

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК [021.6+025.3]:004.031.4

Мяделец  
Константин Константинович

Электронная библиотека с поддержкой онлайн каталогизации

**АВТОРЕФЕРАТ**

на соискание академической степени  
магистра технических наук

по специальности 1-40 80 04 – Математическое моделирование, численные  
методы и комплексы программ

*Подпись магистранта*

Научный руководитель  
Теслюк В.Н.  
к.т.н., доцент

*Подпись руководителя*

Минск 2015

## КРАТКОЕ ВВЕДЕНИЕ

В современном мире роль обмена информацией постоянно увеличивается. Сложно переоценить роль книги как её носителя. Однако развитие электронных библиотек, как основного способа распространения электронных книг, не может удовлетворить спроса читателей. Основная проблема – это цифровое пиратство и отношение издательств к нему.

В настоящее время копирование оцифрованной книги занимает всего несколько секунд и происходит по цене подключения к интернету. Данная особенность является камнем преткновения для правозащитников. Реактивные меры издательств заключаются в поиске и удалении незаконных копий произведений. Однако данные меры, предпринимаемые для защиты правообладателей, как правило негативно сказываются на пользователях. Человеку, купившему такую книгу, необходимо скачивать специальные программы или пользоваться устройствами, поддерживающими данный способ защиты.

Между тем большинство людей не скачивает книги в преступных целях для дальнейшей распечатки и получения прибыли, а хранят их только для личного пользования. Одной из причин «пиратского» скачивания книг является большая сложность нахождения нужной литературы. Не всегда можно понять, понравится ли тебе книга по отзывам других людей, а отдавать деньги за «кота в мешке» нет желания ни у кого.

Одним из выходов, найденных из данной ситуации, является так называемый самиздат. В этом случае автор сам выкладывает написанное или пишущееся произведение на сайт электронной библиотеки и предоставляет доступ всем желающим.

Также в данном случае возможен альтернативный вариант монетизации труда автора – платная подписка. В данном случае между автором и читателем заключается договор, по которому читатель перечисляет на счёт автора денежное вознаграждение, а автор обязуется через определённые интервалы времени высылать обновления пишущегося произведения.

Диссертационная работа посвящена решению данных проблем и созданию электронной библиотеки, которая позволяла бы авторам производить итеративную выкладку своих произведений в сети Интернет, что позволит облегчить их труд, увеличит интерактивность процесса написания произведений и заинтересованность в них читателей.

## Общая характеристика работы

### **Цель и задачи исследования**

*Целью* диссертационной работы является разработка электронной библиотеки с поддержкой каталогизации и онлайн создания произведений.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Определить необходимый функционал для наиболее эффективного использования библиотеки.
2. Разработать архитектуру программного обеспечения.
3. Разработать методы и алгоритмы для реализации требуемого функционала.
4. Реализовать программное обеспечение.
5. Провести экспериментальные исследования разработанной системы.

*Объектом* исследования являются электронные библиотеки.

*Предметом* исследования является программное обеспечение компьютерных систем для решения задач самостоятельной публикации произведений в сети Интернет.

### **Личный вклад соискателя**

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя В. Н. Теслюка, заключается в формулировке целей и задач исследования.

### **Апробация результатов диссертации**

Основные положения диссертационной работы докладывались и обсуждались на 51-ой научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2015).

### **Опубликованность результатов диссертации**

По теме диссертации опубликована 1 печатная работа в сборнике трудов и материалов 51-ой научной конференции аспирантов, магистрантов и студентов БГУИР.

### **Структура и объем диссертации**

Диссертация состоит из общей характеристики работы, введения, пяти глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ существующих аналогов, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Вторая глава посвящена разработке архитектуры ПО и поиску алгоритмов для организации электронной библиотеки. В третьей главе приводится модель данных разрабатываемой системы и предоставляется описание её элементов. В четвертой главе представлена практическая реализация ПО. Пятая глава посвящена тестированию разработанного ПО.

Общий объем работы составляет 69 страниц, из которых основного текста – 53 страниц, 38 рисунков на 14 страницах, список использованных источников из 32 наименований на 3 страницах и 3 приложения на 8 страницах.

## ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** проведён анализ существующих аналогов и анализ их недостатков. Рассмотрены сайты samlib.ru и proza.ru. На их основе определены основные направления дизайна будущей электронной библиотеки: соблюдение баланса между в структуре интерфейса – избегание как крайне низкого уровня организации, когда система не предоставляет никакой помощи в каталогизации и организации произведений, так и излишне жёсткой структуры, которая не подходит для произведений общей тематики.

**Вторая глава** посвящена разработке архитектуры ПО и выбору алгоритмов для реализации его функционала. В ней также более детально описываются требования к системе:

1. Читательская часть:
  - 1.1. Возможность лёгкого доступа к каталогу, желательно с главной страницы.
  - 1.2. Общий дизайн интерфейса должен быть как можно более статичен.

- 1.3. Возможность отслеживать произведения с выводом информации на личной странице пользователя.
  - 1.4. Возможность сортировать и отсеивать содержимое каталога по определённым полям.
  - 1.5. Показ книги и её глав без интерфейса сайта для печати или сохранения для чтения с помощью электронных книг.
  - 1.6. Сохранение книги в один из распространённых форматов электронных книг.
  - 1.7. Возможность комментировать произведение.
  - 1.8. Выдачи рекомендаций на основании явно или неявно собираемой информации о пользователе и его предпочтениях.
  - 1.9. Автоматизация процесса подписки на литературные произведения.
  - 1.10. Возможность коммуникации пользователей и авторов вне системы комментариев к произведениям.
  - 1.11. Система списков произведений, которые читатели могли бы самостоятельно редактировать и делиться ими с другими пользователями.
2. Авторская часть:
- 2.1. Управление главами произведений. Источником текста может выступать как текстовое поле в браузере (онлайн создание книги), так и файл текстового редактора (Word, OpenOffice и другие). Также требуется предоставить механизмы управления видимостью глав.
  - 2.2. Возможность добавлять к историям персонажей с кратким описанием и возможностью создания раскадровок для глав.
  - 2.3. Инструмент автоматизации платной подписки и последующей рассылки обновлений.

Также в данной главе проведён анализ существующих систем хранения форматированного текста в электронном виде и произведён выбор наиболее подходящей для разрабатываемой системы.

В связи с тем, что содержимое разрабатываемой электронной библиотеки будет в большинстве случаев просматриваться в браузере, для уменьшения числа конвертаций основным был выбран формат HTML, а для избегания межсайтовых атак было решено использовать санитизатор для удаления из

текста опасных элементов. В связи с относительно небольшим объёмом отдельных произведений и частому к ним обращению со стороны пользователя контейнером для хранения текстов была выбрана база данных.

Для реализации функционала по выдаче рекомендаций были рассмотрены следующие типы рекомендательных систем:

1. Фильтрации содержимого. Создаются профили пользователей и объектов.
  - 1.1. Профили пользователей могут включать демографическую информацию или ответы на определённый набор вопросов.
  - 1.2. Профили объектов могут включать названия жанров, имена актёров, имена исполнителей и т. п. — в зависимости от типа объекта.
2. Коллаборативная фильтрация. Используется информация о поведении пользователей в прошлом - например, информация о покупках или оценках. В этом случае не имеет значения, с какими типами объектов ведётся работа, но при этом могут учитываться неявные характеристики, которые сложно было бы учесть при создании профиля.

Исходя из специфики библиотеки был выбран алгоритм коллаборативной фильтрации с явным сбором данных о пользователе в виде создания списка отслеживаемых объектов.

Были рассмотрены следующие алгоритмы для оценки удобочитаемости:

1. Индекс Коулман-Лиану. Использует среднюю длину слова и количество предложений на 100 слов. Нет «русифицированной версии».
2. Индекс удобочитаемости Флэша. Использует количество предложений, слов и слогов, присутствующих в тексте. Существует версия, оптимизированная для использования с русским языком.
3. Индекс Ганнинга. Оценивает количество слов, предложений и сложных для понимания слов, находящихся в тексте. Требует словарь сложных слов, а также механизм их поиска по тексту.

Большинство вариантов оценки удобочитаемости привязаны к системе оценки уровня знаний, принятой в США (так называемый US grade level) и будут непонятны неподготовленному русскоговорящему пользователю.

Индекс удобочитаемости Флэша выводит удобочитаемость текста в виде шкалы от 0 до 100, где текст с оценкой 100 крайне прост для чтения, а

чтение литературы с индексом 0 доступно только людям с высшим образованием. Недостатком индекса Флэша является необходимость подсчёта слогов, однако в данном случае нет необходимости в точном определении количества и границ слогов, а только их примерное количество. Для этого можно воспользоваться подсчётом гласных с особой обработкой последовательно идущих сочетаний.

Исходя из вышеперечисленного был выбран индекс удобочитаемости Флэша.

В третьей главе рассматривается модель данных разрабатываемой системы с приведением кратких комментариев по поводу назначения классов. Диаграмма классов приведена на рисунке 1.

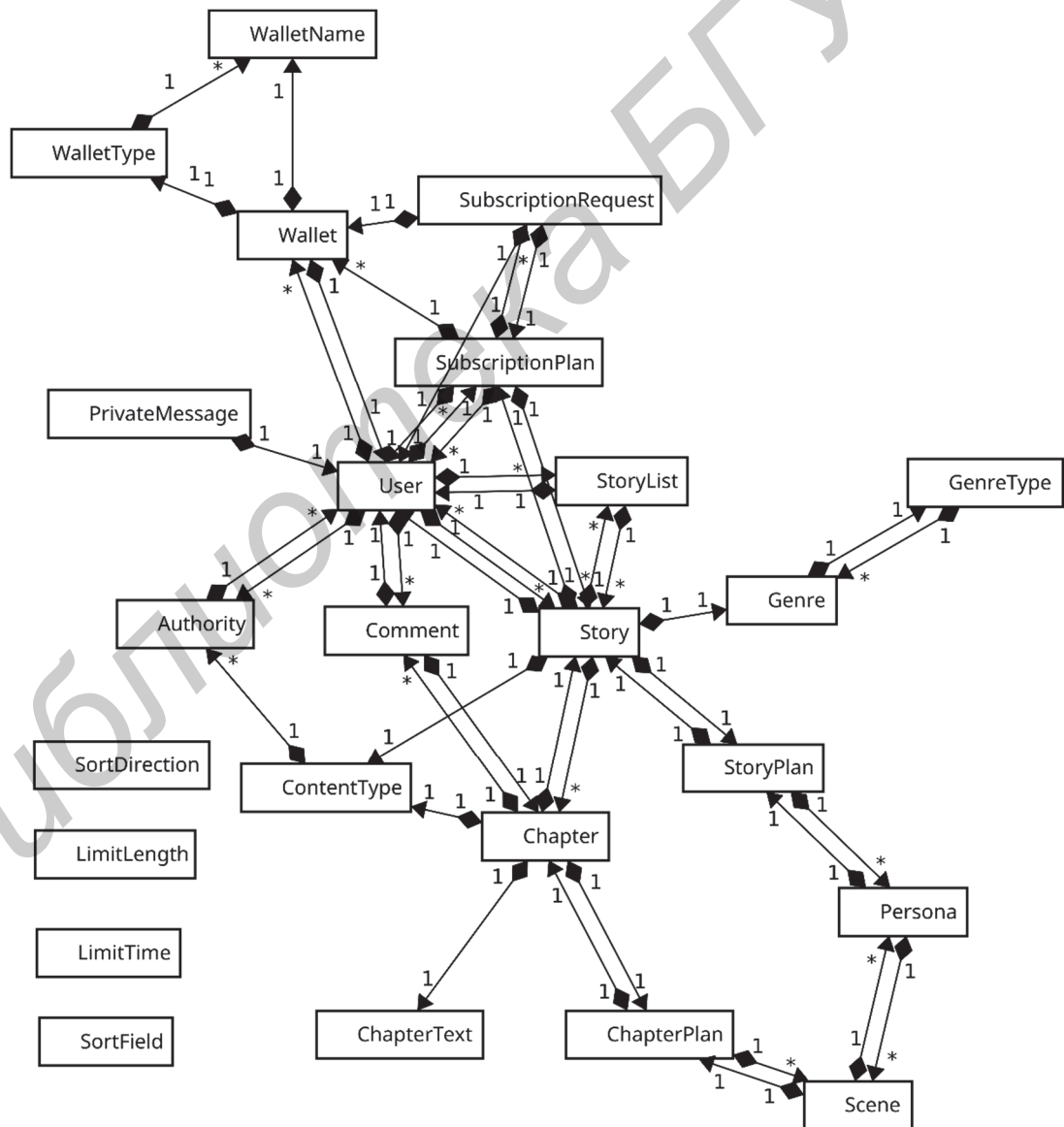


Рисунок 1 – Схема модели данных

В **четвёртой главе** рассматривается реализация электронной библиотеки с описанием слоя данных и рассмотрением наиболее сложных в плане реализации сервисных методов, а также добавление метаданных Dublin Core на страницы произведений для облегчения классификации и поиска по библиотеке сторонними инструментами. Веб-приложение построено по модели MVC с помощью библиотеки Spring Framework и её подпроекта Spring MVC. Серверная часть разделена на три слоя:

1. Слой доступа к данным. Используется для получения данных из внешнего хранилища (базы данных, веб-сервиса и т.д.).
2. Слой бизнес-логики или слой сервисов. Служит для обработки информации, поступающей в приложение из внешнего мира, а также получаемой от слой доступа к данным. Работает с слоем доступа к данным.
3. Слой контроллеров. Обрабатывает внешние подключения к приложению и отдаёт пользователю запрашиваемый контент. Работает со слоем сервисов.

Также в этой главе представлено описание инструмента для автоматизации процесса платной подписки [1].

В **пятой главе** рассматриваются инструменты, использовавшиеся для тестирования системы.

## ЗАКЛЮЧЕНИЕ

Разработанный программный продукт позволяет авторам облегчить процесс написания и публикации произведений, а также получать обратную связь от читателей. Автоматизация процесса подписки на литературные произведения также значительно облегчает процесс писательской деятельности. Для удобства последовательного написания произведения, характерного для сам-издата, в качестве основной структурной единицы произведения была выделена глава. Авторам предоставляется возможность онлайн редактирования текста главы, наряду с его загрузкой из файлов распространённых текстовых редакторов. Для улучшения качества текста предоставляется индекс удобочитаемости, который помогает авторам определять проблемные места в главах.

Читатели же в свою очередь получают удобный механизм поиска и отслеживания интересующей их литературы. Для облегчения нахождения произведений, соответствующих вкусам читателя, применяется рекомендательная система на основе соседства: пользователю рекомендуются произведения,



которые отслеживают другие пользователи со схожими интересами. Также читателям предоставляется удобный механизм просмотра произведений для различных режимов чтения. Пользователь может просматривать отдельную главу или всё произведение без форматирования сайта, а также скачать книгу в формате FB2 для дальнейшего просмотра на электронных книгах или мобильных устройствах в режиме офлайн.

Новизна проекта состоит в том, что его дизайн оптимизирован для лёгкого нахождения произведений и использования без применения стороннего программного обеспечения.

Серверная часть разработанного приложения реализована в виде веб-приложения Java с применением фреймворка Spring, что позволило использовать такие подходы, как инверсия управления и MVC, а также облегчить интеграцию использованных библиотек. В клиентской части применён адаптивный веб-дизайн, что позволяет использовать приложение с одинаковым удобством как на персональных компьютерах, так и на мобильных устройствах.

Для использования системы необходим современный веб-браузер с поддержкой JavaScript, а также минимальный набор навыков владения персональным компьютером.

## **СПИСОК ОПУБЛИКОВАННЫХ РАБОТ**

1. Мяделец К.К. Автоматизация платной подписки на литературные произведения / К.К. Мяделец // Компьютерные системы и сети: материалы 51-ой научной конференции аспирантов, магистрантов и студентов.