

# ОСОБЕННОСТИ АВТОМАТИЗАЦИИ НОРМОКОНТРОЛЯ ТЕКСТОВЫХ ДОКУМЕНТОВ

О. В. Манкевич, П. А. Семеняк

Кафедра информационных систем и технологий, Белорусский государственный технологический университет

Минск, Республика Беларусь

E-mail: lucky\_1994@tut.by, smalpik@ya.ru

В данной работе авторы обращают внимание на проблему процесса нормоконтроля текстовых документов, а также рассматривают возможный вариант автоматизации этого процесса.

## ВВЕДЕНИЕ

Техническая документация должна соответствовать требованиям, предъявляемым к ней в соответствии с различными стандартами. Процесс осуществления нормоконтроля довольно трудоемок и занимает значительное количество времени, в связи с чем создание средства автоматизации этого процесса является весьма актуальной задачей. Можно выделить следующие требования к такому программному средству:

- возможность работы с текстовыми документами в наиболее популярных форматах;
- возможность определения пользователем правил, в соответствии с которыми будет осуществляться нормоконтроль;
- возможность хранения наборов различных правил;
- разграничение прав доступа пользователей к документам и правилам.

## I. ОБЗОР ФОРМАТА OFFICE OPEN XML

Office Open XML – это открытый стандарт для текстовых документов, презентаций и электронных таблиц, который может быть реализован в разных приложениях на различных платформах. Цель стандарта – снять зависимость документов от приложений Microsoft Office, в которых они были созданы, чтобы с ними можно было работать в других приложениях независимо от формата и без потери данных.

Файлы Office Open XML хранятся в ZIP-архиве. Структура архива представлена на рисунке 1.

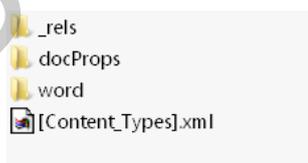


Рис. 1 – Структура документа

Структурно документ Open XML – это пакет, содержащий следующий набор элементов:

- набор явных связей между частями пакета («\_rels»);
- основные свойства документа («docProps»);

- описание главной части документа, его содержимого, стилей, тем и прочего («word»);
- описание типа содержимого всех частей документа («Content\_Types.xml»).

Текстовые документы описываются с помощью разметки WordprocessingML. Поскольку XML – это обычный структурированный текст, просмотреть содержимое части документа возможно с помощью средства чтения текста.

В таблице 1 перечисляются некоторые основные части пакета WordprocessingML, а также соответствующий каждой части одноименный класс, который предоставляет доступ к элементу через программный интерфейс пакета SDK 2.5 Open XML.

Таблица 1 – Основные части пакета WordprocessingML

Часть пакета	Элемент (tag)	Класс пакета SDK 2.5 Open XML
Основной документ document.xml	<document>	Document
Параметры документа settings.xml	<settings>	Settings
Комментарии comments.xml	<comments>	Comments
Концевые сноски endnotes.xml	<endnotes>	Endnotes
Нижний колонтитул footer.xml	<ftr>	Footer
Сноски footnotes.xml	<footnotes>	Footnotes
Верхний колонтитул header.xml	<hdr>	Header
Определения стилей styles.xml	<styles>	Styles

Стоит отметить, что главным файлом является «document.xml», поскольку именно он содержит текст, а также ссылки на остальные элементы документа, такие как стили, комментарии и прочее (рис. 2).

```

<?xml version="1.0" encoding="utf-8"?>
<w:document>
  <w:body>
    <w:p>
      <w:r>
        <w:t>Hello, Word!</w:t>
      </w:r>
    </w:p>
  </w:body>
</w:document>

```

Рис. 2 – Структура файла «document.xml»

Базовая структура «document.xml» состоит из элементов «<document>» и «<body>». Внутри располагаются один или несколько элементов «<p>», которые соответствуют абзацам. Абзац содержит один или несколько элементов «<r>». Пробегом («<r>»), называют область текста с общим набором свойств, таких как форматирование. Пробег может содержать в себе сколько угодно элементов «<t>». Элемент «<t>» содержит текст.

## II. ОСОБЕННОСТИ РЕАЛИЗАЦИИ

Для разработки приложения была выбрана клиент-серверная архитектура. Серверная часть реализуется на основе платформы .NET Framework и технологии ASP.NET. Платформа .NET Framework содержит обширный набор классов, которые упрощают и ускоряют разработку приложений. Технология ASP.NET является частью платформы .NET Framework и предоставляет возможность разрабатывать веб-приложения и веб-сервисы. Клиент-серверная архитектура была выбрана по следующим причинам:

- удобство распространения приложения. Веб-приложение не нужно скачивать и устанавливать;
- простота обновления. В случае добавления нового функционала или исправления ошибки достаточно лишь обновить серверную часть и все пользователи сразу же получают доступ к обновленной версии приложения;
- контроль доступа. Каждый пользователь имеет доступ только к своим документам и наборам правил, в соответствии с которыми осуществляется нормоконтроль.

Для работы с документами Office Open XML используется SDK OpenXML. Пакет SDK предоставляет строго типизированные классы для управления документами, соответствующими спецификации форматов файлов Office Open XML. Пакет SDK 2.5 Open XML позволяет выполнять множество стандартных задач, необходимых при работе с пакетами Open XML, поэтому сложные операции можно осуществлять всего несколькими строками кода.

Однако, этот SDK не лишен недостатков. В частности, к разным частям документа часто необходимо применить разные правила оформления. К примеру, оформление титульного листа или списка литературы будет отличаться от оформления главной части документа. Таким образом, встает вопрос о программном разграничении частей документа, т. к. в Open XML нет даже понятия «Страница».

## ЗАКЛЮЧЕНИЕ

Многие задачи поддаются полной или частичной автоматизации. Задача осуществления нормоконтроля не является исключением. Автоматизация позволяет существенно сэкономить время и человеческие ресурсы, что в итоге приводит к экономии денежных средств, даже с учетом затрат на разработку программного обеспечения.

При проектировании средства автоматизации следует внимательно отнестись к выбору архитектуры проекта и технологий. Клиент-серверная архитектура может упростить развертывание и обслуживание приложения, однако она не лишена минусов: не имея доступа к сети, пользователь не сможет получить доступ к приложению. От выбора технологий зависит скорость и трудоемкость реализации проекта, а также его масштабируемость в дальнейшем.

1. Open XML SDK 2.5 Office [Электронный ресурс] / MSDN Microsoft. – Microsoft, 2015. – Режим доступа: <https://msdn.microsoft.com/ru-ru/library/office/bb456487.aspx>. – Дата доступа: 20.04.2015.
2. Работа с документами WordprocessingML (Open XML SDK [Электронный ресурс] / MSDN Microsoft. – Microsoft, 2015. – Режим доступа: <https://msdn.microsoft.com/ru-ru/library/office/gg278327.aspx>. – Дата доступа: 21.04.2015.