

ИНТЕРАКТИВНОЕ ИНТЕЛЛЕКТУАЛЬНОЕ ПРИЛОЖЕНИЕ ДЛЯ ОФОРМЛЕНИЯ ПЕЧАТНЫХ РАБОТ В СООТВЕТСТВИИ С ГОСУДАРСТВЕННЫМИ СТАНДАРТАМИ

Артемяева И.Л.¹, Трофимов М.В.²

¹ФГБОУ ВПО «Дальневосточный федеральный университет», Владивосток, Россия (690091, г. Владивосток, ул. Суханова, 8, ауд. 138), e-mail: iartemeva@mail.ru.

²ФГБОУ ВПО «Владивостокский государственный университет экономики и сервиса», Владивосток, Россия (690014, г. Владивосток, ул. Гоголя, 41, ауд. 1518), e-mail: bugzex@ya.ru.

В статье обсуждается проблема оформления печатных работ, которые должны соответствовать государственным стандартам. Показано, что проблема заключается в трудности изучения и следования этим стандартам, а также в ограниченности систем подготовки текстов (таких как MS Word, Open Office Writer / Libre Office Writer, TeX и LaTeX), которые обеспечивают поддержку решения только общих задач верстки различных документов. В качестве средства решения проблемы рассматривается задача разработки программной системы, обладающей возможностями оформления печатных работ в соответствии со стандартами. В статье предлагается модель такой программной системы, приводятся сведения о средствах разработки, позволяющих создать программную систему в виде приложения для MS Word, а также описывается прототип такого приложения.

Ключевые слова: программные системы для проверки документов на соответствие стандартам, текстовые процессоры, обработка документа, программные системы, основанные на онтологиях.

INTERACTIVE INTELLIGENT APPLICATION FOR DOCUMENTS DEVELOPMENT IN ACCORDANCE WITH STATE STANDARDS

Artemieva I.L.¹, Trofimov M.V.²

¹Far Eastern Federal University, Vladivostok, Russia (8 Suhanova str., 690950, Vladivostok, office 138), e-mail: iartemeva@mail.ru.

²Vladivostok State University of Economics and Services, Vladivostok, Russia (41 Gogolya str., 690014, Vladivostok, office 1518), e-mail: bugzex@ya.ru.

In the article a problem of document development in accordance with state standards is discussed. It is shown that problems are compliance with standards during document creating, the difficulty of standards studying, and also limitation of word processors (such as MS Word, Open Office Writer / Libre Office Writer, TeX and LaTeX), which support only the common tasks to develop documents but do not support documents checking in accordance with state standards. As the problems solution a special software system is proposed. A structure of this software system is considered. The article also describes a prototype of the system.

Keywords: software systems for documents checking in accordance with state standards, word processors, document processing, software systems based on ontologies.

Введение

В соответствии с госстандартами оформление печатных работ, особенно студенческих (дипломная, курсовая, лабораторная работы), должно соответствовать целому ряду требований. Поскольку эти требования многообразны, то нередко оформление печатной работы выливается в чрезвычайно трудоёмкий процесс. Стандарты по оформлению печатных работ со временем меняются и включают в себе трудность изучения формальных требований. Текстовые процессоры обновляются и предлагают более удобные и функциональные инструменты для верстки документа, но не стремятся решать задачи верстки в соответствии с требованиями ГОСТ. Ввиду этого оформление печатных работ

представляет собой проблему для участников процесса. Поэтому работа над темой является актуальной, а в рамках проблемы ставится задача разработки дополнительного ПО для MS Word, упрощающего процесс оформления печатных работ для их соответствия государственным стандартам. Целью работы является проектирование инструмента, функциональность которого позволяла бы приводить документ в соответствие государственным стандартам в полуавтоматическом или автоматическом режиме.

Обзор наиболее распространенных средств создания документов

В обзоре приводятся сведения по системам MS Word, OpenOffice Writer и TeX. При этом рассматривается, насколько функциональность текстовых процессоров поддерживает создание документа, соответствующего некоторому государственному стандарту.

Основное преимущество MS Word и OpenOffice Writer – простота использования, а преимущество сложного текстового процессора TeX заключается в гибкости настройки форматирования документа. Программы опираются на одну и ту же концепцию верстки документа – использования наборов стилей и шаблонов для форматирования. Программы MS Word и OpenOffice Writer являются аналогичными редакторами, построенными по принципу WYSIWYG. TeX, в отличие от них, является низкоуровневой системой компьютерной верстки, но WYSIWYG-редакторы на основе TeX также существуют. Одним из таких редакторов является LyX [8]. Основное отличие TeX от MS Word и OpenOffice Writer заключается в необходимости использовать внутренний язык для разметки наполнения, а все файлы стилей и шаблонов TeX являются открытыми для изменения и имеют текстовый формат.

Верстка документа

Верстка документа в рассматриваемых системах осуществляется на основе многократного использования стилей и шаблонов, ускоряя, таким образом, процесс форматирования документа. Шаблоны задают структуру элемента наполнения, а стиль - вид этого наполнения. Наборы стилей и шаблонов могут храниться как в документе, так и в отдельном файле для повторного использования. Стили и шаблоны играют важную роль в форматировании наполнения документа и его приведении к государственному стандарту. Подготовкой к наполнению и верстке документа является создание новых стилей и шаблонов на основе встроенных, их настройка в соответствии со спецификацией стандарта. Несмотря на развитость текстовых процессоров, процесс верстки невозможно автоматизировать полностью, и поэтому пользователю приходится прибегать к структуризации и стилизации наполнения без или с частичным использованием встроенных в программу наборов стилей и шаблонов.

Создание списка литературы

Процесс создания списка литературы состоит из двух этапов: формирование и форматирование. Этап формирования библиографического списка заключается в автоматизированной генерации наполнения с учетом правил библиографического стандарта. После формирования списка следует этап форматирования, который заключается в придании внешнего стилистического вида библиографическому списку.

Формирование списка литературы в MS Word

В текстовый процессор MS Word 2007-2010 входит инструмент по работе со списком литературы. Данные для формирования списка литературы берутся из базы данных источников, которую пользователь наполняет самостоятельно или загружает ранее созданную. Библиографические ссылки формируются только по изначально заданным правилам, которые представляют собой стиль (в данном контексте играет роль шаблона) для формирования библиографического списка. В программе такие стили объединяются в набор. Стиль формирования библиографического списка для российского стандарта ГОСТ 7.0.5-2008 в наборе отсутствует. Возможность создания и редактирования существующих библиографических стилей, изменение (добавление и удаление) набора этих стилей инструментом не предлагается. Описание стилей хранится в специализированном формате XML, которое возможно исправить под действующий стандарт с помощью текстового редактора. Такую работу может выполнить только программист. Благодаря открытому формату описания библиографического списка в MS Word, одним из разработчиков был разработан стиль для стандарта ГОСТ 7.0.5-2008 и выложен в свободный доступ [7].

Таким образом, текстовый процессор MS Word предлагает пользователю удобный инструмент для работы с библиографическим списком, но не обладает функциональностью для определения собственных стилей и изменения набора стилей, несмотря на ее необходимость ввиду меняющихся стандартов. Изменение правил стиля под новые требования стандарта доступно только программисту со знаниями технологий XML.

Формирование списка литературы в OpenOffice Writer

Текстовый процессор OpenOffice Writer 3 включает в себя наиболее гибкий по сравнению с MS Word инструмент для работы с библиографическими источниками, который обладает функциональностью для создания базы данных источников и определения стиля формирования (в данном контексте такой стиль является шаблоном) библиографического списка. Программа не имеет возможности создания набора стилей. Предподготовленный стиль формирования библиографического списка для российского стандарта ГОСТ 7.0.5-2008 также отсутствует. Поэтому пользователю необходимо самостоятельно задать стиль,

используя предоставленный инструмент. При этом составленный пользователем стиль не может быть доступен в других документах, так как его описание хранится в документе.

Таким образом, текстовый процессор OpenOffice Writer 3, в отличие от MS Word, обладает инструментом с необходимой функциональностью для определения стиля формирования библиографического списка для любого библиографического стандарта. Настройка инструмента не требует от пользователя знаний программирования, но хорошее знание пользователем стандарта является обязательным.

Формирование списка литературы в TeX

Для автоматического формирования списка литературы используется программа BibTeX. Программа BibTeX позволяет создавать и использовать в дальнейшем библиографическую базу. Список источников формируется на основе этой базы и заранее описанных правил формирования – стилей. Стиль BibTeX для формирования библиографического списка по ГОСТ 7.0.5-2008 был разработан сообществом разработчиков и имеется в свободном доступе [1].

Таким образом, текстовый процессор TeX наряду с программой BibTeX позволяет формировать библиографический список на основе подготовленной базы и стиля. Для описания своих собственных стилей пользователю необходимы навыки программирования на встроенном языке.

Оформление подписей к рисункам, таблицам и формулам

При оформлении таблиц в соответствии с ГОСТ 2.105-95 с помощью программ MS Word и OpenOffice Writer в основном возникает трудность в создании надписей к таблицам и ссылок на них. Во-первых, трудность заключается в дополнительном редактировании ссылки, сформированной по шаблону, так как применяемый шаблон относится как для надписи, так и для ссылки. Такая трудность также присутствует при оформлении рисунков и формул. Во-вторых, программы не позволяют автоматически разбивать большие таблицы на страницы и создавать надписи вида «Продолжение таблицы 1». В текстовом процессоре TeX таких трудностей не возникает, потому что встроенный язык позволяет задавать более сложные правила форматирования для формирования надписей и ссылок.

Формирование рамок для конструкторских частей документа

В рассматриваемых системах не существует инструмента для автоматического формирования рамок для конструкторских частей документа в соответствии с ЕСКД. Готовые шаблоны таких рамок, отвечающих и частично отвечающих новым версиям стандартов, все же имеются в сети Интернет. Рамки для конструкторских частей документа создаются с помощью сложных таблиц. Зачастую рамки сложно редактировать, не обладая навыками соответствующей верстки. Так как рамка в рассматриваемых программах не

является отдельной сущностью, то такая полезная функциональность, как автоматическая нумерация листов графического материала и подстановка данных автора в поля данных таблицы, отсутствует.

Реализация отсутствующей функциональности на встроенном языке

Для нереализованной функциональности в рассматриваемых текстовых процессорах предусмотрен встроенный язык. В MS Word для написания программ предусмотрен язык VBA, для OpenOffice Writer предусмотрена поддержка нескольких языков – OpenOffice Basic, JavaScript, BeanShell, Python, а для TeX - собственный одноименный язык. Встроенный язык чаще всего используется для автоматизации ручной работы при необходимости внести множество однотипных изменений в наполнение документа. Для реализации отсутствующей функциональности необходимы хорошие навыки программирования, и поэтому такую работу может выполнить только программист.

Для описанных в статье проблем, связанных с тем, что существующие системы подготовки текстов не стремятся предоставить инструменты для оформления документа в соответствии с действующими стандартами, существуют решения на встроенных языках, называемых макросами. Немалое их количество имеется в свободном доступе [3]. Имеющиеся решения задач не являются комплексными, практичными в использовании и фрагментарны. Несмотря на существующие программы, объединяющие в себе наборы макросов [4], фактически не учитываются спецификации государственных стандартов, а сами программы являются сложными в настройке и использовании.

Выводы

Стандарты по оформлению печатных работ со временем меняются, что влечет за собой трудность их изучения и следования им. Системы подготовки текстов обновляются и предлагают более удобные и функциональные инструменты для верстки документа, но лишь обеспечивают поддержку решения общих задач верстки различных документов. Отсутствуют необходимые или в полной мере реализованные инструменты, которые бы обеспечивали приведение наполнения документа в соответствие государственным стандартам.

Ввиду этого оформление печатных работ в соответствии с требованиями некоторого стандарта является проблемой. Поэтому разработка программной системы, упрощающей процесс оформления печатных работ и обеспечивающей проверку их соответствия стандартам, является актуальной задачей. Целью данной статьи является описание одного из подходов к созданию такой программной системы.

Концепция программной системы

Программный анализ документа, обнаружение и исправление несоответствий госстандартам – нетривиальные задачи. Для решения задач необходимо формализовать требования стандартов и наполнение документа, используя некоторую модель описания. Кроме того, должна быть формализована задача определения несоответствия документа стандарту и задача исправления этих несоответствий. На рисунке 1 изображена концептуальная модель программной системы, предназначенной для решения указанной проблемы.

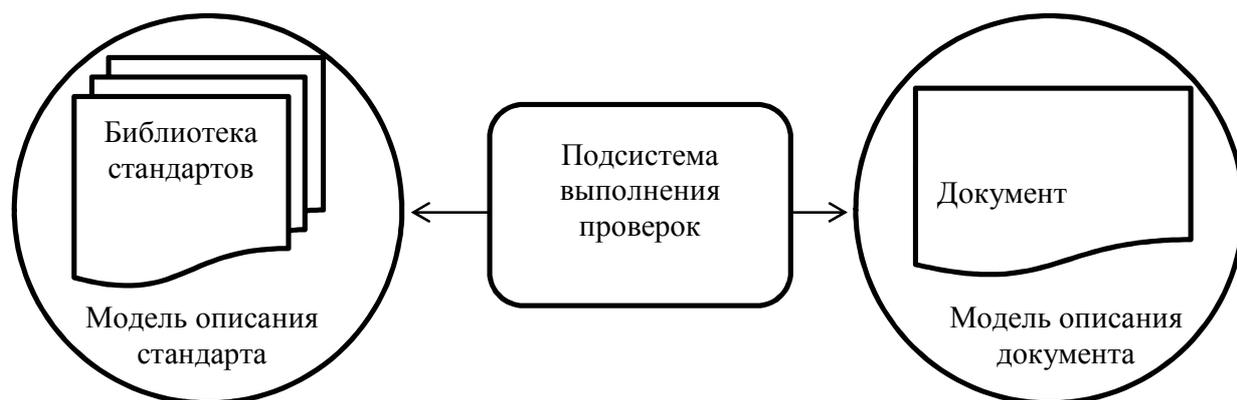


Рис. 1. Концептуальная модель программной системы

Программная система должна содержать библиотеку описаний множества стандартов, составленных на основе некоторой модели описания стандарта. Задачи определения и исправления несоответствий документа реализуются с помощью подсистемы выполнения проверок, которая, зная некоторые правила, производит анализ документа и описания стандарта. Ниже следует информация для описания стандарта и документа.

Информация для описания стандарта

Каждый стандарт оформления документа задает:

- множество стилей форматирования для соответствующих видов наполнения;
- параметры страниц документа.

Каждый стиль определяет:

- начертание текста (принимает одно или несколько значений из набора: жирный, курсив, подчеркнутый или обычный – ни один из приведенных);
- название шрифта;
- регистр букв текста (принимает одно значение из набора: все строчные, все прописные, как в предложениях, каждое слово с прописной буквы);
- размер шрифта (вещественное значение в пунктах – pt);
- межстрочный интервал текста (вещественное значение в pt);

- интервал перед абзацем (вещественное значение в pt);
- интервал после абзаца (вещественное значение в pt);
- выравнивание абзаца (принимает одно значение из набора: по левому краю, по правому краю, по центру, по ширине);
- абзацный отступ (вещественное значение в сантиметрах);
- положение объекта (свойственно только объектам нетекстового типа, например рисункам и формулам; принимается одно значение из набора: в тексте, за текстом, перед текстом, вокруг рамки, по контуру);
- стиль следующего абзаца (наименование стиля абзаца, следующего за текущим, например стиль «Обычный»).

К параметрам страниц документа относятся свойства:

- поля страницы (левое, правое, верхнее, нижнее поля с вещественными значениями в сантиметрах);
- ориентация страницы (книжная или альбомная);
- размер страницы (ширина и высота с вещественными значениями в сантиметрах);
- число колонок (принимает целое значение);
- отступ между колонками (принимает вещественное значение в сантиметрах);
- перенос строк (принимает одно значение из набора: автоматический, ручной, без переносов).

Информация для описания документа

Наполнение документа разделяется на виды: текст, таблица, формула, рисунок, список.

Вид наполнения «текст» имеет свойства:

- состоит только из набора букв, цифр, знаков препинания, стандартных и нестандартных символов;
- является основным видом наполнения.

Вид наполнения «таблица» имеет свойства:

- надпись, состоящая из номера и названия таблицы (название таблицы может отсутствовать, надпись может располагаться перед или после таблицы);
- таблица состоит из частей: наименования колонок (шапка таблицы) и наполнения в колонках (тело таблицы).

Вид наполнения «формула» имеет свойства:

- номер формулы (может отсутствовать, если формула размещена в тексте);
- описание обозначений в формуле (может отсутствовать).

Вид наполнения «рисунок» имеет свойства:

– подпись, состоящая из номера и пояснения к рисунку (пояснение к рисунку может отсутствовать).

Вид наполнения «список» имеет свойства:

- уровень вложенности;
- допустимые обозначения элементов списка.

Физически все наполнение документа состоит из абзацев. Абзацы разделяются символом переноса (в рассматриваемых текстовых процессорах символ переноса ставится с помощью клавиши Ввод/Enter). В одном абзаце могут чередоваться все виды наполнения.

Логически наполнение документа структурируется и представляет собой набор структурных элементов, называемых фрагментами:

- реквизиты документа [5] (например: наименование организации, заголовок документа, город и год публикации);
- основной текст (фрагмент, который состоит из одного вида наполнения или, реже, из комбинации записи нескольких видов наполнения);
- заголовки основного текста с уровнем вложенности от 1 до 9 (обычно используется до четырех уровней);
- фрагменты, содержащие таблицы, рисунки, формулы, списки и т.д.

Для каждого фрагмента задается стиль форматирования, которому дается название на основе содержания фрагмента.

Обобщив информацию для описания стандарта и документа, можно выделить сущности «Стандарт», «Стиль», «Параметр документа», «Фрагмент», связи между ними и представить их в виде реляционной модели.

Реляционная модель данных для представления информации о стандарте

Представим информацию для описания стандарта и документа в виде реляционной модели (рис. 2).

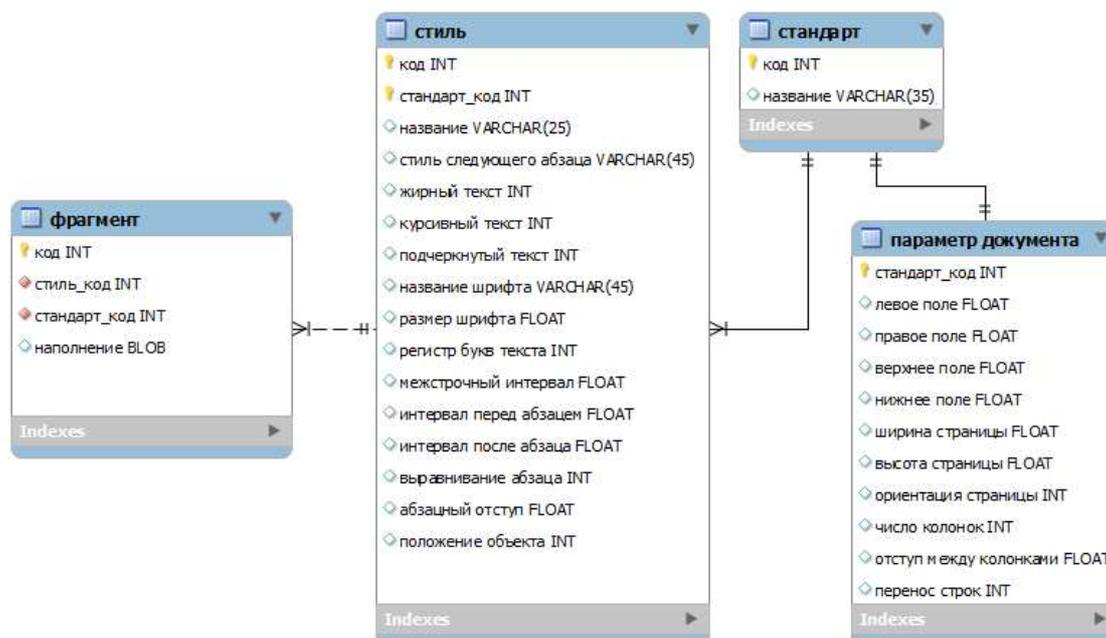


Рис. 2. Реляционная модель данных для представления стандарта

Реляционная модель приведена в денормализованном виде. Сущность «Стандарт» связана с сущностью «Стиль» в отношении «один ко многим», а сущность «Параметр документа» связана с сущностью «Стандарт» в отношении «один к одному». Сущность «Стиль» связана с сущностью «Фрагмент» в отношении «один ко многим».

Базу знаний подсистемы выполнения проверок можно представить различными способами, например с помощью продукций, использованием онтологий или регулярных выражений. Продукции хорошо подходят для анализа наполнения документа. Информация для описания стандарта и документа используется при задании условий продукций. Ниже приводится несколько примеров.

Правило продукции, определяющее рисунки с подрисуночной подписью в тексте:

«**Если** (в наполнение абзаца включено только изображение) **и** (наполнение следующего абзаца является текстом **и** абзац выровнен по центру), **то** текущий и следующий абзацы являются рисунком и подрисуночной подписью соответственно».

Продукции также хорошо подходят для приведения форматирования распознанных абзацев к стандарту.

Правило продукции для приведения форматирования рисунка и подрисуночной подписи к требованиям стандарта:

«**Если** абзац является рисунком, **то** применить стиль рисунка из стандарта»;

«**Если** абзац является подрисуночной подписью, **то** применить стиль подрисуночной подписи из стандарта».

При этом описание стиля некоторого стандарта для форматирования рисунка может быть представлено следующим образом:

- выравнивание абзаца: по центру;
- абзацный отступ: отсутствует;
- интервал перед абзацем: 12 pt;
- интервал после абзаца: 6 pt;
- положение объекта: в тексте;
- стиль следующего абзаца: «Подрисуночная подпись».

Выводы

В предыдущих разделах была приведена основная информация для описания стандарта и документа. Для формализации такой информации в терминах для компьютерной обработки хорошо подходит как реляционная модель представления данных, так и модель данных на основе онтологий. За счет онтологического подхода достигается системность, единообразие и научность в автоматизации задачи [2], и поэтому данный подход является более предпочтительным для разработки программной системы.

Если для решения проблемы нахождения в документе несоответствий стандартам использовать онтологический подход, то должны быть разработаны онтологии, определяющие системы понятий для описания конкретного стандарта и документа. При этом онтологии описания стандарта и документа используются для задания условий продукции. Такой подход позволит выполнять корректуру документов в полуавтоматическом и автоматическом режимах работы программы.

Рабочие процессы программной системы

Наиболее значимыми компонентами программной системы являются:

- библиотека описаний стандартов;
- подсистема выполнения проверок документа на соответствие стандартам и автоматической корректуры.

Рабочие процессы программной системы изображены на рис. 3.

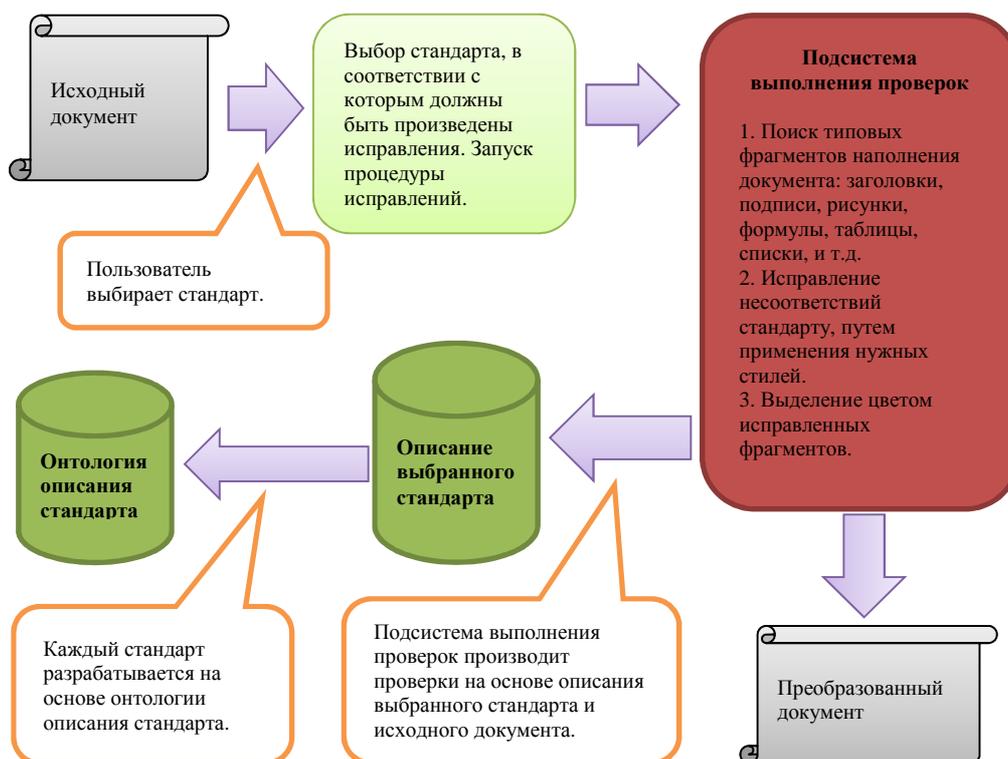


Рис. 3. Рабочие процессы программной системы

Процесс при работе пользователя с программной системой для приведения документа к требуемому стандартам виду состоит из пяти этапов:

- загрузка исходного документа в программу;
- выбор стандарта, в соответствии с которым должны быть произведены исправления;
- запуск процедуры автоматического преобразования, которая реализуется с помощью подсистемы выполнения проверок;
- оценивание пользователем результатов преобразования, путем просмотра выделенных цветом исправленных фрагментов, и, возможно, исправление ошибочных преобразований фрагментов в полуавтоматическом режиме (то есть с помощью выбора из предоставленного набора подходящего вида преобразования, рис. 4);
- сохранение преобразованного документа.

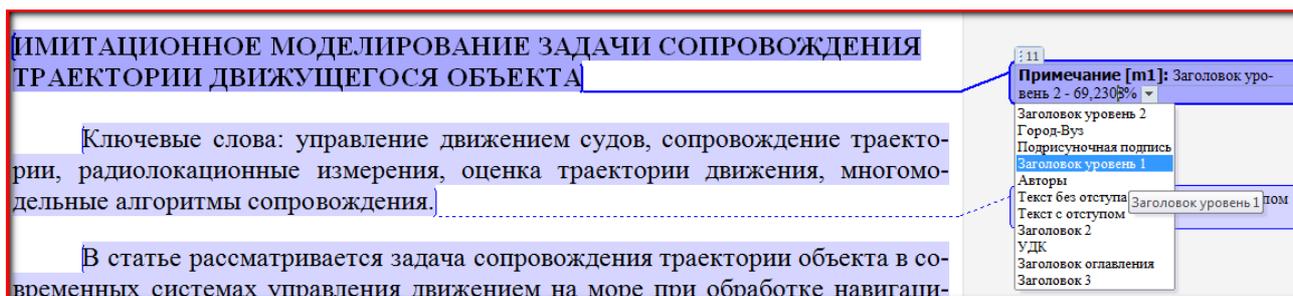


Рис. 4. Снимок экрана после выполнения процедуры автоматической корректуры документа

Текстовый процессор MS Word наиболее популярен в России, имеет мощные средства разработки с помощью VSTO [6] и VBA (встроенный язык), поэтому реализация прототипа программной системы и встраивание ее в MS Word в виде отдельного приложения (Add-In) стала наиболее возможной.

Заключение

В статье были рассмотрены существующие недостатки инструментов популярных текстовых процессоров и предложена модель программной системы, реализация которой могла бы решить проблему оформления печатных работ в соответствии с требованиями стандартов. Такая программная система может применяться для приведения любых печатных работ к требуемому виду, но является наиболее востребованной для студентов, занимающихся курсовым и дипломным проектированием, и научных деятелей, занимающихся активной публикацией научных трудов.

Список литературы

1. Акиншин А. LaTeX-шаблон для русской кандидатской диссертации // GITHUB.COM: build software better, together. - URL: <https://github.com/AndreyAkinshin/Russian-Phd-LaTeX-Dissertation-Template> (дата обращения: 15.12.2012).
2. Гаврилова Т.А. Онтологический подход к управлению знаниями при разработке корпоративных систем автоматизации // BIG.SPB.RU: БИГ-Петербург. - URL: http://www.big.spb.ru/publications/bigspb/km/ontol_podhod_to_uz.shtml (дата обращения: 24.12.2012).
3. Макросы // WORDEXPERT.RU: профессиональная работа с текстом. - URL: <http://wordexpert.ru/category/makrosy> (дата обращения: 17.12.2012).
4. Макросы // METHODICHKA.RU: пакет макросов Word. - URL: <http://methodichka.ru/macro.html> (дата обращения: 17.12.2012).
5. Состав и правила оформления реквизитов документов // EDOU.RU: центр компетенций по вопросам документационного обеспечения управления и архивного дела. - URL: http://www.edou.ru/enc/razdel11/?COURSE_ID=3&LESSON_ID=2 (дата обращения: 22.12.2012).
6. Уайтчепел Э. VSTO: создание решений на базе Microsoft Office с помощью WPF, WCF и LINQ // MSDN.MICROSOFT.COM : журнал MSDN Magazine. – 2007. - URL: <http://msdn.microsoft.com/ru-ru/magazine/cc163292.aspx> (дата обращения: 27.12.2012).

7. Файл стиля библиографического списка для оформления библиографии в MS Word по ГОСТ 7.0.5-2008 // DISAI.RU: академическая социальная сеть кафедры ИСПИ ВГУЭС. - URL: <http://disai.ru/studtracker/downloads/litgost.zip> (дата обращения: 13.12.2012).
8. Home // LYX.ORG: LyX – The Document Processor. - URL: <http://www.lyx.org/Home> (дата обращения: 07.12.2012).

Рецензенты:

Кривошеев В.П., д.т.н., профессор, профессор кафедры информационных систем и прикладной информатики Владивостокского государственного университета экономики и сервиса Минобразования РФ, г. Владивосток.

Мазелис Л.С., д.э.н., к.т.н., доцент, заведующий кафедрой математики и моделирования Владивостокского государственного университета экономики и сервиса Минобразования РФ, г. Владивосток.