Дудченко С.В. УДК 02+025.4(0.064) ЭЛЕКТРОННЫЙ ДОКУМЕНТ КАК ПОЛНОТЕКСТОВЫЙ, СПРАВОЧНЫЙ И БИБЛИОГРАФИЧЕСКИЙ ПЕРВОИСТОЧНИК

В статье исследуется проблема создания полнотекстовых электронных книг и журналов на основе новой структуры электронного документа, показана эффективность использования информации в предложенном формате представления, приведены схемы созданных электронных документов, доказывается их соответствие печатным аналогам и предлагается путь перевода из электронного «самиздата» в электронный первоисточник.

Ключевые слова: электронная книга, электронный журнал, первоисточник, электронный документ.

The article presents a study of creating a full textual version of an electronic book or journal by means of a new technique of electronic document structure formation; it demonstrates the efficiency of using information in the suggested presentation format and provides developed schemes of electronic documents, proving the conformity of the latter with their printed analogues and suggesting a way of transforming them from an electronic "samizdat" / "reprinted" / version to an electronic original source of information.

Key words: electronic book, electronic journal, original source of information, electronic document.

В последнее время компьютерные технологии все больше и больше входят в издательские и библиотечные процессы. Активно создаются электронные библиотеки различного тематического направления, при этом возникает «проблема быстрого создания электронных ресурсов» и проблема прямого соответствия электронных документов своему целевому назначению [1].

Для решения этой задачи необходимо ответить на несколько вопросов, один из которых — что же такое электронная книга? То же самое, что и книга, отпечатанная традиционным способом — на бумаге, но переведенная на электронный носитель информации. Такой ответ на этот вопрос можно получить, рассмотрев проблему, кому нужна электронная книга и каков критерий оценки документов подобного типа (в данном случае учебного и научного содержания) у пользователя.

Общеизвестно, что, кроме первичной информации (непосредственно содержательного наполнения – текста), в «научной литературе» существует и вторичная информация [2], как-то расположение первичной информации (текста) на странице документа, и ссылки на документ (источник).

Можно ли сделать ссылку на электронный документ, у которого зачастую отсутствует такое понятие, как страница? Да, можно, ² но электронные документы, особенно те, которые располагаются на страницах электронных библиотек различных сайтов, из-за «нестабильности» Интернета (сегодня сайт есть – завтра нет) и при 90%-ом отсутствии ответственности за точность передачи информации, не пользуются авторитетом: практически все пользователи, занимающиеся научными изысканиями, делают ссылки на традиционные «бумажные» первоисточники. Помимо всего прочего, электронный или бумажный документ должен быть зарегистрирован. Общеизвестно, что каждый печатный документ имеет регистрационный номер, для книг ISBN, для периодических изданий ISSN, в противном случае это самиздат.

Что же мы видим в различного рода электронных библиотеках? Перепечатка либо оцифровка текстов ограничивается лишь представлением текста (так называемого «содержательного наполнения»), не всегда проверенного с точки зрения соответствия печатному оригиналу.

Как известно, распознавание текста различными программами, предназначенными для этих целей, происходит не идеально. Поэтому многие структуры в высших учебных заведениях, информационных центрах, научных библиотеках, занимающиеся созданием электронных библиотек учебно-научного содержания, столкнувшись с неправильным распознаванием текстов печатных источников, решили отказаться от этого трудоемкого процесса. Любимый формат таких создателей электронных библиотек DjVu. Казалось бы, замечательный формат: точная передача изображения страницы, малый «вес» файла, создание гиперссылок по тексту, чем не электронная книга?

Теперь рассмотрим, как этот формат представляется на подавляющем большинстве сайтов электронных библиотек. ³ Это «фотография» страницы, с которой можно произвести следующие действия:

- прочитать текст с экрана компьютера;
- распечатать страницы на принтере;

¹ «Научная литература – совокупность произведений, освещающих научные проблемы и служащих первоисточниками для научного исследования»[2, с.109].

² Нужно привести электронный адрес страницы в Интернете и указать время снятия (копирования) информации с сайта.

³ Из этого списка исключим библиотеки, использующие данный формат для создания баз данных «старопечатных книг», а также библиотеки, работающие с форматом DjVu в полном объеме, то есть с распознаванием текста на странице. В этом случае работа над созданием электронного текста в формате DjVu затягивается и становится по времени создания равной работе по распознаванию текста (Π pumeu. автора).

 распознать файл в формате DjVu в программе распознавания текстов типа ABBYY FineReader (для тех, кто использует такого рода программы).

Из опыта обслуживания читателей по электронным документам, сделанным в различных форматах, формат DjVu читатель, не владеющий специализированными программами для распознавания текстов, приравнивает к таким форматам, как jpg, bmp, tif, и прочим форматам «картинок». Если еще к этому прибавить то, что создатели таких библиотек зачастую пренебрегают сканированием неинформативных, по их мнению, титульных страниц, страниц с введением, содержанием, то в результате получается электронный аналог печатной книги с оторванной коркой и вырванными страницами. Единственным достоинством такого документа является точное совпадение текста «оставшихся» страниц электронного документа со страницами печатного оригинала. В результате имеем дело не с электронным документом, а с электронной копией печатного документа⁴. Теперь читатель может сослаться на необходимую страницу, как на печатный источник. Но все полезные свойства этого электронного документа на этом заканчиваются — различные виды поисков, применяемые в электронных документах, в электронной копии недоступны.

Возникло некоторое противоречие. «Быстросозданная» электронная копия печатного документа, которая не содержит «ошибок» сканирования, но не имеет некоторых полезных свойств электронного документа, таких, как копирование текста и его вставка в другой документ и различные виды поисков по тексту с использованием компьютерных средств.

Если добавить к этому еще и отсутствие правильного библиографического описания (которое также бывает необходимо для составления, например, списка использованной литературы), то говорить об авторском праве для создания электронного документа уже не приходится.

Существует мнение о том, что автор печатного издания вправе создать еще и электронную версию, причем создать ее где угодно. Совершенно верное высказывание, но под «где угодно» следует понимать все-таки издательство, которое может присвоить уже электронной версии ISBN, в противном случае мы будем иметь дело с самиздатом, но уже электронным. При наличии ISBN электронный документ уже может отличаться от печатного, поскольку он может быть выпущен (зарегистрирован) другим издательством.

Но большинство электронных библиотек создается путем сканирования печатных документов [3], т.е. создаются самиздатные электронные копии.

К сожалению, по данной теме вопросов больше чем ответов: электронный документ или электронная копия печатного документа, регистрированное издание или самиздат, авторское право, требования создателей электронных документов и требования читателей?

Решение, которое рассматривается далее, в своей первоначальной редакции, в основном сформировано с учётом требований читателей к электронным документам:

- 1. Точное совпадение электронной копии с бумажным оригиналом (постранично).
- 2. Специально созданная поисковая система (от примитивного поиска по странице электронной копии печатного документа до «продвинутого» поиска по любому слову из текста документа).
- 3. Возможность копирования текста или части текста в новый документ.
- Наличие правильного библиографического описания (для читателей, занимающихся научной работой).

Кроме вышеизложенных читательских требований, существуют требования, исходящие из внедрения в библиотечную практику новых компьютерных технологий и применения для научной обработки документов различных автоматизированных систем и программ. Поэтому сейчас, когда практически все библиотеки ведут электронные каталоги, необходимо вместе с печатной или электронной версией поставлять на электронном носителе и библиографическое описание в MARC-совместимом формате, пригодном для импорта в электронный каталог той или иной библиотеки. Эта проблема назрела и давно требует решения. Если издательства наконец-то подключатся к составлению библиографических описаний, создание электронных каталогов сдвинется с мертвой точки, так как пока подавляющее большинство библиотек создаёт свои электронные каталоги, ограничиваясь введением информации о новых поступлениях [4].

Поэтапно рассмотрим технологию и методику создания электронного варианта продолжающегося издания, которая учитывает все вышеперечисленные требования к электронному документу. Сделаем это на примере создания электронной версии журнала «Культура народов Причерноморья» (Приложение №1):

1. Для подготовки печатного варианта готовится оригинал-макет, который затем перерабатывается в электронную копию печатного документа, в любом из «устойчивых» форматов («устойчивым» считается формат, когда представление документа не изменяется в зависимости от просмотровщика; к «неустойчивым» форматам, например, относится формат MS WinWord, в котором представление документа зависит не только от версии MS WinWord, но и от шаблона, в котором сделан документ).

⁴ Автор статьи предлагает ввести в научный оборот термин «электронная копия печатного документа»

Для журнала «Культура народов Причерноморья» был принят формат документа pdf, т.к. данный формат позволяет копировать не только выделенный фрагмент в качестве рисунка, но и в виде текста с возможностью копирования его в другие документы.

- 2. Электронные копии отдельных печатных номеров издаются по мере выхода в свет каждого печатного номера. Отдельные номера записываются на компакт-диски типа «визитка» (название данного компакт-диска говорит об его прямом предназначении, в данном случае диск используется в качестве носителя электронной копии печатного номера журнала).
- 3. По мере выхода номеров журнала происходит накопление информации и создается сводный выпуск с определенным количеством номеров (первый выпуск электронной версии журнала, который был выпущен в 2003 году, содержал №№ с 1 по 35; сейчас готовится второй выпуск, который охватывает №№ с 36 по 50).
- 4. Сводный электронный выпуск оснащается поисковыми системами:
 - 4.1. Первая с использованием Интернет-технологий: поиск по ключевым словам, по словам из библиографических описаний, по любым словам из аннотаций, представленных в сводном выпуске.
 - 4.2. Вторая с применением библиотечных технологий и с использованием системы «ИРБИС»: поиск по электронному каталогу с возможностью экспорта данных в электронные каталоги потенциальных пользователей.

То, что электронная версия журнала делается с оригинал-макетов печатного издания, исключает расхождение между электронной и печатной версиями журнала. Это предоставляет читателю возможность использовать электронный документ и при этом ссылаться на печатный. Свойства электронного документа и поисковые системы, использованные при создании электронной версии сводных выпусков, а также система обработки информации (полная аналитико-синтетическая обработка документа), дают возможность читателю получить не только полнотекстовую, но краткую и полную библиографическую информацию

Рассмотрим технологию создания электронного варианта другого печатного издания: Петров О.В.⁵ Гніздовий тлумачно-словотворчий словник композитів. – Сімферополь, 2003. – 286 с. – Рос. мовою. – ISBN 966-572-386-3 (Приложение №2)

- 1. Для подготовки электронного варианта у автора был взят оригинал-макет (по которому в издательстве был создан печатный документ), и на его основе создана электронная копия печатного документа, в формате *pdf*.
- 2. Электронная книга оснащена поисковой системой с использованием Интернет-технологий, имеется поиск по ключевым словам, по списку использованной литературы, по любому слову из словообразовательных гнезд.
- 3. Кроме того, на книгу составлено электронное библиографическое описание, которое представлено в различных библиографических форматах UNIMARC, USMARC, RUSMARC, а также в формате библиотечной системы «ИРБИС».
- 4. Электронная копия записана на компакт-диск типа «визитка».

Подведем итог исследования:

Разработана методика и технология формирования электронного документа, включающего в себя, кроме электронной копии печатного документа, дополнительные функциональные а также информационные возможности.

Предложена последовательность создания электронного документа, который по своей сути является полной копией печатного документа и сам может быть использован как первоисточник:

Электронная копия печатного документа создается из **оригинал-макета** печатного издания и представляется в любом из **«устойчивых»** форматов, но с возможностью работы с текстовой частью документа

Электронная копия оснащается поисковой системой (в зависимости от вида издания, начиная от простейшего поиска по содержанию и заканчивая сложным поиском по любому слову из текста документа).

Составляется полное библиографическое описание документа в MARC-совместимых форматах.

Электронная версия дополняется сведениями с сопутствующей информацией (сведения об авторе, список дополнительных источников по данной тематике: библиография работ автора, аудиовизуальная информация и т.п.)

На основе составляющих компонентов электронного документа разрабатывается титульный интерфейс электронной версии документа.

Созданный электронный документ записывается на компакт-диск типа «визитка» и прилагается к печатному оригиналу в качестве электронной копии на другом носителе.

В случае издания электронной и печатной версии в одном издательстве регистрации электронной вер-

⁵ См.: Приложение №3 – отзыв к.ф.н., доцента Петрова А.В.

сии не требуется.

Самостоятельное издание электронной версии требует регистрации.

При соблюдении данных технологических требований к созданию электронных документов уровень представления информации на электронных носителях переходит на уровень первоисточника.

Источники и литература

- 1. Ядрова Г.В. Сотрудничество и координация как инструмент создания электронных образовательных ресурсов и новый путь распространения программы «ИРБИС» / Г.В. Ядрова // Культура народов Причерноморья. Симферополь: Межвузовский центр «Крым». 2004. N48. Т.2. С. 10-14. Библиогр.: 5 назв.
- 2. Терминологический словарь по библиотечному делу и смежным отраслям знания / Сост.: 3.Г. Высоцкая (отв. ред.), В.А. Врубель, А.Б. Маслов, Л.К. Розеншильд; РАН. Б-ка по естественным наукам. – М., 1995. – 268 с.
- 3. Земсков А.И., Шрайберг Я.Л. Электронные библиотеки: Учебник для студентов вузов культуры и искусств и др. высших учеб. заведений / А.И. Земсков, Я.Л. Шрайберг. М.: Либерия, 2003. 352 с. Библиогр.: 110 назв.
- 4. Ядрова Г.В. Электронный каталог: проблема полного отражения фондов библиотеки вуза / Г.В. Ядрова // Культура народов Причерноморья. Симферополь: Межвузовский центр «Крым». 2003. N40. С. 20-24. Библиогр.: 9 назв.

Приложение №1



Рис.1. Интерфейс второго сводного выпуска журнала «Культура народов Причерноморья».

ПОИСК ПО АВТОРУ, СЛОВАМ ИЗ БИБЛИОГРАФИЧЕСКОГО ОПИСАНИЯ, КЛЮЧЕВЫМ СЛОВАМ
ПОИСК

Показанные результаты **1-10** найдено **80** искалось **36** на поиск ушло **0.09** секунд.

Вахрушев И.Б. Опасные геоморфологические явления и экологическая ситуация в Крыму / И.Б. Вахрушев // Культура народов Причерноморья. - Симферополь: Межвузовский центр 'Крым'. - 2002. - N36. - С. 9-12. - Библиогр.: 16 назв.

Кл.слова:reorрафия Крыма; экология Крыма; карст; абразия; сели; Крымские горы; ополэни; обвалы; природные катастрофы; геоморфологические процессы knp36/knp36_9_12.pdf

Воронін І.М. Телекомунікаційні системи України: економіко-географічний аналіз/ І.М. Воронін // Культура народов Причерноморья. -Симферополь: Межвузовский центр 'Крым'. - 2002. - N36. - С. 13-16. - Библиогр.: 8 назв.

Кл.слова:связь Украина 2000 г.; телекоммуникационные системы Украины; преподаватель ТНУ; Интернет; компьютерная связь; спутниковая связь; твейджинг; транкинг; пейджинг; ОСС; "Киевстар GSM"; сотовая связь; UМС; "Украинская мобильная связь; экономико-географический анализ; информационная индустрия; география связи; мобильная связь; телефонная связь; виды связи knp36/knp36_13_16.pdf

Пилипенко І.О. Зміна природних умов формування рибних ресурсів в Азовському морі/ І.О. Пилипенко // Культура народов Причерноморья. - Симферополь: Межвузовский центр 'Крым'. - 2002. - N36. - С. 16-19. - Библиогр.: 4 назв.

Кл.слова:водное хозяйство Крыма; рыбные ресурсы; биологическое загрязнение; морское хозяйство; природные условия; Азовское море

knp36/knp36_16_19.pdf

Рис.2. Фрагмент поиска по втором сводному выпуску журнала «Культура народов Причерноморья».

Приложение №2

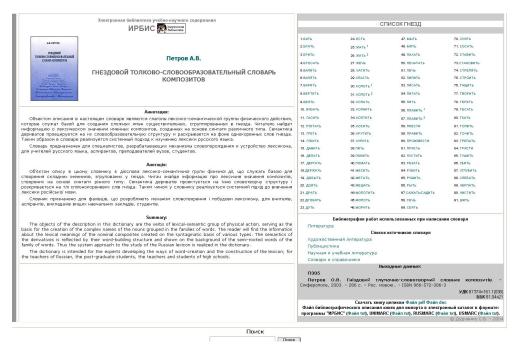


Рис.3. Интерфейс книги Петров О.В. Гніздовий тлумачно-словотворчий словник композитів.



Рис.4. Вариант поиска по производным словообразовательных гнезд словаря.

Приложение №3

Отзыв автора печатного издания «Гнездового толково-словообразовательного словаря композитов» к.ф.н., доцента кафедры русского языкознания Таврического национального университета им. В.И. Вернадского А.В. Петрова

Предложенная методика создания электронного документа сохраняет все свойства печатной версии документа: наблюдается полное постраничное совпадение бумажного варианта словаря и электронного, что позволяет использовать электронный вариант словаря в качестве первоисточника. Представленная методика формирования электронного документа вносит целый спектр новых полезных свойств, которые в полной мере могут быть использованы в учебной и научной работе.

Предлагаемая методика позволяет исследовать документ с помощью различных видов поиска:

- по гнездам (список гнезд вынесен на титульный лист электронного словаря, что дает полноценную картину структуры словаря);
- по производным словам гнезда;
- по корневой части сложного слова;
- по источнику библиографических сведений.

Перечисленные достоинства электронного документа свидетельствуют о том, что данная методика создания электронного документа может быть эффективно использована в научно-исследовательской работе, а также при создании новых типов электронных словарей с расширенными поисковыми возможностями.

JON -