

Сравнительный анализ форматов файлов электронных книг

Самбулов Д.В.

Аннотация — Популярность электронных книг сегодня переживает второе рождение в связи с появлением и развитием новых устройств, таких как смартфоны или планшетные компьютеры. Появляются новые сервисы по распространению электронных книг. Сами форматы файлов за последние несколько лет пережили многие изменения, а их количество возросло. Некоторые форматы поддерживают не только статическую графику, но и такой контент как аудио и видео. Каждый формат предоставляет свою адаптацию и оптимизацию для различных данных и их представлений.

Ключевые слова — электронные книги, форматы файлов.

I. Введение

На сегодняшний день электронные книги пользуются огромной популярностью среди людей, интересующихся чтением и новинками литературных произведений. Это заставляет писателей и разработчиков электронных книг заниматься постоянным обновлением литературной базы для удовлетворения потребностей читателей.

Самбулов Дмитрий Владимирович – студент 2-го курса магистратуры факультета вычислительной математики и кибернетики Московского Государственного Университета имени М.В.Ломоносова

Именно поэтому стремительно развивается рынок электронных книг и их форматов. Сегодня насчитывается более десятка форматов электронных книг, предназначенных для различного программного обеспечения и устройств.

Популярность электронных книг подтверждается и тем фактом, что в США они впервые обогнали по продажам книги в мягких обложках. Так, согласно данным Ассоциации американских издателей, доходы от продажи электронных книг в феврале 2011 года выросли до 90,3 миллиона долларов, что почти на 170% выше аналогичного показателя годом раньше.[1] Издатели признаются, что они намного удобнее.

Сегодня рынок цифровых ридеров растет. Так, в 2012 году было продано около 2 млн читалок[2]. Интерес читателей к бумажным книгам будет падать. Исследовательская группа Forrester считает, что к 2015 году объем рынка электронных книг увеличится до 3 миллиардов долларов.

II. Характеристика электронных книг

Электронные книги – это книги в цифровом варианте. Они замечательная альтернатива бумажным книгам. В основном электронные книги получили свое широкое распространение в глобальной сети интернет. По оценкам экспертов, ежегодно количество пользователей электронных книг увеличивается на 80%.[3] Электронные книги, в основном, востребованы молодыми людьми в возрасте до 30 лет. Именно этой категории проще ориентироваться в интернете и находить в электронном хранилище необходимую информацию (познавательную, развлекательную, новостную и т.п.).

Электронные книги имеют огромный потенциал для дальнейшего развития.

Стоимость таких книг практически нулевая: они не нуждаются в бумаге для печати, не нужно привлекать типографии, так как это всего лишь файлы, которые можно просмотреть на любом персональном компьютере.

Для хранения электронных книг не требуется больших полок. На пару дисков можно поместить приличную библиотеку, насчитывающую сотни, а то и тысячи книг, и займет это совсем немного места. Это замечательное свойство также позволяет брать с собой, например, в путешествие столько книг, сколько захочется.

В любое удобное время электронные книги можно скачать в интернете. Поиск, даже среди тысяч наименований не составит ни малейшего труда.

Электронные книги удобнее в обращении, чем их бумажные аналоги. В научных работах приходится использовать много цитат, и если вы используете бумажные версии, то цитаты надо искать и выписывать. При использовании электронных книг, поиск и цитирование – это одно удовольствие. За пару секунд можно скопировать целую страницу текста и не тратить много времени на переписывание. В электронной книге очень просто перейти на нужную главу, страницу или даже найти какой-нибудь абзац. Использование электронных книг сильно упрощает и ускоряет процесс создания научной работы.

Несомненным удобством является возможность мгновенной доставки необходимого файла. Электронные книги, значительно облегчающие проблему выбора изданий и предусматривающие возможность заказа через Интернет, открывают новые возможности перед современными книжными магазинами.[4]

Не стоит забывать, что в формате электронной книги можно распространять журналы, брошюры или даже газеты. То есть, понятие электронная книга несколько шире, чем просто книга.

Следует отметить, что книги могут храниться в файлах различных форматов (как и любые данные в цифровом виде). Каждый из них имеет свои особенности и лучше или хуже подходит для хранения книг, а также для чтения на портативных устройствах. Поэтому следует разбираться в различных книжных форматах: это поможет и при выборе файла для загрузки из онлайн-библиотеки или магазина, и при подборе программы для чтения, и при покупке устройства для чтения электронных книг, если вы решите им обзавестись.[5]

III. Форматы файлов электронных книг

Существует несколько групп форматов электронных книг – графические и растровые форматы. Растровые форматы представляют объект в виде множества точек, векторные – в виде простейших геометрических фигур (точки, линии, многоугольники). Таблица внизу классифицирует их именно по этому признаку.[6]

Однако, как и любая другая, эта классификация условна. Например, PostScript (.ps, .eps), PDF могут быть чисто векторными. Так же PDF может содержать только растровые отсканированные картинки или же наоборот – содержать только текст.[6]

Существует разделение форматов электронных книг для чтения онлайн и для сохранения на компьютере пользователя. К первой группе относятся файлы такого формата, как HTML и TXT, ко второй – RTF, DjVu, PDF и др.[6]

Графические растровые форматы:

- **TIFF** - Формат для хранения изображений с большой глубиной цвета. Чаще всего используется при сканировании, отправке факсов, распознавании текста, в полиграфии, широко поддерживается графическими приложениями. Глубина изображения делает **TIFF** незаменимым форматом для тех, кто занимается оцифровкой бумажных изданий, так как процесс распознавания текста упрощается во много раз.[6]
- **JPEG** - Данный формат использует сильное сжатие, уровень которого можно регулировать, с потерей информации. Применяется для хранения высококачественных изображений, позволяет использовать до 16 миллионов цветов, именно поэтому электронные книги в этом формате также имеют место – для печати он непригоден, а для демонстрации на дисплее подходит идеально.[6]
- **DjVu** - Узкоспециализированный растровый графический формат, предназначенный для передачи текста именно в том виде, в котором он был подготовлен. Прекрасно подходит для хранения рукописей, текстов с рукописными заметками и т. д.[7]

Графические векторные форматы с оформлением:

- **RTF** - Формат, поддерживаемый большинством текстовых редакторов, созданный для документов в среде ОС Windows. обеспечивает возможность сохранения структуры текстовых документов, позволяет выделять их фрагменты (курсивом или жирным шрифтом, создавать колонки и т. п.). Книги в формате **RTF** доступны для скачивания во многих книгообменных сайтах.[6]

- **HTML** - Стандартный язык разметки документов во Всемирной паутине. Разметка позволяет вставку спецсимволов, изменение размера, типа, начертания шрифта, создание гиперссылок и таблиц, выравнивание текста. Электронные книги в формате **HTML** чаще всего выставляются в полнотекстовом виде на сайте, в этом случае в скачивании на компьютер нет необходимости.[6]
- **OPF FlipBook** - Открытый формат электронных книг – стандартная версия программы **FlipAlbum** создает альбомы-книги **FlipBooks** с расширением **OPF (Open Electronic Book Package Format)**. Эти файлы могут быть просмотрены с помощью самой программы или бесплатной утилиты для работы с **OPF**-файлами.[6]
- **CHM** - Книга – файл формата **.chm (HTMLHelp)**. Содержит в себе набор **HTML**-страниц, может также включать в себя содержание со ссылками на страницы, предметный указатель, а также базу для полнотекстового поиска по содержимому страниц. Используется для создания справки в ОС Windows и, по аналогии, для создания книг, с помощью специального программного обеспечения.[6]
- **SGML** - Стандартный общий язык разметки для документов. Широко используется в издательском деле – рукописи зачастую передаются в электронном **SGML**-структурированном виде, что значительно облегчает работу наборщиков и дизайнеров.[6]
- **XML** - Текстовый формат, предназначенный для хранения структурированных данных, для обмена информацией между программами, а также для создания на его основе более специализированных языков разметки. **XML** позволяет легко создавать

документы, готовые к непосредственному использованию и программной обработке (конвертации, хранению, управлению) в любой среде, поэтому на его основе создан формат **FictionBook**. [6]

- **FB2** - Открытый формат, основан на **XML**. Основное преимущество **FictionBook(fb2)** – возможность без труда создавать (в том числе и автоматически) книги в этом формате из файлов всех популярных текстовых форматов (*.txt, *.doc, *.rtf, *.html и пр.). Кроме того, **FictionBook** поддерживается большинством распространенных программ и устройств для чтения. [6]
- **FB3** - Формат fb3 разрабатывается для удовлетворения возросших требований к форматированию текста книг и возможностям их каталогизации. Современное состояние стандартов и технологий позволяет, помимо простого расширения набора тегов, привести формат к совместимости с наиболее удачными и перспективными наработками. В частности, в основу fb3 положен стандарт Open Packaging Convention (ECMA-376 Part 2). Помимо этого в качестве картинок можно использовать SVG-файлы (и GIF-файлы, патент на которые закончился). [6]
- **TEX** – система верстки, разработанная в целях создания компьютерной типографии. В неё входят средства для секционирования документов, для работы с перекрёстными ссылками и для набора сложных математических формул. Документы набираются на собственном языке разметки в виде файлов, содержащих информацию о форматировании текста или выводе изображений. Электронные книги обычно имеют расширение .tex, конвертируются специальной программой в файлы «.dvi», которые могут быть отображены на экране или напечатаны. **DVI**-книги можно

специальными программами преобразовать в **PostScript**, **PDF**, или другой электронный формат. [6]

- **PDF** - Формат электронных документов, созданный фирмой **Adobe Systems** с использованием ряда возможностей языка **PostScript**. Обычно используется для книг с иллюстрациями (в т.ч. схемы и математические формулы). **PDF** – формат, позволяющий внедрять необходимые, векторные и растровые изображения, формы и мультимедиа-вставки. [6]
- **LIT** - Фирменный формат Microsoft (**.lit**), достаточно распространенный, в том числе из-за поддержки производителя. Существует некоторое затруднение с прочтением, поскольку формат читается в основном с помощью **Microsoft Reader** (официального ПО). **Microsoft LIT** включает ряд функций, из-за которых его считают очень похожим на **Adobe PDF**: закладки, аннотации и масштабирование. [6]
- **PostScript и EPS** - **PostScript (.ps)**- язык составления документов, главная функция – описание страниц, чтобы при выводе на печать они выглядели так же, как на экране. Формат **EPS (Encapsulated PostScript)** был создан для сохранения графики (содержащей изображения всех типов, контуры, текстовые поля и др. – в том числе и электронных книг), предназначенной для печати на «PostScript-принтерах». **EPS**, в котором, как правило, сохраняют конечный результат работы, это упрощенная версия формата **PostScript**. [6]
- **ExeBook** - Книга в формате **ExeBook**, по сути, **Win**-приложение, скомпилированное с помощью специального программного обеспечения. Основной особенностью книги в формате **EXE** является ее максимальная приближенность к бумажному аналогу. [6]

- **RBS - (Radix-Tools REBUS)** – это защищенный формат, основанный на **HTML** и включающий в себя текстовый, графический, мультимедиа-контент и интерактивные составляющие. Использование в формате **HTML**-языка позволяет легко конвертировать файлы практически любого формата в **RBS**-вид с помощью специального ПО. В электронном виде в формате **RBS** может быть представлено множество типов бумажных изданий, как то – учебные и контрольные тесты, анкеты – результаты опросов, картографические документы, справочники, энциклопедические материалы и художественная литература.[6]
- **PRC** - Сжатый формат **PRC** – так называемый **Palm**-формат, использует механизм шифрования контента электронной книги, что обеспечивает защиту интеллектуальной собственности при ее продаже. Файлы с расширением **PRC** используются для электронных книг в формате **Mobipocket Reader** для чтения книг на КПК и телефонах.[6]
- **ePUB** - Формат, основанный на **XML**. Позволяет издателям производить и распространять цифровую публикацию в одном файле, обеспечивая совместимость между программным и аппаратным обеспечением, необходимым для воспроизведения незашифрованных цифровых книг и других публикаций с плавающей версткой.[6]

Простой текст (plain text):

- **ТХТ** - Файлы формата **ТХТ** имеют наименьший вес, поэтому зачастую использовались в самом начале эры электронных интернет-библиотек, у посетителей которых была предельно низкая скорость скачивания, равно как и объемы жестких дисков, а также

различных носителей. Для того, чтобы скачать крупное издание, создавались ZIP-архивы, объединяющие все главы книги. Основной плюс **ТХТ** – его совместимость со всеми устройствами и практически любым программным продуктом, Хотя данный формат и не поддерживает графику и сложное форматирование.[6]

IV. Открытый и закрытый формат

Форматы файлов электронных книг бывают открытыми и закрытыми.

Под открытым форматом файла подразумевается общедоступная спецификация хранения цифровых данных, обычно разрабатываемая некоммерческой организацией по стандартизации, свободная от лицензионных ограничений при использовании. В частности, должна быть возможность включать поддержку открытых форматов как в свободное(открытое), так и в проприетарное ПО, распространяемое по лицензиям, характерным для каждого из этих типов. В отличие от открытых, проприетарные форматы создаются и контролируются обычно частными компаниями и служат их интересам. Открытые форматы являются подмножеством открытых стандартов.[8]

Главная цель открытых форматов – гарантировать возможность доступа к данным в течение долгого времени безо всякой оглядки на лицензионные права и технические спецификации. Другая цель – активизировать конкуренцию вместо того, чтобы позволять компании – автору проприетарного формата препятствовать конкурирующим продуктам. В последние годы правительственные организации многих стран проявляют всё больший интерес к

открытым форматам[8]. Примером открытого формата файлов электронных книг являются:

- ОЕВ – формат, основанный на XHTML (XML) и созданный в свое время консорциумом ведущую роль в котором играла компания Microsoft. По сути, ОЕВ-книга представляет собой ZIP-архив xhtml- (xml-), графических файлов и файла с информацией о документе в целом. В своем текущем состоянии формат практически мертв, однако сейчас OpenReader Consortium строит грандиозные планы по расширению возможностей ОЕВ и его превращению в действительно универсальный формат;[7]
- DjVu – был разработан для облегчения распространения литературы через сеть Интернет. Представляет собой специальную технологию сжатия изображения. Алгоритм устроен таким образом, что запоминает изображение повторяющихся символов и хранит только одну их копию.[9]

Этот формат является разработкой компании AT&T. Он отлично подходит для сканированных текстов: книг, исторических документов, рукописей, журналов, рисунков, формул и везде, где распознавание текста является слишком трудоемким процессом. Тем не менее, DjVu позволяет хранить текстовый слой для поиска по словам, гиперссылки и даже оглавление. DjVu часто можно встретить в разнообразных онлайн-библиотеках, но для электронных читалок он является таким же "тяжелым", как и PDF.[9]

На сегодня с этим форматом более-менее научились справляться лишь некоторые устройства для чтения с диагональю экрана в 6 и более дюймов, например, PocketBook.[9]

- FB2 и ePub – оба этих формата электронных книг были разработаны на основе языка разметки XML. Эти стандарты документов призваны максимально упростить кроссплатформенность и масштабруемость. Они содержат в себе всю необходимую информацию для книги: сведения об авторе, структурированный текст, оглавление, иллюстрации, аннотацию и даже обложку. Формат ePub является открытым стандартом и даже поддерживает каскадные таблицы стилей CSS (Cascading Style Sheets – формальный язык описания внешнего вида документа, написанного с использованием языка разметки) и технологию защиты авторских прав DRM (Digital rights management). Интересной особенностью обоих этих форматов является то, что они не декларируют, как должен выглядеть документ, оставляя это на откуп программе-просмотрщику.[9]

Преимущество ePub над FB2 заключается в том, что в него можно легко вставлять математические формулы и уникальные шрифты, а в FB2 формулу можно вставить только как изображение.

- PDF (Portable Document Format) – межплатформенный формат электронных документов, созданный фирмой Adobe Systems.[10] Формат PDF был разработан для полиграфии. В нем распространяется огромное количество технической документации. Он позволяет внедрять в документ необходимые шрифты, векторные и растровые изображения, вставлять JavaScript-сценарии, формы и даже мультимедиа-контент. PDF документы поддерживают несколько типов сжатия растровой информации и переходы по гиперссылкам. Существует так же формат PDF с электронной

подписью для коммерческого использования и защиты авторских прав. После сканирования текста со сложной структурой, математическими формулами и схемами его проще перевести и хранить именно в этом формате. Недостаток формата – это его громоздкость и высокие требования к вычислительным ресурсам. [11]

PDF является открытым стандартом с 1 июля 2008 года ISO 32000.[10]

К закрытым форматам файлов электронных книг относятся:

- LIT – основан на OEB, разработан Microsoft. Индексации не поддается, воспринимается только «родной» программой MS Reader. Microsoft LIT включает ряд инструментов, которые делают его очень похожим на Adobe PDF: в частности, доступны закладки, аннотации и масштабирование. Степень сжатия текста – 50-60%;[11]
- Некоторые другие мало распространенные на сегодня форматы, например zTXT (степень компрессии до 45%, допускает возможность вставки закладок, поддерживает гиперссылки), FBR (читаются программой Franklin Reader).

Главное отличие открытого формата – то, что вся информация о нём опубликована. Её можно свободно получить и создать программу, которая будет полноценно работать с файлами этого формата – считывать информацию из них, сохранять в них новую информацию и т.п.[12]

К сожалению, многие посылают по Интернету файлы закрытых форматов (в частности, Microsoft Office). Этого не следует делать, если вы не уверены, что у получателя установлена нужная программа.

Использование открытых форматов – единственный способ обеспечить полноценный обмен информацией (в частности, через Интернет) между компьютерами во всём мире, на которых работают самые разные программы[12].

V. Заключение

У электронных книг довольно много форматов файлов. Это связано с тем, что на сегодняшний день еще не сложились единые стандарты, и многие производители предлагают свои форматы, используемые в их программах и устройствах.

Тем не менее, постепенно некоторые форматы набирают критическую массу и выходят в лидеры[13].

FB2 – это специальный формат для электронных книг, который расшифровывается как «FictionBook». В свою очередь, слово «fiction» в английском языке используют для названия художественной литературы, от слова «фикция», «выдумка». То есть литература, в основе которой лежит художественный вымысел автора. Сам формат FB2 был придуман российскими разработчиками и получил широкое распространение именно на российских книжных просторах. Поэтому многие книги в российских интернет-библиотеках и магазинах представлены именно в этом формате.[13] Для чтения книг в таком формате также нужны специальные программы. Их достаточно много.

Однако на сегодняшний день наиболее перспективным для электронных книг формат ePub, который расшифровывается как «Electronic Publication» (ePub) и представляет собой формат электронных версий книг, разработанный международным форумом по цифровым публикациям IDPF. В последнее время этот формат становится стандартом для издания электронных книг. Например,

книжный магазин iBook Store компании Apple принимает книги только в этом формате.[13]

В отличие от PDF и DjVu формат EPUB проще и легче, и не требует больших ресурсов компьютера и других устройств. Кроме того, он ближе по своим техническим характеристикам к HTML-страницам, из которых состоят сайты. Это является одной из причин широкой популярности электронных книг именно в формате EPUB, поскольку они легче интегрируются с интернетом.[13]

Для чтения электронных книг в формате EPUB, так же как и в предыдущих случаях, нужны свои программы. Среди них есть и специальный плагин для браузера Firefox, который называется EPUBReader.[13]

Таким образом, на сегодняшний день ведущими форматами электронных книг являются DjVu, pdf, fb2 и ePub.

Библиография

- 1 Электронные книги по продажам обгоняют книги в мягких обложках. [Электронный ресурс] – 2011-2013 – <http://www.segodnya.ua/world/elektronnye-knihi-po-prodazham-obhonzajut-knihi-v-mjahkikh-oblozhkakh.html>
- 2 Евгений Милица: Новая эра развития электронных книг началась. [Электронный ресурс] – 2013 – <http://www.cnews.ru/reviews/?2013/04/17/526040>
- 3 Электронная книга на 22-й Московской международной книжной выставке-ярмарке [Электронный ресурс] – 2009-2013 – <http://www.iqlib.ru/news/news/66586>
- 4 Электронные книги [Электронный ресурс] – 2013 – [013/02/10/%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5%D0%BA%D0%BD%D0%B8%D0%B3%D0%B8/](http://idyllwildphonebook.carseatfront.com/2013/02/10/%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5%D0%BA%D0%BD%D0%B8%D0%B3%D0%B8/)
- 5 Электронная книга (документ). [Электронный ресурс] – 2013 – http://ru.wikipedia.org/wiki/%D0%AD%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BA%D0%BD%D0%B8%D0%B3%D0%B0_%28%D0%B4%D0%BE%D0%BA%D1%83%D0%BC%D0%B5%D0%BD%D1%82%29
- 6 Форматы электронных книг. [Электронный ресурс] – 2008-2013 – http://reeed.ru/ru_formats.php
- 7 Форматы для электронных книг. [Электронный ресурс] – 2008-2013 – <http://old.computerra.ru/print/35629/>
- 8 Открытый формат. [Электронный ресурс] – 2008-2013 – http://ru.wikipedia.org/wiki/%D0%9E%D1%82%D0%BA%D1%80%D1%8B%D1%82%D1%8B%D0%B9_%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82
- 9 Форматы электронных книг – беспредел разнообразия. [Электронный ресурс] – 2008-2013 – <http://blog.ibooki.com.ua/2010/08/formaty-elektronnyx-knig-bespredel-raznoobraziya/>
- 10 Portable Document Format [Электронный ресурс] – 2008-2013 – http://ru.wikipedia.org/wiki/Portable_Document_Format
- 11 Форматы электронных книг. [Электронный ресурс] – 2008-2013 – http://www.ph4.ru/ebook_Form.ph4

- 12 Закрытые и открытые форматы. Стандартные форматы. [Электронный ресурс] – 2008-2013 – <http://gimnnik.narod.ru/open-office/TextProcessor/p130aa1.html>
- 13 Форматы электронных книг. [Электронный ресурс] – 2011-2013 – <http://www.izdat-knigu.ru/formaty-elektronnyh-knig/>