникновения ошибки и давались бы рекомендации по её исправлению;
* стандарт, ГОСТ в соответствии с которым была порождена ошибка, чтобы пользователь, в случае если у него возникнут дополнительные вопросы касающиеся ошибки, мог узнать в каком именно стандарте следует искать информацию;
* подпункт стандарта, для более удобной навигации пользователя в случае необходимости изучения ГОСТа.

Так как задачей класса, реализующего IDocumentStandard, является только возможность проверки частей документа на соответствие стандарту, представляемому данным классом, то необходимо также выделить сущность, которая бы делила документ на составные части, определяла бы чем именно они являются, например заголовком, таблицей или же подписью к изображению, и отправляла на проверку в соответствующие методы IDocumentStandard, а также имела бы возможность добавлять комментарии о найденных ошибках в документ.

Сущностью, которая берёт на себя эту обязанность, стал класс Editor, реализующий два основных метода EditDocument(), проводящий покомпонентную проверку документа на соответствие стандарту и сохраняющий найденные ошибки в памяти, и AddComments(), который добавлял бы Open XML комментарии, в соответствии с найденными ошибками, в проверяемый документ.

Взаимодействие описанных выше компонентов можно проследить на рисунке 6.

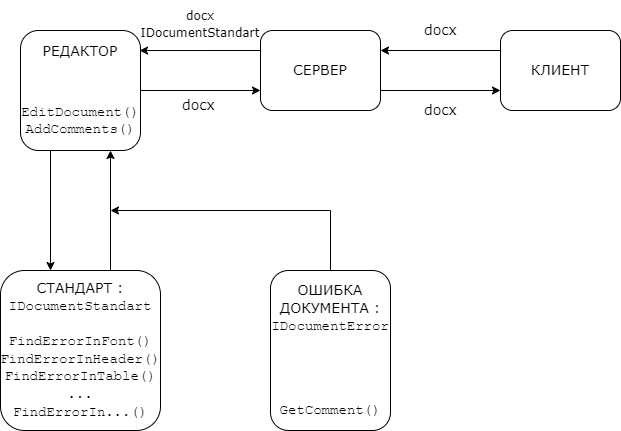


Рисунок 6 – Схема работы функции проверки

Так же для взаимодействия с базой данных следует реализовать классы, которые соответствовали сущностям, хранящимся в таблицах, они будут называться соответственно Person, UploadedDocument, EditedDocument. Набор полей этих классов должен соответствовать столбцам связанных с ними таблиц. Запросы, связанные с добавлением документов в базу загрузкой их из базы, а также регистрацией и авторизацией пользователей сервер будет обрабатывать в контроллере.

Перед началом разработки необходимо выбрать также язык программирования, на котором она будет вестись. Для разработки был необходим язык программирования, подходящий для написания серверных приложений, а также поддерживающий работу с Open XML SDK – набором инструментов от Microsoft, создателей формата Open XML, для работы DOCX, PPTX и XLSX документами. Также при выборе языка программирования следовало учитывать имеющийся опыт разработки на нём.

По всем критериям подходил язык программирования С# 10. Для серверной разработки на этом языке существует официальная, поддерживаемая Microsoft библиотека ASP .NET Core 6, которая предоставляет все необходимые инструменты для комфортной и быстрой серверной разработки [15]. Также в языке C# поддерживается и работа с Open XML SDK ­– основным инструментом для программной работы с Open XML документами [16].

В качестве среды разработки приложения была выбрана Visual Studio Community 2022, IDE от Microsoft так как и язык программирования и основные инструменты серверной разработки и работы с документами использовались от этой компании. Данная среда разработки поддерживает C# десятой версии, а также ASP .NET Core последней 6 версии, что и было необходимо для работы [17].

Подходя к выбору, метода разработки были учтены следующие особенности предстоящей работы:

* была понятна общая концепция задачи, однако тонкости реализации оставалось ещё выяснять;
* определение соответствия стандарту каждой отдельной категории частей документа слабо зависит от определения корректности оформления иных категорий, что позволяет их рассматривать по раздельности;
* получить первый результат нужно было как можно быстрее, для подведения выводов и прогнозирования потенциальных затрат по времени на реализацию каждой отдельной функции;
* не все элементы документа, форматирование которых предписывается ГОСТом, имеют одинаковую сложность в проверке на корректность оформления.

Проанализировав различные популярные модели разработки программного обеспечения, был сделан вывод что наиболее эффективной для предстоящей станет «Incremental Model». Инкрементная модель предполагает прохождение полного цикла разработки при реализации каждой функциональной части приложения, начиная от планирования и предъявления требований к ней, заканчивая её тестированием и оценкой [18].

Так же был сделан вывод что наиболее рациональным методом разработки данного приложения станет следующий алгоритм:

* выделение части документа, оформление которой предписывается ГОСТом;
* сбор информации о требованиях, предъявляемых стандартом к выбранной части документа;
* реализация необходимого функционала проверки;
* тестирование реализованного функционала на DOCX файлах, содержащих ошибку в выбранных частях документа.

Повторяя данный алгоритм для всех частей документа, описываемых гостом, можно прийти к итоговому виду приложения.

* 1. **Проектирование клиентской части приложения**

При проектировании клиентской части приложения необходимо было определиться с его дизайном и технологиями, которые бы применялись при её разработке.

Для разработки дизайна интерфейса клиентской части приложения был использован онлайн-сервис для прототипирования и разработки интерфейсов Figma. В нём были спроектированы прототипы всех окон веб-приложения таких как:

* окно регистрации;
* окно входа;
* главный экран;
* экран выбора стандарта для проверки соответствия.

Главный экран приложения вы можете видеть на рисунке 7. Дизайн окон регистрации и авторизации можно увидеть в приложении Л.

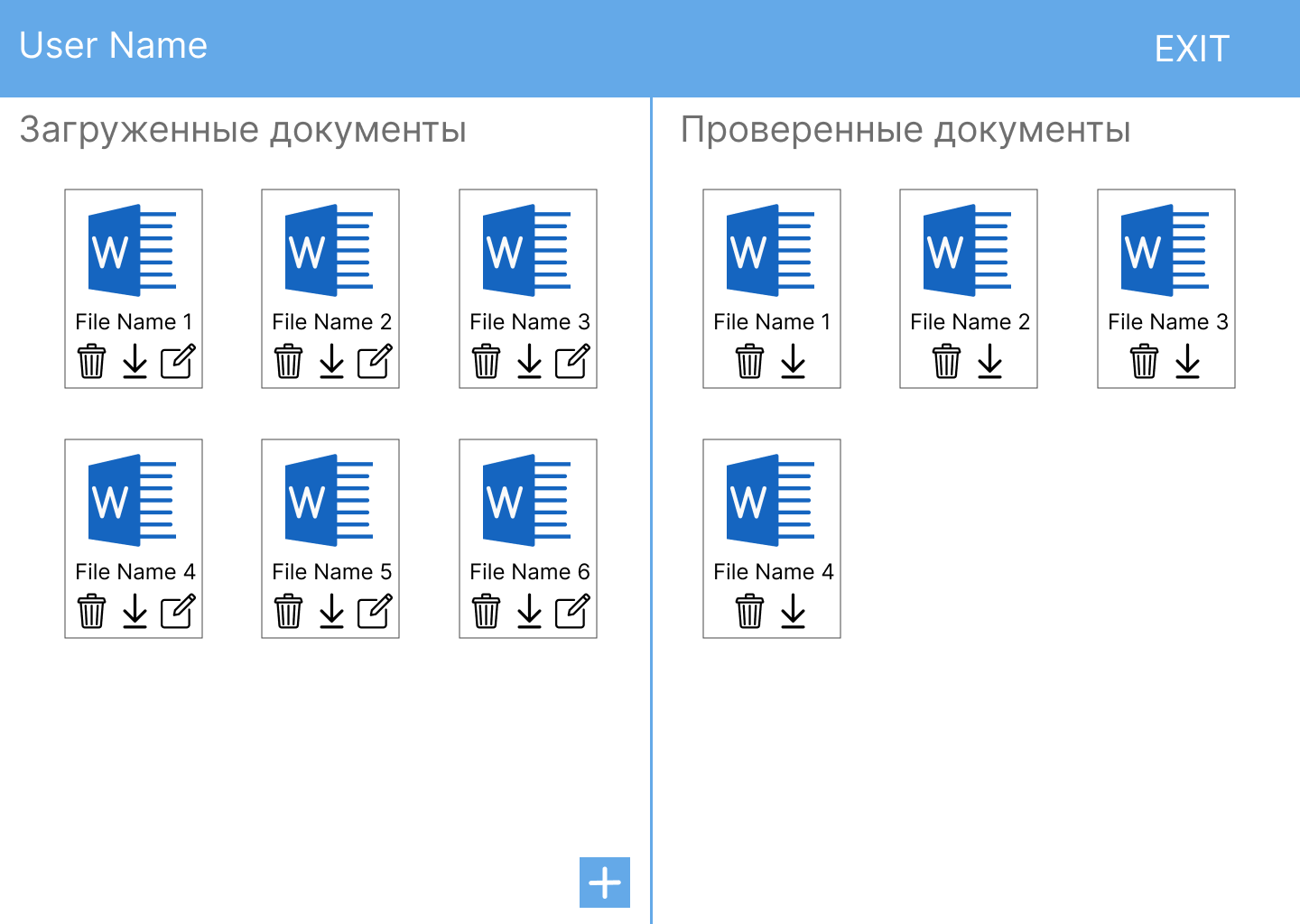


Рисунок 7 – Набросок главного экрана приложения

Так как клиентская часть приложения не реализует никакого сложного функционала было принято решение использовать при её разработке HTML, CSS и Java Script без каких-либо сторонних фреймворков, которые бы лишь замедлили разработку, а также негативно сказались бы на скорости отклика самого интерфейса.

1. **РАЗРАБОТКА ПРОТОТИПА ПРИЛОЖЕНИЯ**
   1. **Создание класса редактора**

После проектирования архитектуры приложения и описания всех используемых в нём интерфейсов наступил этап написания базовых реализаций классов. Для начала работы по выявлению ошибок в различных частях документа, необходимо было реализовать механизм, позволяющий разделять файл на эти составные части. Так началось создание класса редактора.

Следуя построенной ранее архитектуре была начата разработка метода EditDocument(), задачей которого было деление документа на составные части, их проверка на соответствие ГОСТу и сохранение найденных ошибок. Как отмечалось ранее, в главе об анализе структуры docx документа, основным структурным элементом файла document.xml является Paragraph, именно в нём хранятся все фрагменты текста. Поэтому было принято решение разделить документ по параграфам.

Open XML SDK представляет элементы Open XML документа в качестве древовидной структуры данных [19]. Исходя из этого, чтобы получить доступ к коллекции, содержащей все элементы Paragraph, содержащиеся в документе, необходимо запросить все дочерние элементы, XML тега Body и отфильтровать из них те, которые являются абзацами. После полученную коллекцию можно поэлементно проверять в цикле.

На данном этапе, после выделения абзацев, уже можно проводить частичную проверку документа, так как каждый Paragraph хранит в себе ParagraphProperties, а значит можно получить информацию о выравнивании текста на странице междустрочном интервале отступе первой строки и т. д. Однако этого недостаточно для того, чтобы проверить все настройки форматирования документа, предписываемые стандартом.

Следующим этапом разбиения документа стало выделение из Paragrah фрагментов форматированного текста, представленного XML элементом Run. Для этого также нужно получить список всех дочерних элементов каждого абзаца и отфильтровать из них те, которые представлены XML элементом Run.

Как было описано ранее в главе анализа структуры docx документа, элемент Run хранит в себе RunProperties, которые являются настройками форматирования текста. Поэтому уже на этом этапе класс редактора может проверить большинство параметров задаваемых ГОСТом. Код метода EditDocument() можно увидеть в приложении А.

Для того чтобы сохранять ошибки найденные в процессе проверки документа был инициализирован словарь, ключами которого стали параграфы, в которых были обнаружены ошибки, а значениями список, хранящий в себе IDocumentError.

После реализации функционала по обнаружению ошибок началась разработка метода AddComments(), который должен встраивать полученные из IDocumentError Open XML комментарии.

Для начала работы с open XML комментариями необходимо определить присутствуют ли они в документе на момент редактирования, так как каждый Open XML комментарий имеет свой уникальный идентификатор. В формате docx комментарии хранятся отдельно от структуры документа, в файле comments.xml, в document.xml на них оставляются только ссылки. Пример файла comments.xml вы можете увидеть на рисунке 8.

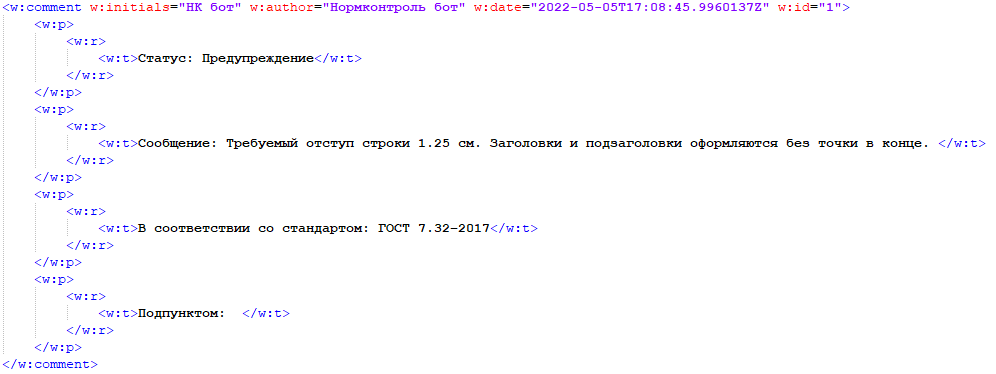


Рисунок 8 – Пример комментария из comments.xml

Метод поиска пригодного для работы идентификатора комментария выглядит следующим образом:

//Проверяется есть ли в документе комментарии

if(document.MainDocumentPart.GetPartsCountOfType<WordprocessingCommentsPart>() > 0)

{

//В случае наличия комментариев они сохраняются

comments =

document.MainDocumentPart.WordprocessingCommentsPart.Comments;

//Получается идентификатор с самым большим значением

if (comments.HasChildren)

id = comments.Descendants<Comment>()

.Select(e => e.Id.Value)

.Max();

}

else

{

//В случае отсутствия комментариев они создаются и идентификатор равен 0

WordprocessingCommentsPart commentPart = document.MainDocumentPart.AddNewPart<WordprocessingCommentsPart>();

commentPart.Comments = new Comments();

comments = commentPart.Comments;

id = 0;

}

После нахождения пригодного для использования идентификатора можно приступить к добавлению комментариев. Для этого необходимо создать новый объект класса Comment, предоставляемый Open XML SDK и являющийся представлением Open XML комментария в коде. Для этого объекта инициализировать значения идентификатора, автора комментария, инициалов автора и времени создания комментария а также добавить Paragraph содержащий текст комментария в качестве его дочернего элемента, данный функционал берёт на себя метод GetComment() интерфейса IDocumentError().

После получения объекта комментария необходимо встроить его в структуру docx документа. Для этого необходимо добавить его в качестве дочернего элемента ранее полученному объекту comments, представляющему собой comments.xml. Чтобы указать в каком именно абзаце документа была найдена комментируемая ошибка, следует создать объекты типа CommentRangeStart и CommentsRangeEnd, хранящие в себе идентификатор встраиваемого комментария, и установить их в качестве дочерних элементов параграфа, в котором были найдены ошибки. CommentsRangeStart встраивается перед первым элементом Run абзаца. CommnetsRangeEnd встраивается после последнего элемента Run абзаца. Код метода AddComments() можно увидеть в приложении Б.

После реализации данных методов класса Editor начался этап разработки класса ошибки, реализующей интерфейс IDocumentError.

* 1. **Создание класса ошибки**

Главным методом класса ошибки стал GetComment(), преобразующий её в Open XML комментарий, поэтому основная работа при разработке этого класса была связана именно с реализацией этого метода.

Ошибка хранит в себе текст сообщения, стандарт в соответствии с которым проводится проверка, пункт стандарта для упрощения навигации пользователя, а также свойства приоритета ошибки и категории ошибки.

Введение свойства приоритета связано с тем, что государственный стандарт не всегда жёстко задаёт параметры оформления тех или иных частей текста, иногда даже носит скорее рекомендательный нежели запретительный характер. Исходя из этого были выделены четыре приоритета ошибок, которые можно задать при инициализации объекта данного класса, а именно:

* error, строгая ошибка, нарушение конкретных правил оформления предоставляемых стандартом, например нарушение абзацного отступа;
* warning, предупреждение, ошибка появляющаяся в случае отступления от рекомендованных ГОСТом значений, однако предусматриваемых им рамках, например использование шрифта отличного от Times New Roman;
* info, информация, используется в случае, когда определить является ли использование какого-либо форматирования нарушением правил предписываемых ГОСТом, например использование курсивного начертания текста разрешённого только для именования объектов и написания терминов;
* notAnError, приоритет показывающий, что данный объект не является по сути своей ошибкой, необходим для корректной обработки элементов документа не содержащих в себе ошибок.

Так же были введены категории ошибок, они необходимы так как при использовании некоторых шрифтов в текстовом редакторе, структура документа сильно искажается, в результате чего каждое отдельное слово оборачивается в элемент Run, что в свою очередь в случае возникновения ошибок в форматировании текста, приводит к созданию множества однотипных комментариев, принадлежащих одному абзацу. Категории ошибок позволяют избежать этого исключая появление схожих ошибок в рамках одного элемента Paragraph.

В реализации метода GetComment() особое внимание было уделено форматированию выводимого пользователю комментария. Несмотря на то, что возможности настройки комментариев в Open XML ограничены, можно делить выводимый текст на абзацы. Была спроектирована следующая структуру комментария:

* приоритет ошибки;
* текст сообщения об ошибке;
* стандарт, в соответствии с которым проходила проверка;
* подпункт этого стандарта.

Для её создания необходимо было каждое выводимое в комментарии сообщение оборачивать в XML элементы Paragraph, Run и Text. Код метода GetComment() можно увидеть в приложении В.

После создания класса ошибки начался этап разработки класса стандарта.

* 1. **Создание класса стандарта**

В данной работе было решено использовать стандарт ГОСТ 7.32–2017 «Отчёт о научно-исследовательской работе» [20]. Этот стандарт был выбран по той причине, что для его тестирования можно было бы найти множество реальных документов, так как именно в соответствии с этим стандартом оформляются выпускные квалификационные работы, и по завершению разработки имелась бы возможность протестировать корректность работы прототипа приложения на фрагментах работ других обучающихся. Также для проверки некоторых элементов документа использовалась книга «Методические рекомендации для подготовки и оформления выпускных квалификационных работ», кафедры интеллектуальных информационных технологий Уральского федерального университета, для дополнения критериев оформления, задаваемых ГОСТом [21].

Началась работа с реализации метода по выявлению ошибок в начертании шрифтов. Были выделены следующие требования ГОСТ 7.32–2017 предъявляемые к начертанию текста [20]:

* полужирным шрифтом выделяются только заголовки разделов и подразделов;
* курсивный шрифт используется только в написании объектов и терминов.

Для определения используется ли в проверяемом фрагменте текста полужирное или курсивное начертание необходимо проверить содержит ли исследуемый элемент RunProperties среди своих дочерних элементов Bold или Italic. В случае если какое-либо из начертаний используется, то в текст ошибки добавляется информация об этом. После чего формируется объект класса ошибки, ей в конструкторе передаётся текст будущего комментария, приоритет Info, а также подпункт стандарта, в котором описаны данные требования. Код метода можно увидеть в приложении Г.

После тестирования работы описанного функционала на примере документа, содержащего данные ошибки, началась разработка следующих методов.

Для реализации метода определения корректности выбранного шрифта был рассмотрен Open XML элемент RunFonts. Он содержит в себе информацию о названии шрифта используемого для вывода фрагмента текста. ГОСТ 7.32–2017 рекомендует использовать при написании отчётов шрифт Times New Roman [20]. Поэтому необходимо проверить что в свойстве «ascii» элемента RunFonts хранится именно это значение, в ином случае необходимо отправить ошибку с приоритетом Warning. Код метода можно увидеть в приложении Г.

Для проверки размера шрифта следует проверить свойство value элемента FontSize. ГОСТ предписывает использовать шрифт не менее 12 пт [20]. Внутренними единицами измерения формата Open XML являются точки. На основании исследования структуры document.xml было выяснено, что точки соотносятся с пт в отношении два к одному, а значит искомое значение value 24. В Open XML SDK все значения, полученные из структуры документа, хранятся в строках, поэтому для сравнения необходимо перевести их в целые числа. В случае если свойство value исследуемого фрагмента текста меньше 24, следует вернуть ошибку с приоритетом Error. Код метода можно увидеть в приложении Г.

Для проверки цвета шрифта необходимо обратиться к свойству value элемента Color. ГОСТ предписывает использовать шрифт только чёрного цвета [20]. Чёрный цвет в формате Open XML задаётся как «000000», а также является цветом по умолчанию, поэтому для проверки необходимо узнать присутствует ли элемент Color в RunProperties, если нет, то значит цвет задан по умолчанию, если присутствует, то необходимо проверить чтобы значение его свойства value было рано «000000». Код метода можно увидеть в приложении Г.

Для поиска ошибок в оформлении абзацев необходимо обратиться к следующим XML элементам:

* SpacingBetweenLines, отвечает за междустрочный интервал;
* Indentation, хранит информацию об абзацных отступах.

Данные XML элементы не всегда присутствуют в ParagraphProperties, при этом их значения по умолчанию отличаются от предписываемых ГОСТом. Поэтому помимо проверки их значений на соответствие ГОСТу необходимо также проверять наличие этих элементов, их отсутствие эквивалентно ошибке. В случае если элемент SpacingBetweenLines присутствует в ParagraphProperties, необходимо проверить значение его свойства Line. ГОСТ 7.32–2017 предписывает, что междустрочный интервал должен равняться полутора пунктам, что задаётся в формате Open XML как значение Line равное 360 [20]. Если элемент Indentation присутствует необходимо проверить значение его свойства FirstLine, в котором содержится информация об абзацном отступе первой строки. Стандарт устанавливает значение отступа равным 1,25 см, что в формате Open XML выражается как значение свойства FirstLine равное 709 [20]. В случае нахождения несоответствий с ГОСТом, данным методом формируется ошибка с приоритетом Error. Код метода можно увидеть в приложении Д.

Для поиска более сложных ошибок форматирования, например в списках, таблицах и заголовка, необходимо было добавить функционал распознавания данных элементов в структуре документа классом редактора. Так в результате анализа структуры document.xml были сделаны следующие выводы:

* таблица задаётся в документе элементом Table, хранящим внутри себя элементы TableRow, представляющие строки таблицы, дочерними элементами TableRow являются XML тэги TableCell, описывающими ячейки таблицы, и уже в них в свою очередь содержатся обычные элементы Paragraph;
* перечисление в документе можно определить по наличию в свойствах абзаца элемента NumberingProperties;
* изображение задаётся элементом Drawing;
* информация о полях документа храниться в XML элементе SectionProperties;
* заголовки и подзаголовки можно охарактеризовать как элементы перечисления, выделенные полужирным начертанием, поэтому для их поиска в документе нужно обратиться к элементам NumeringProperties и Bold.

На основании этих данных в класс редактора был добавлен функционал по определению вышеописанных частей документа для корректной передачи их в класс стандарта.

Для поиска ошибок в форматировании полей следует обратиться к элементу SectionProperties. В его дочернем элементе PageMargin находятся свойства left, right, top, bottom, в которых хранятся значения левого, правого, нижнего и верхнего полей соответственно. ГОСТ 7.32–2017 предписывает использовать для полей следующие значения [20]:

* левое – 3 см;
* правое 1,5 см;
* нижнее – 2 см;
* верхнее – 2см.

В единицах измерения формата Open XML данные размеры преобразуются в 1700 для левого поля, 850 для правого, 1134 для верхнего и нижнего. Необходимо проверить равны ли значения свойств элемента PageMargin данным значениям и в если нет вернуть ошибку с приоритетом Error. Код метода можно увидеть в приложении Е.

В процессе тестирования данного метода на примере реального документа выяснилось, что данные значения в формате Open XML не всегда равны для одних и тех же полей. Правое поле равное 1,5 см заданное в текстовом редакторе Microsoft Word по умолчанию преобразуется в 850 точек, однако если задать те же значения полей вручную и сохранить документ, его значение изменится на 851. Было выдвинуто предположение, что данная погрешность связана с тем, что поле задаётся не в целых числах. Для компенсации данного факта было решено сравнивать поля документа с припуском в пять точек.

В методе поиска ошибок в перечислениях необходимо обратиться к элементу Text абзаца, распознанного как элемент перечисления. ГОСТ предписывает чтобы каждый элемент простого перечисления оканчивался на запятую, а каждый элемент сложного (многоуровневого) перечисления оканчивался на точку с запятой [20]. Поэтому была произведена проверка на то, чтобы свойство Text элемента Text оканчивалось на данные символы. В ином случае отправляется ошибка с приоритетом Info. Код метода можно увидеть в приложении Ж.

Для поиска ошибок в оформлении изображений необходимо было обратиться к абзацам предшествующему изображению и следующим за пояснением к иллюстрации, так как они не должны были содержать текста. Для проверки корректности оформления подписи к иллюстрации происходит обращение к элементу ParagraphProperties абзаца следующего за элементом Drawing. Стандарт предписывает что подпись к иллюстрации должна быть расположена в следующем за ней абзаце, по центру документа и не должна иметь точки в конце [20]. Для определения положения абзаца на странице следует обратиться к элементу Justification и его свойству value принимающему следующие значения:

* both, текст располагается по ширине страницы;
* left, текст выравнивается по левому краю;
* right, текст выравнивается по правому краю;
* center, выравнивание текста происходит по центру страницы.

Необходимо было проверить чтобы значение было равно center. Для проверки оканчивается ли подпись к иллюстрации на точку нужно обратиться к элементу Text содержащемуся в элементе Paragraph подписи. Требование стандарта по использовании в подписи к иллюстрации шрифта равного 14 пт можно проверить через элемент FontSize, его свойство value должно равняться 28. Если несоответствия были найдены, то метод возвращает ошибку с приоритетом Error. Код метода можно увидеть в приложении И.

При проверке оформления таблиц необходимо обратить внимание не только на содержимое таблицы, но и на оформление её подписи, поэтому мною было принято решение разделить эти две проверки на разные методы.

Проверяя на соответствие стандарту содержимое таблицы, особое внимание следовало обратить на её структуру. Стандарт требует писать название столбцов и строк с прописной буквы и без точек в конце [20]. Для проверки того, что проверяемая ячейка является заголовком столбца следует обратиться к её родительскому элементу TableRow и убедится, что ему не предшествуют другие строки таблицы. Для определения того является ли ячейка заголовком строки необходимо проверить наличие предшествующих ей элементов TableCell в рамках рассматриваемой строки. В случае если были найдены несоответствия со стандартом метод отправляет ошибку с приоритетом Error. Код метода можно увидеть в приложении К.

Название таблицы, согласно стандарту, должно отвечать следующим требованиям [20]:

* располагаться сверху слева над таблицей, в предшествующем ей абзаце;
* не должно оканчиваться на точку;
* не должно иметь абзацного отступа.

Для проверки того располагается ли название слева, необходимо проверить свойство Value элемента Justification абзаца названия. Оно должно иметь значение left. Чтобы узнать отсутствует ли абзацный отступ у названия таблицы необходимо обратиться к элементу Indentation и, если он присутствует, проверить что его свойство Value имеет значение 0. Проверка на то оканчивается ли название таблицы точкой производится путём обращение к элементу Text проверяемого абзаца и проверки его последнего символа. В случае если были найдены несоответствия метод возвращает ошибку с приоритетом Warning. Код метода можно увидеть в приложении К.

При проверке корректности оформления заголовков и подзаголовков необходимо проверить следующее [20][21]:

* что они выравнены по левому краю;
* выделены полужирным начертанием;
* не оканчиваются точкой;
* что им предшествует пустая строка
* что за ними следует пустая строка;
* что междустрочный интервал равен 1;
* что они написаны с абзацного отступа.

Для проверки выравнивания необходимо убедиться, что значение свойства Value элемента Justification равно left. Для проверки абзацного отступа необходимо убедиться, что значение свойства FirstLine элемента Indentation равно 709. Для проверки полужирного начертания необходимо выяснить является ли элемент Bold дочерним элементом проверяемого абзаца. Междустрочный интервал проверяется через элемент SpacingBetweenLines и его свойство Line значение которого должно равняться 240. Проверка наличия пустых строк, предшествующих заголовку и следующих за ним производится путём проверки наличия текста в соответствующих абзацах. Проверка на то, что заголовок не оканчивается точкой производится путём проверки, что свойство Text элемента Text не оканчивается точкой. В случае нахождения несоответствий метод возвращает ошибку с приоритетом Error. Код метода можно увидеть в приложении Л.

* 1. **Разработка серверных контроллеров**

Обязанность по обработке запросов с клиента на сервере лежит на классах контроллерах. Так как в разработке серверной части приложения был использован фреймворк ASP .NET Core 6, то контроллер представляет из себя C# класс, которому при поступлении запроса на сервер система маршрутизации передаёт данные запроса, после чего контроллер проводит над ними необходимые операции и отсылает результат своей работы обратно [22].

В используемой серверной архитектуре было выделено два основных класса контроллера:

* AccountController, занимающийся обработкой запросов связанных с учётной записью пользователя;
* DocumentController, занимающийся обработкой запросов связанных с работой с документами.

В AccountController были реализованы методы для регистрации нового пользователя, авторизации пользователя, а также метод для удаления учётной записи пользователя.

Так как приложению нет необходимость хранить о пользователе какие-либо дополнительные данные при регистрации используется только логин и пароль. Они сохраняются в базу данных, чтобы в дальнейшем можно было авторизовать уже зарегистрированного пользователя.

При разработке контроллеров я использовал библиотеку MediatR, позволяющую отделить логику обработки запроса клиента, от логики приёма запроса и отправки ответа. В используемой архитектуре контроллер отвечает только за приём запроса, переправку его параметров соответствующему классу обработчику и возврат полученного от обработчика ответа.

Для корректной работы класса обработчика, реализующего интерфейс IRequestHandler<TRequest, TResponse> необходимо создать класс, представляющий запрос, который должен реализовать интерфейс IRequest<TResult>. В классе запроса для регистрации нового пользователя достаточно определить два свойства Login и Password, оба из которых являются строками.

Классу обработчику в конструкторе передаётся контекст данных, унаследованный от стандартного класса библиотеки EntityFramework – DbContext, для получения возможности работы с данными в базе. После чего обработчик добавляет в базу данные нового пользователя, полученные из класса запроса, и сохраняет изменения. В случае если данные были добавлены успешно сервер вернёт ответ с кодом 200, если же выполнить добавление нового пользователя не удастся будет возвращён ответ с кодом 400.

Авторизация пользователей на сервере производится при помощи JSON Web Tokens. При отправке запроса по авторизации на сервер, будет проведена проверка существует ли такой пользователь в базе данных, и в том случае если он существует, будет сгенерирован специальный ключ, который будет передан клиенту [23]. В последствии клиент должен будет сохранить этот ключ и отправлять его в HTTP заголовке Authorization. Так сервер будет распознавать авторизированных пользователей.

В классе запроса для метода авторизации также необходимо реализовать два свойства Login и Password по которым будет проходить поиск пользователя в базе данных.

Также в этом случае появилась необходимость в создании класса ответа, в котором необходимо реализовать следующие свойства:

* Id, для передачи идентификатора пользователя на клиент, необходимо для работы с документами;
* Login, для отображения логина пользователя на клиенте в соответствии с набросками сайта;
* Token, непосредственно сам ключ для авторизации пользователя.

Обработчик запроса принимает данные о пользователе, проверяет их наличие в базе данных, после чего генерирует ключ инициализирует объект класса ответа передаёт ему данные об идентификаторе, логине и ключе, и отправляет соответствующему методу контроллера. Контроллер в свою очередь отправляет ответ в формате JSON на клиент.

В классе DocumentController реализован весь необходимый функционал для работы с документами, а именно:

* добавление нового документа в базу данных;
* получение всех документов, загруженных пользователем;
* удаление загруженного документа из базы данных;
* проверка загруженного документа на соответствие стандарту;
* получение всех проверенных документов пользователя;
* удаление проверенного документа из базы данных.

Так как в спроектированной ранее архитектуре документы хранятся непосредственно в файловой системе сервера сохранение документов, полученных с клиента, производится в директорию upload.

Для реализации класса обработчика загрузки документов необходимо реализовать класс запроса, который будет содержать свойство Form с типом данных IFormCollection, которое представляет из себя данные с формы загрузки файла.

В конструктор обработчика запроса передаётся контекст данных, для работы с базой. Метод извлекает коллекцию файлов из формы, переданной в запросе, после чего поэлементно сохраняет файлы в директорию upload и добавляет запись с именем файла, путём до него на сервере и идентификатором пользователя, загрузившего файл в базу данных. В случае успешного добавления данных в базу отправляется HTTP ответ с кодом 200, в случае если данные добавить не удалось с кодом 400.

Класс запроса на получение документов пользователя содержит только одно свойство Id с идентификатором пользователя. Обработчик запроса получает информацию обо всех документах, загруженных пользователем, после чего передаёт полученную коллекцию документов на клиент.

Класс запроса на удаление документа из базы данных содержит одно свойство Id, представляющее идентификатор документа для удаления. Класс обработчика запроса на удаление получает доступ к базе данных, к таблице с документами, удаляет запись с соответствующим идентификатором и сохраняет изменения. В случае если удалось удалить запись из базы сервер вернёт HTTP ответ с кодом 200, в ином случае с кодом 400.

Класс запроса на проверку документа соответствующему стандарту содержит три свойства Id, целое число, представляющее идентификатор пользователя, DocumentId, целое число, представляющее идентификатор документа для проверки, и Standard, строка, представляющая название стандарта, в соответствии с которым будет проводиться проверка.

Класс обработчика запроса на проверку стандарта инициализирует объект класса стандарта, в соответствии с которым будет проводиться проверка, инициализирует объект класса редактора, передав ему в конструкторе стандарт и копию выбранного для обработки документа, сохранённую в директорию edited. После чего вызывает методы редактора EditDocument() и AddComments(). Добавляет в базу, в таблицу с проверенными документами, запись о названии проверенного документа, путь к нему на сервере и идентификатор пользователя, которому принадлежит документ. В случае если удалось добавить запись в базу сервер вернёт HTTP ответ с кодом 200, в ином случае с кодом 400.

Класс запроса на получение проверенных документов пользователя содержит свойство Id, целое число, представляющее идентификатор пользователя. Обработчик запроса получает информацию обо всех проверенных документах, принадлежащих пользователю, после чего передаёт полученную коллекцию документов на клиент.

Класс запроса на удаление проверенного документа из базы данных содержит Id, целое число, представляющее идентификатор проверенного документа для удаления. Класс обработчика запроса на удаление получает доступ к базе данных, к таблице с проверенными документами, удаляет запись с соответствующим идентификатором и сохраняет изменения. В случае если удалось удалить запись из базы сервер вернёт HTTP ответ с кодом 200, в ином случае с кодом 400.

В результате работы были созданы два API контроллера, позволяющие обмениваться запросами с клиентом. После их реализации начался этап разработки клиента приложения.

* 1. **Разработка клиента приложения**

Разработка клиентской части приложения заключалась в вёрстке HTML страниц, настройке стилей для них, а также в настройке отправки запросов и корректного отображения полученных результатов посредством Java Script.

Были созданы три HTML страницы:

* страница регистрации пользователя;
* страница авторизации пользователя;
* главную страница приложения;

При входе на сайт пользователь попадает на главную страницу приложения. Для работы ему необходимо авторизоваться на сайте. При нажатии на кнопку регистрации происходит перенаправление на страницу регистрации. На ней пользователь вводит логин и пароль, который он в дальнейшем будет использовать для входа на сайт. После регистрации пользователя перенаправляет на страницу авторизации, где он после ввода корректного логина и пароля перенаправляется обратно на главную страницу сайта и может начинать работу.

Все страницы были созданы в соответствии с разработанными на этапе проектирования макетами.

При написании Java Script кода было необходимо реализовать механизм отправки запросов на сервер. Запросы на сервер отправляются при помощи метода fetch (url, [options]) [24]. Он принимает в себя в качестве параметров url – адрес, на который будет отправлен запрос, а также [options] – параметры запроса в формате JSON, содержащие в себе HTTP метод, заголовки и тело запроса. Fetch по результатам своей работы возвращает ответ сервера, который можно преобразовать в JSON формат для дальнейшей работы с ним.

Для реализации регистрации пользователя необходимо считывать данные из тегов input в которые пользователь вводил свои логин и пароль, формировать на их основе данные формы, и отправлять на сервер с запросом на регистрацию, в качестве метода использовался POST, в теле запроса передавались сформированные данные формы. В случае получения от сервера кода 200 производил перенаправление на страницу авторизации пользователя.

Для реализации регистрации пользователя также считываются данные из тегов input, на их основе формируются данные формы и отправляются на сервер. В качестве параметров fetch указывается метод POST, и тело запроса – данные формы. Получив ответ от сервера, программа сохраняет ключ для авторизации в веб-хранилище sessionStorge, после чего производит перенаправление на главную страницу сайта.

При формировании содержимого главной страницы через Java Script подаются запросы на сервер на получение загруженных пользователем документов и обработанных документов пользователя. Для отправки запросов используется метод fetch без параметров, идентификатор пользователя передаётся непосредственно в url. После получения ответа от сервера поэлементно для каждого документа формирует свой блок для отображения на странице.

На главной странице при нажатии на кнопку загрузки нового документа происходит открытие дополнительного окна, в котором пользователь выбирает документ, который он хочет загрузить на сервер и подтверждает загрузку. Метод загрузки файла на сервер при помощи fetch создаёт запрос на сервер, в параметрах запроса передаётся метод POST и тело запроса – загружаемый файл.

Для удаления документов Java Script получает данные об идентификаторе документа, который нужно удалить, и методом fetch отправляет DELETE запрос на сервер. Идентификатор документа передаётся в url.

После реализации клиентской части начался этап тестирования общей работоспособности приложения.

* 1. **Тестирование приложения**

Тестирование сайта проводилось при непосредственной работе с ним. Все элементы отображались корректно, никаких наложений или смещений элементов замечено не было. Все перенаправления работали в соответствии с описанным ранее поведением. Загрузка и скачивание документов также работали корректно без каких-либо проблем.

По результатам тестирования функционала редактирования документов так же не было замечено никаких отклонений или ошибок в работе обозначенного ранее функционала редактора. Пример отредактированного документа вы можете видеть на рисунке 9.

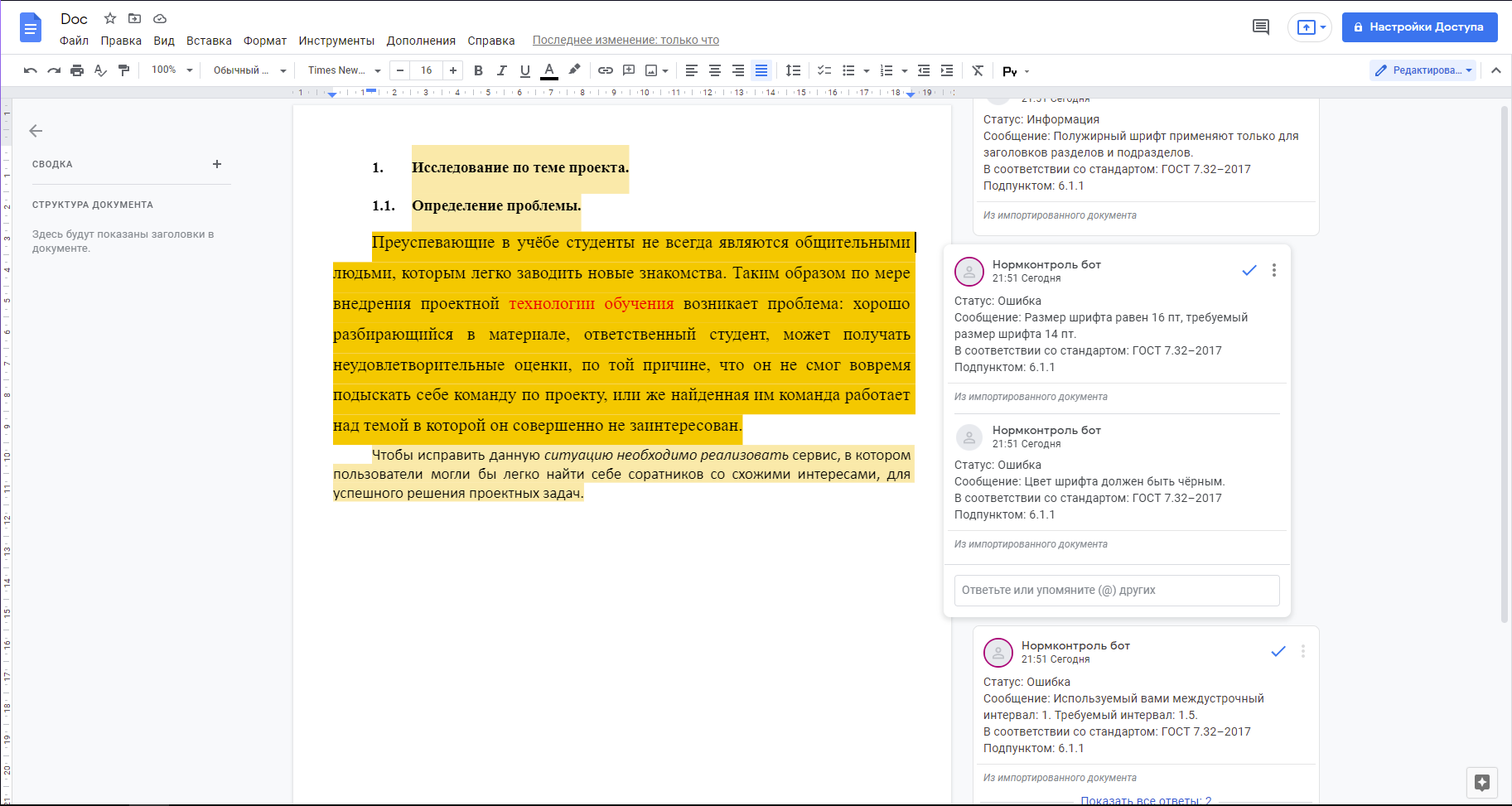


Рисунок 9 – Результат работы программы

Для тестирования сайта был использован браузер Opera GX, для просмотра отредактированных документов использовалась веб-программа Google Документы.

**ЗАКЛЮЧЕНИЕ**

В ходе работы были проанализированы разные актуальные на данный момент форматы хранения электронных документов, были сделаны выводы касаемо популярности каждого из них, также была произведена оценка тенденции развития их популярности.

Был проведён краткий обзор формата Open XML как основного в данной работе, а также пакета Open XML SDK для работы с ним.

Были проанализированы научные статьи, так или иначе затрагивающие рассматриваемую тему, также были рассмотрены существующие на рынке потенциальные конкуренты. По итогам анализа можно отметить, что данная тема слабо освящена на текущий момент времени.

Целью данной работы была разработка приложения для определения соответствия документа заданному ГОСТу. Для этого были осуществлены следующие этапы работы:

1. Произведён анализ структуры docx документов, а также структуры основного файла разметки DOCX документов document.xml. На основании полученных данных были выделены основные структурные элементы docx документов, а также числовые значения свойств соответствующие предписываемым стандартом правилам форматирования документов. Получено представление о внутренней структуре DOCX документов.
2. Была спроектирована архитектура приложения и базы данных, а также выбран метод разработки программного обеспечения и созданы макеты клиентской части приложения. Были описаны сущности базы данных, основные сущности программы, необходимые для реализации интерфейсы и классы.
3. В соответствии с проектом был создан прототип приложения. Был реализован функционал редактирования документов в соответствии со стандартом ГОСТ 7.32–2017, функционал авторизации пользователей, хранения и работы с загруженными и обработанными документами. Весь заявленный функционал реализован успешно.
4. Проведено тестирование работоспособности приложения. Протестированы все заявленный функции сайта и редактора документов. Приложение прошло все тесты.

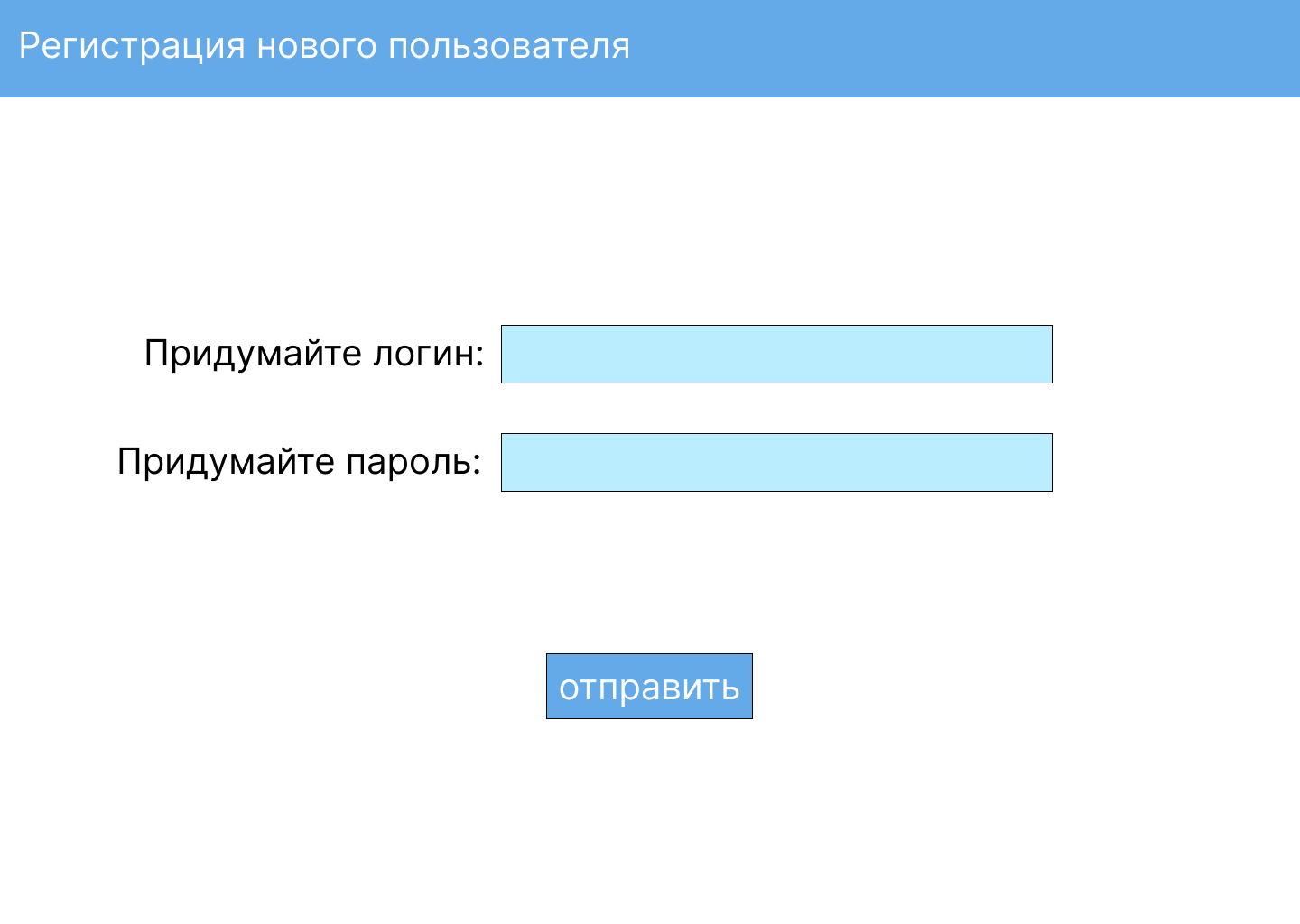
По итогам работы был получен готовый к использованию прототип веб приложения для проверки документов на соответствие ГОСТ 7.32–2017, с возможностью хранения ранее загруженных и проверенных файлов.

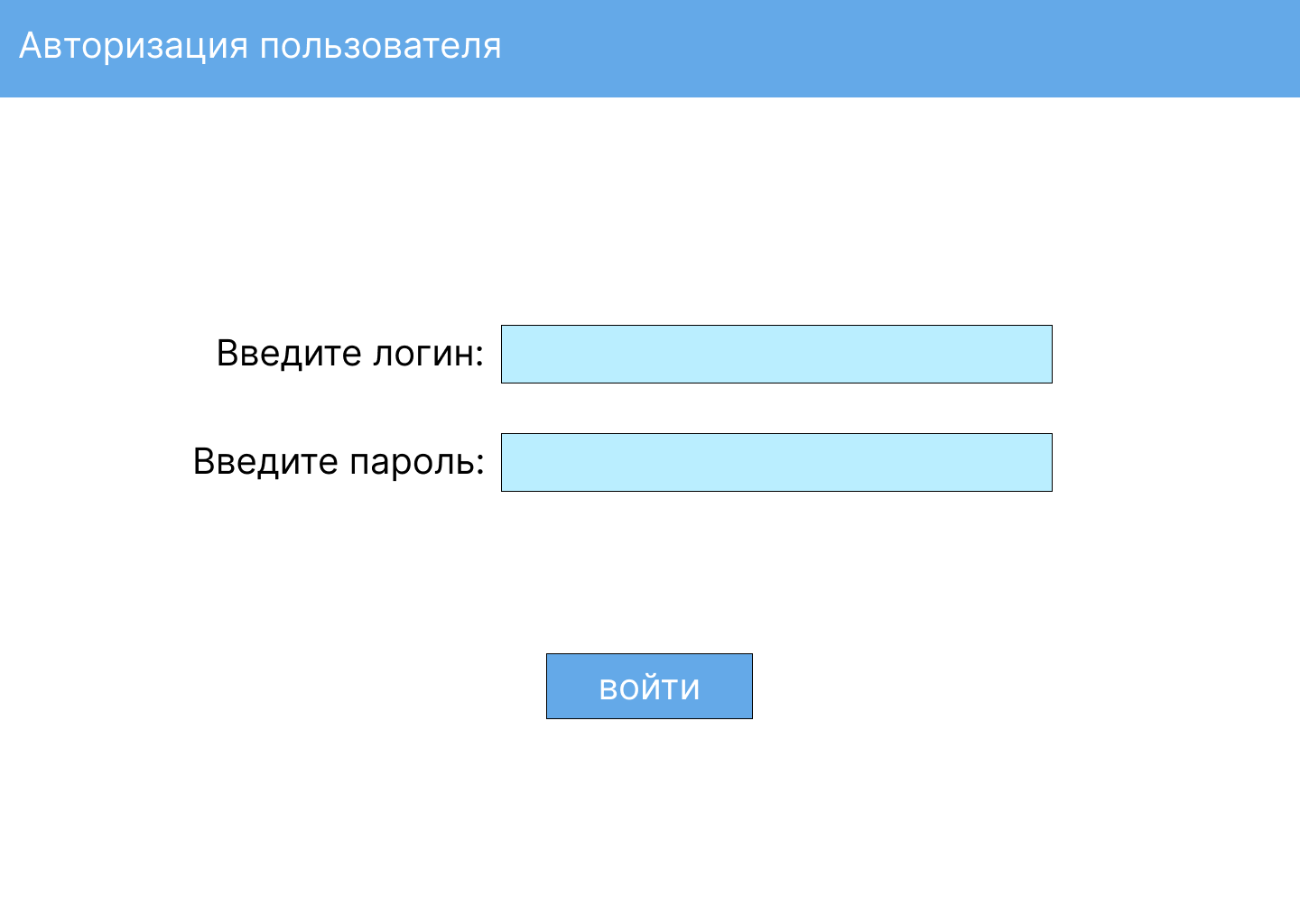
**БИБЛИОГРАФИЧЕСКИЙ СПИСОК**

1. Разница между PDF и DOC [Электронный ресурс]: https://ru.spot-the-difference.info/difference-between-pdf – (дата обращения 08.05.2022).
2. Петрянин Д. Л., Сидорова Е. А., Зорькин А. В. Сжатие текстовых данных [Текст] / Д. Л. Петрянин, Е. А. Сидорова, А. В. Зорькин // Новые информационные технологии в автоматизированных системах. – 2014. – № 17. – С. 505-507.
3. Format race 2018: DOCX over DOC and ODT? [Электронный ресурс]: https://www.onlyoffice.com/blog/2018/09/format-race-2018-docx-over-doc-and-odt/ – (дата обращения 08.05.2022).
4. Microsoft's Open Specification Promise: No Assurance for GPL [Электронный ресурс]: https://softwarefreedom.org/resources/2008/osp-gpl.html – (дата обращения 08.05.2022).
5. ISO/IEC 29500-1:2016. Information technology — Document description and processing languages — Office Open XML File Formats [Текст]: дата введения 26.10.2016. – М.: Switzerland: International Organization for Standardization, 2016. – 5024 с.
6. Broersma M. Microsoft releases first Open XML SDK [Электронный ресурс]: SDK 1.0 now available for developers to create, access and manipulate Open XML documents. – https://www.arnnet.com.au/article/224702/microsoft\_releases\_first\_open\_xml\_sdk/ – (дата обращения 08.05.2022).
7. Introduction to markup compatibility (Open XML SDK) [Электронный ресурс]: https://docs.microsoft.com/en-us/office/open-xml/introduction-to-markup-compatibility – (дата обращения 08.05.2022).
8. Научная электронная библиотека «КиберЛенинка» [Электронный ресурс]: https://cyberleninka.ru – (дата обращения 10.05.2022).
9. Харлов И. А. РАЗРАБОТКА МОДУЛЯ АТОМАТИЧЕСКОГО ФОРМИРОВАНИЯ ОТЧЁТОВ ДЛЯ АВТОМАТИЗИРОВАННОЙ ИНФОРМАЦИННОЙ СИСТЕМЫ «АБИТУРИЕНТ» ХГУ ИМ. Н. Ф. КАТАНОВА [Текст] / И. А. Харлов // Вестник Хакасского государственного университета им. Н. Ф. Катанова. – 2019. – № 30. – С. 25-28.
10. Миронов В. В., Шакирова Г. Р. ТЕХНОЛОГИЯ ПЕРСОНАЛИЗАЦИИ ДОКУМЕНТОВ ФОРМАТА OPEN OFFICE XML НА ОСНОВЕ XSL-ТРАНСФОРМАЦИИ [Текст] / В. В. Миронов, Г. Р. Шакирова // Вестник Уфимского государственного авиационного технического университета. – 2009. – № 2 том 12 – С. 188-197.
11. Система Торговых Аналитических Решений [Электронный ресурс]: https://star-pro.ru – (дата обращения 10.05.2022).
12. Difference Between DOC and DOCX: Which Should You Use? [Электронный ресурс]: https://www.folderit.com/blog/difference-between-doc-and-docx-which-should-you-use/ – (дата обращения: 13.05.2022).
13. ECMA-376. Fundamentals And Markup Language Reference [Текст]: дата введения декабрь 2006 г. – М.: Rue du Rhône 114: Ecma International, 2016. – 5029 с.
14. ГОСТ Р 7.0.97–2016. Национальный стандарт Российской Федерации. Система стандартов по информации, библиотечному и издательскому делу. Организационно-распорядительная документация. Требования к оформлению документов [Текст]: дата введения 01.07.2018. – М.: Госстандарт России: Стандартинформ, 2018. – 32 с.
15. Новые возможности в ASP.NET Core 6.0 [Электронный ресурс]: https://docs.microsoft.com/ru-ru/aspnet/core/release-notes/aspnetcore-6.0 – (дата обращения 13.05.2022).
16. Welcome to the Open XML SDK 2.5 for Office [Электронный ресурс]: https://docs.microsoft.com/en-gb/office/open-xml/open-xml-sdk – (дата обращения 13.05.2022).
17. What's new in Visual Studio 2022 [Электронный ресурс]: https://docs.microsoft.com/en-us/visualstudio/ide/whats-new-visual-studio-2022 – (дата обращения 13.05.2022).
18. Ещё раз про семь основных методологий разработки [Электронный ресурс]: https://habr.com/ru/company/edison/blog/269789/ – (дата обращения 13.05.2022).
19. OFFICE OPEN XML OVERVIEW [Электронный ресурс]: https://www.ecma-international.org/wp-content/uploads/Office-Open-XML-White-Paper.pdf – (дата обращения 13.05.2022).
20. ГОСТ 7.32–2017. Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления [Текст]: дата введения 01.07.2018 – М.: Госстандарт России: Стандартинформ, 2018. – 33 с.
21. Алферьева Т. И., Васина В. Н., Шадрин Д. Б. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПОДГОТОВКИ И ОФОРМЛЕНИЯ ВЫПУСКНЫХ КВАЛИФИКАЦИОННЫХ РАБОТ [Электронный ресурс]: Подготовлено кафедрой интеллектуальных информационных технологий / Т. И. Алферьева, В. Н. Васина, Д. Б. Шадрин. – Электронный текстовый ресурс. – Екатеринбург, 2019. – 1 электрон. текст. док. – Загл. с экрана.
22. Контроллеры [Электронный ресурс]: https://metanit.com/sharp/aspnet5/5.1.php – (дата обращения 13.05.2022).
23. Авторизация с помощью JWT-токенов [Электронный ресурс]: https://metanit.com/sharp/aspnet5/23.7.php – (дата обращения 13.05.2022).
24. Fetch [Электронный ресурс]: https://learn.javascript.ru/fetch – (дата обращения 13.05.2022).

**ПРИЛОЖЕНИЕ А**

**Дизайн окон регистрации и авторизации**





**ПРИЛОЖЕНИЕ Б**

**Код метода EditDocument()**

public void EditDocument()

{

//Инициализируем список ошибок

var errs = new List<IDocumentError>();

var skipParagraphCount = 0;

//Инициализируем объект свойств раздела

var sectionProperties = document.MainDocumentPart.Document.Descendants<SectionProperties>().First();

//Проверяем корректность оформления полей

CheckBorders(sectionProperties, errs);

//Итерируемся по абзацам

foreach (var paragraph in document.MainDocumentPart.Document.Descendants<Paragraph>())

{

//Пропускаем абзацы при необходимости

if (skipParagraphCount > 0)

{

skipParagraphCount--;

continue;

}

//Пропускаем пустые абзацы

if (paragraph.ParagraphProperties == null || !paragraph.Descendants<Run>().Any()) continue;

//Проверяем корректность иллюстраций

if (CheckImage(paragraph, errs))

{

SaveErrs(paragraph, errs);

skipParagraphCount++;

сontinue;

}

//Проверяем корректность оформления названий таблиц

if (CheckTableName(paragraph, errs))

{

SaveErrs(paragraph, errs);

continue;

}

//Проверяем корректность оформления таблиц

if (!CheckTableCell(paragraph, errs))

CheckParagraph(paragraph, errs);

//Итерируемся по фрагментам текста

foreach (var run in paragraph.Descendants<Run>())

{

//Пропускаем пустые фрагменты текста

if (run.RunProperties == null) continue;

//Проверяем заголовки

if (CheckHeader(paragraph, run, errs)) continue;

//Проверяем корректность оформления шрифтов

CheckFontFamily(run, errs);

CheckFontSize(paragraph, run, errs);

CheckFontColor(run, errs);

CheckFontStyle(run, errs);

//Проверяем перечисления

CheckList(paragraph, run, errs);

}

//Сохраняем найденные ошибки

SaveErrs(paragraph, errs);

}

}

**ПРИЛОЖЕНИЕ В**

**Код метода AddComments()**

public void AddComments()

{

//Итерируемся по абзацам с обшибками

foreach (var paragraph in errors.Keys)

{

/Итерируемся по списку ошибок

foreach (var error in errors[paragraph].DistinctBy(x => x.Category))

{

//Преобразуем ошибку в комментарий

var cmt = error.GetComment();

//Добавляем комментарий в документ

cmt.Id = currentCommentId;

comments.AppendChild(cmt);

comments.Save();

paragraph.InsertBefore(new CommentRangeStart()

{ Id = cmt.Id }, paragraph.GetFirstChild<Run>());

var cmtEnd = paragraph.InsertAfter(new CommentRangeEnd()

{ Id = cmt.Id }, paragraph.Elements<Run>().Last());

paragraph.InsertAfter(new Run(new CommentReference() { Id = cmt.Id }), cmtEnd);

IncreaseCurrentId();

}

}

}

**ПРИЛОЖЕНИЕ Г**

**Код метода GetComment()**

public Comment GetComment()

{

//Создаём текст сообщения

var firstP= new Paragraph(new Run(new Text("Статус: " + GetErrorStatus())));

var secondP = new Paragraph(new Run(new Text("Сообщение: " + Message)));

var thirdP = new Paragraph(new Run(new Text("В соответствии со стандартом: " + Standart)));

var fourthP = new Paragraph(new Run(new Text("Подпунктом: " + StandartItem)));

//Создаём комментарий

var comment = new Comment()

{

Author = "Нормконтроль бот",

Initials = "НК бот",

Date = DateTime.UtcNow.AddHours(5)

};

//Добавляем текст в комментарий

comment.AppendChild(firstP);

comment.AppendChild(secondP);

comment.AppendChild(thirdP);

comment.AppendChild(fourthP);

//возвращаем комментарий

return comment;

}

**ПРИЛОЖЕНИЕ Д**

**Код методов поиска ошибок в шрифтах**

public IDocumentError FindErrorInFontColor(Color color)

{

//Проверяем цвет шрифта

if (color.Val != "000000")

return new ErrorS7\_32\_2017("Цвет шрифта должен быть чёрным.", "6.1.1", Prioritys.Error, Categories.FontColor);

return NotAnError.GetError();

}

public IDocumentError FindErrorInFontName(RunFonts font)

{

//Проверяем название шрифта

if (font.Ascii != "Times New Roman")

return new ErrorS7\_32\_2017(

$"Вы используете шрифт {font.Ascii}, рекомендуемый тип шрифта основного текста Times New Roman. " +

"Использование иного начертания шрифта может применяться для выделения текста.",

"6.1.1",

Prioritys.Warning,

Categories.FontName);

return NotAnError.GetError();

}

public IDocumentError FindErrorInFontSize(FontSize fontSize)

{

int fontSizeInPt;

int.TryParse(fontSize.Val, out fontSizeInPt);

fontSizeInPt /= 2;

//Проверяем размер шрифта

if (fontSize.Val != "28")

return new ErrorS7\_32\_2017(

$"Размер шрифта равен {fontSizeInPt} пт, требуемый размер шрифта 14 пт.",

"6.1.1",

Prioritys.Error,

Categories.FontSize);

return NotAnError.GetError();

}

public IDocumentError FindErrorInFontStyle(RunProperties runProperties)

{

var errorMessage = string.Empty;

//Проверяем полужирное начертание

if (runProperties.Descendants<Bold>().Any())

errorMessage += "Полужирный шрифт применяют только для заголовков разделов и подразделов. ";

//Проверяем курсивное начертание

if (runProperties.Descendants<Italic>().Any())

errorMessage += "Использование курсива допускается для обозначения объектов и написания терминов.";

if (errorMessage != string.Empty)

return new ErrorS7\_32\_2017(errorMessage, "6.1.1", Prioritys.Info, Categories.FontStyle);

return NotAnError.GetError();

}

**ПРИЛОЖЕНИЕ Е**

**Код метода проверки абзацев**

public IDocumentError FindErrorInParagraph(ParagraphProperties paragraphProperties)

{

var errorMessage = string.Empty;

//Проверяем, что заданы настройки междустрочного интервала

if (paragraphProperties.SpacingBetweenLines?.Line != null && paragraphProperties.SpacingBetweenLines?.LineRule != null)

{

double lineSpDouble = int.Parse(paragraphProperties.SpacingBetweenLines.Line) / 240.0;

//Проверяем междустрочный интервал

if (paragraphProperties.SpacingBetweenLines.Line != "360")

errorMessage += $"Используемый вами междустрочный интервал: {lineSpDouble}. Требуемый интервал: 1.5. ";

}

else errorMessage += $"Требуемый междустрочный интервал: 1.5. ";

//Проверяем что заданы настройки отступа первой строки

if (paragraphProperties.Indentation?.FirstLine != null)

{

double firstLineIndDouble = int.Parse(paragraphProperties.Indentation.FirstLine) / 567.2;

//Проверяем отступ первой строки

if (paragraphProperties.Indentation.FirstLine != "709")

errorMessage += $"Используемый вами отступ первой строки: {firstLineIndDouble} см. Требуемый отступ: 1.25 см.";

}

else errorMessage += $"Требуемый отступ первой строки: 1.25 см.";

if (errorMessage != string.Empty)

return new ErrorS7\_32\_2017(errorMessage, "6.1.1", Prioritys.Error, Categories.Paragraph);

return NotAnError.GetError();

}

**ПРИЛОЖЕНИЕ Ж**

**Код метода проверки полей**

public IDocumentError FindErrorInBorders(PageMargin borders)

{

var leftBorder = int.Parse(borders.Left);

var rightBorder = int.Parse(borders.Right);

var topBorder = int.Parse(borders.Top);

var bottomBorder = int.Parse(borders.Bottom);

//Проверяем поля с припуском

if (Math.Abs(leftBorder - 1701) > BorderAllowance ||

Math.Abs(rightBorder - 850) > BorderAllowance ||

Math.Abs(topBorder - 1134) > BorderAllowance ||

Math.Abs(bottomBorder - 1134) > BorderAllowance)

return new ErrorS7\_32\_2017(

$"Вы используете поля: левое – {leftBorder / 56} мм, правое – {rightBorder / 56} мм, " +

$"верхнее – {topBorder / 56} мм, нижнее – {bottomBorder / 56} мм. " +

"Следует соблюдать следующие размеры полей: левое – 30 мм, правое – 15 мм, верхнее и нижнее – 20 мм",

"6.1.1",

Prioritys.Error,

Categories.Border);

return NotAnError.GetError();

}

**ПРИЛОЖЕНИЕ И**

**Код метода проверки перечислений**

public IDocumentError FindErrorInList(ParagraphProperties paragraphProperties, Text text)

{

var errorMessage = "Перед каждым элементом перечисления должно быть тире или строчная буква русского алфавита" +

", если требуется указать ссылку на пункт перечисления. ";

//Проверяем что элемент оканчивается на «,» или «.»

if (!text.Text.Trim().EndsWith(',') || !text.Text.Trim().EndsWith(';'))

errorMessage += "Каждый элемент перечисления, кроме последнего, должен оканчиваться на \",\" или \";\" если это сложное перечисление.";

if (errorMessage != string.Empty)

return new ErrorS7\_32\_2017(errorMessage, "6.4.6", Prioritys.Info, Categories.Lists);

return NotAnError.GetError();

}

**ПРИЛОЖЕНИЕ К**

**Код метода проверки иллюстраций**

public IDocumentError FindErrorInImage(Paragraph paragraph)

{

var errorMessage = string.Empty;

var prevParagraph = paragraph.ElementsBefore().LastOrDefault(x => x is Paragraph);

//Проверка на предшествующую пустую строку

if (prevParagraph != null)

{

var prevParagraphRun = prevParagraph.Descendants<Run>().LastOrDefault();

if (prevParagraphRun != null)

{

var prevParagraphText = prevParagraphRun.Descendants<Text>().LastOrDefault();

if (prevParagraphText != null)

errorMessage += "Перед иллюстрацией следует оставлять одну пустую строку. ";

}

}

//Проверка корректности оформления названия

var imageDesc = paragraph.ElementsAfter().FirstOrDefault(x => x is Paragraph);

if (imageDesc != null)

{

var descParagraph = (Paragraph)imageDesc;

var descJc = descParagraph.ParagraphProperties?.Justification;

if (descJc != null && descJc.Val != "center")

errorMessage += "Название иллюстрации следует размещать по центру. ";

var descRun = descParagraph.Descendants<Run>().LastOrDefault();

if (descRun != null)

{

var descRunProp = descRun.RunProperties;

if (descRunProp != null && descRunProp.FontSize.Val != "28")

errorMessage += "Размер шрифта в названии иллюстрации должен быть равен 14 пунктам. ";

var descText = descRun.Descendants<Text>().LastOrDefault();

if (descText != null && descText.Text.EndsWith('.'))

errorMessage += "Точку в конце названия рисунка ставить не следует. ";

}

}

var nextParagraph = imageDesc.ElementsAfter().FirstOrDefault(x => x is Paragraph);

//Проверка на последующую пустую строку

if (nextParagraph != null)

{

var nextParagrapRun = nextParagraph.Descendants<Run>().FirstOrDefault();

if (nextParagrapRun != null)

{

var nextParagraphText = nextParagrapRun.Descendants<Text>().FirstOrDefault();

if (nextParagraphText != null)

errorMessage += "После названия иллюстрации следует оставлять одну пустую строку. ";

}

**Продолжение ПРИЛОЖЕНИЯ К**

}

if (errorMessage != string.Empty)

{

//Рекомендация по оформлению названия в случае, если были ошибки

errorMessage += "После иллюстрации должно располагаться её название и номер в формате \"Рисунок 1 - Название рисунка\". ";

return new ErrorS7\_32\_2017(errorMessage, " ", Prioritys.Error, Categories.Image);

}

return NotAnError.GetError();

}

**ПРИЛОЖЕНИЕ Л**

**Код методов проверки таблиц и подписей к ним**

public IDocumentError FindErrorInTable(Paragraph paragraph)

{

var errorMessage = string.Empty;

var tc = paragraph.Parent;

var tr = tc?.Parent;

var tbl = tr?.Parent;

var firstRun = paragraph.Elements<Run>().FirstOrDefault();

var firstText = firstRun?.Elements<Text>().FirstOrDefault();

var lastRun = paragraph.Elements<Run>().LastOrDefault();

var lastText = lastRun?.Elements<Text>().LastOrDefault();

if (tr != null && !tr.ElementsBefore().Any(x => x is TableRow))

{

//Проверка названий столбцов

if (firstText != null && char.IsLower(firstText.Text.First()))

errorMessage += "Названия столбцов должны начинаться с цифры или прописной буквы. ";

if (lastText != null && lastText.Text.EndsWith('.'))

errorMessage += "В конце названия столбцов не следует ставить точку. ";

}

if (tc != null && !tc.ElementsBefore().Any(x => x is TableCell))

{

//Проверка названий строк

if (firstText != null && char.IsLower(firstText.Text.First()))

errorMessage += "Названия строк должны начинаться с цифры или прописной буквы. ";

if (lastText != null && lastText.Text.EndsWith('.'))

errorMessage += "В конце названия строк не следует ставить точку. ";

}

if (errorMessage != string.Empty)

return new ErrorS7\_32\_2017(errorMessage, " ", Prioritys.Warning, Categories.Table);

return NotAnError.GetError();

}

public IDocumentError FindErrorInTableName(Paragraph paragraph)

{

var errorMessage = string.Empty;

if (paragraph != null)

{

var paragraphProp = paragraph.ParagraphProperties;

//Проверка выравнивания названия

if (paragraphProp != null && paragraphProp.Justification != null && paragraphProp.Justification?.Val != "left")

errorMessage += "Название таблицы следует располагать над таблицей слева. ";

//Проверка абзацного отступа названия

if (paragraphProp != null && paragraphProp.Indentation != null && paragraphProp.Indentation.FirstLine != "0")

errorMessage += "Название таблицы пишется без абзацного отступа. ";

//Проверка на точку в конце названия

var paragraphRun = paragraph.Descendants<Run>().LastOrDefault();

if (paragraphRun != null)

**Продолжение ПРИЛОЖЕНИЯ Л**

{

var paragraphText = paragraphRun.Descendants<Text>().LastOrDefault();

if (paragraphText != null && paragraphText.Text.EndsWith('.'))

errorMessage += "В конце названия таблицы точка не ставится. ";

}

}

if (errorMessage != string.Empty)

return new ErrorS7\_32\_2017(errorMessage, " ", Prioritys.Warning, Categories.Table);

return NotAnError.GetError();

}